UC Santa Cruz
UC Santa Cruz Previously Published Works

Title

The contribution of historical processes to contemporary extinction risk in placental mammals

Permalink

https://escholarship.org/uc/item/1rq5p0s8

Journal

Science, 380(6643)

ISSN

0036-8075

Authors

Wilder, Aryn P
Supple, Megan A
Subramanian, Ayshwarya
et al.

Publication Date

2023-04-28

DOI

10.1126/science.abn5856

Peer reviewed

*Corresponding authors: awilder@sdzwa.org (APW), msupple@ucsc.edu (MAS), oryder@sdzwa.org (OAR), and bashapir@ucsc.edu (BS).

†co-lead authors

‡see list at end of manuscript

§co-senior authors

Author contributions:

Conceptualization: APW, MAS, ASA, CS, KPK, DPG, EKK, KLT, TMB, ZC, OAR, BS

Data analysis: APW, MAS, AS, AM, RS, ASA, VMF, KF, WKM

Interpretation of results: APW, MAS, AS, AM, OAR, BS, with input from all authors

Writing – original draft: APW, MAS, BS

Writing – review & editing: all authors

Competing interests:

Authors declare that they have no competing interests.

‡Zoonomia Consortium Members:

Gregory Andrews[1], Joel C. Armstrong[2], Matteo Bianchi[3], Bruce W. Birren[4], Kevin R. Bredemeyer[5], Ana M. Breit[6], Matthew J. Christmas[3], Hiram Clawson[2], Joana Damas[7], Federica Di Palma[8,9], Mark Diekhans[2], Michael X. Dong[3], Eduardo Eizirik[10], Kaili Fan[1], Cornelia Fanter[11], Nicole M. Foley[5], Karin Forsberg-Nilsson[12,13], Carlos J. Garcia[14], John Gatesy[15], Steven Gazal[16], Diane P. Genereux[4], Linda Goodman[17], Jenna Grimshaw[14], Michaela K. Halsey[14], Andrew J. Harris[5], Glenn Hickey[18], Michael Hiller[19,20,21], Allyson G. Hindle[11], Robert M. Hubley[22], Graham M. Hughes[23], Jeremy Johnson[4], David Juan[24], Irene M. Kaplow[25,26], Elinor K. Karlsson[1,4,27], Kathleen C. Keough[17,28,29], Bogdan Kirilenko[19,20,21], Klaus-Peter Koepfli[30,31,32], Jennifer M. Korstian[14], Amanda Kowalczyk[25,26], Sergey V. Kozyrev[3], Alyssa J. Lawler[4,26,33], Colleen Lawless[23], Thomas Lehmann[34], Danielle L. Levesque[6], Harris A. Lewin[7,35,36], Xue Li[1,4,37], Abigail Lind[28,29], Kerstin Lindblad-Toh[3,4], Ava Mackay-Smith[38], Voichita D. Marinescu[3], Tomas Marques-Bonet[39,40,41,42], Victor C. Mason[43], Jennifer R. S. Meadows[3], Wynn K. Meyer[44], Jill E. Moore[1], Lucas R. Moreira[1,4], Diana D. Moreno-Santillan[14], Kathleen M. Morrill[1,4,37], Gerard Muntané[4], William J. Murphy[5], Arcadi Navarro[39,41,45,46], Martin Nweeia[47,48,49,50], Sylvia Ortmann[51], Austin Osmanski[14], Benedict Paten[2], Nicole S. Paulat[14], Andreas R. Pfenning[25,26], BaDoi N. Phan[25,26,52], Katherine S. Pollard[28,29,53], Henry E. Pratt[1], David A. Ray[14], Steven K. Reilly[38], Jeb R. Rosen[22], Irina Ruf[54], Louise Ryan[23], Oliver A. Ryder[55,56], Pardis C. Sabeti[4,57,58], Daniel E. Schäfer[25], Aitor Serres[24], Beth Shapiro[59,60], Arian F. A. Smit[22], Mark Springer[61], Chaitanya Srinivasan[25], Cynthia Steiner[55], Jessica M. Storer[22], Kevin A. M. Sullivan[14], Patrick F. Sullivan[62,63], Elisabeth Sundströ[3], Megan A. Supple[59], Ross Swofford[4], Joy-El Talbot[64], Emma Teeling[23], Jason Turner-Maier[4], Alejandro Valenzuela[24], Franziska Wagner[65], Ola Wallerman[3], Chao Wang[3], Juehan Wang[16], Zhiping Weng[1], Aryn P. Wilder[55], Morgan E. Wirthlin[25,26,66], James R. Xue[4,57], Xiaomeng Zhang[4,25,26]

Affiliations:

[1]Program in Bioinformatics and Integrative Biology, UMass Chan Medical School; Worcester, MA 01605, USA.

[2]Genomics Institute, University of California Santa Cruz; Santa Cruz, CA 95064, USA.

[3]Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University; Uppsala, 751 32, Sweden.

[4]Broad Institute of MIT and Harvard; Cambridge, MA 02139, USA.

[5]Veterinary Integrative Biosciences, Texas A&M University; College Station, TX 77843, USA.

[6]School of Biology and Ecology, University of Maine; Orono, ME 04469, USA.

[7]The Genome Center, University of California Davis; Davis, CA 95616, USA.

[8]Genome British Columbia; Vancouver, BC, Canada.

[9]School of Biological Sciences, University of East Anglia; Norwich, UK.

[10]School of Health and Life Sciences, Pontifical Catholic University of Rio Grande do Sul; Porto Alegre, 90619–900, Brazil.

[11]School of Life Sciences, University of Nevada Las Vegas; Las Vegas, NV 89154, USA.

[12]Biodiscovery Institute, University of Nottingham; Nottingham, UK.

[13]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University; Uppsala, 751 85, Sweden.

[14]Department of Biological Sciences, Texas Tech University; Lubbock, TX 79409, USA.

[15]Division of Vertebrate Zoology, American Museum of Natural History; New York, NY 10024, USA.

[16]Keck School of Medicine, University of Southern California; Los Angeles, CA 90033, USA.

[17]Fauna Bio Incorporated; Emeryville, CA 94608, USA.

[18]Baskin School of Engineering, University of California Santa Cruz; Santa Cruz, CA 95064, USA.

[19]Faculty of Biosciences, Goethe-University; 60438 Frankfurt, Germany.

[20]LOEWE Centre for Translational Biodiversity Genomics; 60325 Frankfurt, Germany.

[21]Senckenberg Research Institute; 60325 Frankfurt, Germany.

[22]Institute for Systems Biology; Seattle, WA 98109, USA.

[23]School of Biology and Environmental Science, University College Dublin; Belfield, Dublin 4, Ireland.

[24]Department of Experimental and Health Sciences, Institute of Evolutionary Biology (UPF-CSIC), Universitat Pompeu Fabra; Barcelona, 08003, Spain.

[25]Department of Computational Biology, School of Computer Science, Carnegie Mellon University; Pittsburgh, PA 15213, USA.

[26]Neuroscience Institute, Carnegie Mellon University; Pittsburgh, PA 15213, USA.

The contribution of historical processes to contemporary extinction risk in placental mammals

Aryn P. Wilder[1,†,*], **Megan A Supple**[2,3,†,*], **Ayshwarya Subramanian**[4], **Anish Mudide**[5], **Ross Swofford**[4], **Aitor Serres-Armero**[6], **Cynthia Steiner**[1], **Klaus-Peter Koepfli**[7,8,9], **Diane P. Genereux**[4], **Elinor K. Karlsson**[4,10], **Kerstin Lindblad-Toh**[4,11], **Tomas Marques-Bonet**[6,12,13,14], **Violeta Munoz Fuentes**[15], **Kathleen Foley**[16,17], **Wynn K. Meyer**[17], **Zoonomia Consortium**[‡],

Oliver A. Ryder[1,18,§,*], **Beth Shapiro**[2,3,§,*]

[1]Conservation Genetics, San Diego Zoo Wildlife Alliance; Escondido, CA 92027, USA.

[27]Program in Molecular Medicine, UMass Chan Medical School; Worcester, MA 01605, USA.

[28]Department of Epidemiology & Biostatistics, University of California San Francisco; San Francisco, CA 94158, USA.

[29]Gladstone Institutes; San Francisco, CA 94158, USA.

[30]Center for Species Survival, Smithsonian's National Zoo and Conservation Biology Institute; Washington, DC 20008, USA.

[31]Computer Technologies Laboratory, ITMO University; St. Petersburg 197101, Russia. [32]Smithsonian-Mason School of Conservation, George Mason University; Front Royal, VA 22630, USA.

[33]Department of Biological Sciences, Mellon College of Science, Carnegie Mellon University; Pittsburgh, PA 15213, USA.

[34]Senckenberg Research Institute and Natural History Museum Frankfurt; 60325 Frankfurt am Main, Germany.

[35]Department of Evolution and Ecology, University of California Davis; Davis, CA 95616, USA.

[36]John Muir Institute for the Environment, University of California Davis; Davis, CA 95616, USA.

[37]Morningside Graduate School of Biomedical Sciences, UMass Chan Medical School; Worcester, MA 01605, USA.

[38]Department of Genetics, Yale School of Medicine; New Haven, CT 06510, USA.

[39]Catalan Institution of Research and Advanced Studies (ICREA); Barcelona, 08010, Spain.

[40]CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST); Barcelona, 08036, Spain.

[41]Department of Medicine and LIfe Sciences, Institute of Evolutionary Biology (UPF-CSIC), Universitat Pompeu Fabra; Barcelona, 08003, Spain.

[42]Institut Catalàde Paleontologia Miquel Crusafont, Universitat Autòoma de Barcelona; 08193, Cerdanyola del Vallè, Barcelona, Spain.

[43]Institute of Cell Biology, University of Bern; 3012, Bern, Switzerland.

[44]Department of Biological Sciences, Lehigh University; Bethlehem, PA 18015, USA.

[45]BarcelonaBeta Brain Research Center, Pasqual Maragall Foundation; Barcelona, 08005, Spain.

[46]CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST); Barcelona, 08003, Spain.

[47]Department of Comprehensive Care, School of Dental Medicine, Case Western Reserve University; Cleveland, OH 44106, USA.

[48]Department of Vertebrate Zoology, Canadian Museum of Nature; Ottawa, Ontario K2P 2R1, Canada.

[49]Department of Vertebrate Zoology, Smithsonian Institution; Washington, DC 20002, USA.

[50]Narwhal Genome Initiative, Department of Restorative Dentistry and Biomaterials Sciences, Harvard School of Dental Medicine; Boston, MA 02115, USA.

[51]Department of Evolutionary Ecology, Leibniz Institute for Zoo and Wildlife Research; 10315 Berlin, Germany.

[52]Medical Scientist Training Program, University of Pittsburgh School of Medicine; Pittsburgh, PA 15261, USA.

[53]Chan Zuckerberg Biohub; San Francisco, CA 94158, USA.

[54]Division of Messel Research and Mammalogy, Senckenberg Research Institute and Natural History Museum Frankfurt; 60325 Frankfurt am Main, Germany.

[55]Conservation Genetics, San Diego Zoo Wildlife Alliance; Escondido, CA 92027, USA.

[56]Department of Evolution, Behavior and Ecology, School of Biological Sciences, University of California San Diego; La Jolla, CA 92039, USA.

[57]Department of Organismic and Evolutionary Biology, Harvard University; Cambridge, MA 02138, USA.

[58]Howard Hughes Medical Institute; Chevy Chase, MD, USA.

[59]Department of Ecology and Evolutionary Biology, University of California Santa Cruz; Santa Cruz, CA 95064, USA.

[60]Howard Hughes Medical Institute, University of California Santa Cruz; Santa Cruz, CA 95064, USA.

[61]Department of Evolution, Ecology and Organismal Biology, University of California Riverside; Riverside, CA 92521, USA.

[62]Department of Genetics, University of North Carolina Medical School; Chapel Hill, NC 27599, USA.

[63]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet; Stockholm, Sweden.

[64]Iris Data Solutions, LLC; Orono, ME 04473, USA.

[65]Museum of Zoology, Senckenberg Natural History Collections Dresden; 01109 Dresden, Germany.

[66]Allen Institute for Brain Science; Seattle, WA 98109, USA

[2]Department of Ecology and Evolutionary Biology, University of California Santa Cruz; Santa Cruz, CA 95064, USA.

[3]Howard Hughes Medical Institute, University of California Santa Cruz; Santa Cruz, CA 95064, USA.

[4]Broad Institute of MIT and Harvard; Cambridge, MA 02139, USA.

[5]Phillips Exeter Academy; Exeter, NH 03833, USA.

[6]Institute of Evolutionary Biology, Department of Experimental and Health Sciences, Universitat Pompeu Fabra; Barcelona, 08003, Spain.

[7]Smithsonian-Mason School of Conservation, George Mason University; Front Royal, VA 22630, USA.

[8]Center for Species Survival, Smithsonian Conservation Biology Institute, National Zoological Park; Washington, DC, 30008, USA.

[9]Computer Technologies Laboratory, ITMO University; St. Petersburg, 197101, Russia.

[10]Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School; Worcester, MA 01605, USA.

[11]Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University; Uppsala, 751 32, Sweden.

[12]Catalan Institution of Research and Advanced Studies; Barcelona, 08010, Spain.

[13]Centre for Genomic Regulation, Barcelona Institute of Science and Technology; Barcelona, 08028, Spain.

[14]Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona; Barcelona, 08193, Spain.

[15]European Molecular Biology Laboratory-European Bioinformatics Institute, Wellcome Genome Campus; Hinxton, UK.

[16]College of Law, University of Iowa; Iowa City, IA 52242, USA.

[17]Lehigh University, Biological Sciences; Bethlehem, PA 18015, USA.

[18]Department of Evolution, Behavior and Ecology, Division of Biology, University of California, San Diego; La Jolla, CA 92039 USA.

Abstract

Species persistence can be influenced by the amount, type, and distribution of diversity across the genome, suggesting a potential relationship between historical demography and resilience. Here, we surveyed genetic variation across single genomes of 240 mammals comprising the Zoonomia alignment to evaluate how historical effective population size (N_e) impacts heterozygosity and deleterious genetic load and how these factors may contribute to extinction risk. We find that species with smaller historical N_e carry a proportionally larger burden of deleterious alleles due to long-term accumulation and fixation of genetic load, and have higher risk of extinction. This suggests that historical demography can inform contemporary resilience. Models that included

genomic data were predictive of species' conservation status, suggesting that, in the absence of adequate census or ecological data, genomic information may provide an initial risk assessment.

One-Sentence Summary:

Genomic data from 240 species show that information encoded within a single genome can provide a conservation risk assessment.

The current rate of biodiversity loss amounts to a sixth mass extinction(1) and is compounded by substantial population declines across nearly one third of vertebrate species(2). Many species need immediate conservation intervention, a process that is especially challenging for the more than 20,000 species currently listed as "Data Deficient" by the International Union for Conservation of Nature (IUCN). Fortunately, genomic data, which are increasingly available for a broad taxonomic range of species, may hold promise for helping to identify at-risk species by providing readily accessible information on demography and fitness-relevant genetic variation(3, 4). It remains poorly explored, however, to what extent genomic data on their own are sufficient to help triage endangered species for conservation intervention.

Population genetic diversity and individual heterozygosity are long recognized correlates of fitness-relevant functional variation(5, 6). Our previous analysis of 124 placental mammalian genomes showed that lower heterozygosity and stretches of homozygosity are more common in species in threatened IUCN Red List categories(7). However, functional diversity, including estimates of adaptive variation and genetic load, may also be useful correlates of population resiliency. Such measures are increasingly accessible with emerging genomic tools(8) and comparative genomics resources such as the Zoonomia alignment of placental mammalian genomes (table S1)(7). The Zoonomia alignment provides high-resolution constraint scores and reconstructed ancestral sequences that can help to identify deleterious alleles at functionally important sites(7, 9).

Here, we surveyed the distribution of neutral and functional genetic variation across 240 species in the Zoonomia alignment to determine how historical effective population sizes (N_e) have influenced heterozygosity and deleterious genetic load (fig. S1). We test the value of genomic data to more precisely target species for conservation efforts by comparing the outcome of predictive models of conservation status that use ecological data, genomic data, or both. While we acknowledge the limitations of assuming that single genomes are representative of a species, our approach capitalizes on the unique resource provided by the Zoonomia consortium to explore whether genomic data can provide initial risk assessments that may be useful to triage data-deficient species and guide resource allocation for conservation intervention.

Historical population size is relevant to contemporary extinction risk

Species with historically small N_e tend to be classified in threatened IUCN Red List categories (Fig. 1). Species classified as Near Threatened (NT), Vulnerable (VU), Endangered (EN) or Critically Endangered (CR) had significantly smaller harmonic

mean N_e (mean$_{threatened}$=18,950) compared to non-threatened species (Least Concern (LC); mean$_{non-threatened}$=27,839; p<3.3e-5 when accounting for relationships across the phylogeny; Fig. 1B; figs. S2). N_e was also significantly smaller in threatened compared to non-threatened species within two of three taxonomic orders with sufficient numbers of species to test (Cetartiodactyla: mean$_{threatened}$=18,336, mean$_{non-threatened}$=22,648, p=0.023; and Carnivora: mean$_{threatened}$=9,636, mean$_{non-threatened}$=26,195, p=2.4e-5; but not Primates: mean$_{threatened}$=22,508, mean$_{nonthreatened}$=24,373, p=0.31; fig. S3). Within these two orders in particular, large-bodied herbivores and carnivores have declined in both geographic range and population size during the Anthropocene[10, 11]. Smaller populations are expected to have higher extinction risk, yet these historical N_e estimates reflect periods more than 10,000 years in the past, suggesting that long-term characteristics of ancestral populations can be informative about population size and extinction risk today. These results support the utility of metrics of genome-wide diversity in conservation assessments, a topic that is currently debated[12, 13].

Estimates of historical N_e can also identify previously large populations that have experienced contemporary declines. Specifically, if the estimate of historical N_e is large while N_c is small, this inflates the N_e/N_c ratio. In a study of pinnipeds, for example, most species that had undergone recent declines had smaller population census sizes (N_c) than expected based on their historical N_e [14]. To test this across the taxonomic range of the Zoonomia alignment, we examined the ratio of deep historical N_e to contemporary N_c for 89 species with population census information available in PanTHERIA[15]. Species in threatened IUCN categories had larger N_e/N_c ratios, i.e. smaller contemporary N_c relative to historical N_e (mean$_{threatened}$=1.07e-3; mean$_{non-threatened}$=4.29e-4; p=0.012; Fig. 1C). The relationship was also significant within Primates (phylolm, mean$_{threatened}$=3.46e-3; mean$_{non-threatened}$=1.11e-3; p=0.029), the only order with available N_e/N_c estimates and sufficient numbers of taxa in the two threat categories, indicating that the pattern holds among species with similar life-history traits. Across taxa, the largest N_e/N_c ratios included American bison (*Bison bison*), giant panda (*Ailuropoda melanoleuca*), and hirola (*Beatragus hunteri*), all of which have declined due to recent human activities[16–18].

Historically smaller populations carry proportionally larger burdens of genetic load

Historical N_e is correlated with the proportion of deleterious substitutions in mammalian genomes, reflecting the accumulation and fixation of genetic load over long evolutionary time periods. We called derived, single nucleotide substitutions for each species relative to the reconstructed sequence of the nearest ancestral phylogenetic node and called heterozygous sites from resequencing data mapped to the focal genome. We inferred the impacts of derived substitutions and heterozygous variants assuming that mutations at sites that are conserved across taxa (phyloP>2.27)[9] and nonsynonymous mutations are predominantly deleterious (fig. S1)[19]. Assuming most substitutions are fixed and mutation rates are similar across the phylogeny [20][21], the proportion of substitutions that are deleterious should be correlated with the total number of fixed deleterious mutations in the genome. Deleterious substitutions should therefore largely reflect fixed drift load that

reduces the mean fitness of the population, whereas heterozygous deleterious variants reflect segregating mutational load(22).

We found that species with smaller N_e had proportionally more substitutions at evolutionarily conserved sites genome-wide (phylolm, p=9.65e-3) and proportionally more missense substitutions in genes (phylolm, p=7.76e-5; fig. S4). Phylop kurtosis, which describes the extreme phyloP outliers in the tail of the distribution across substitutions, was positively correlated with N_e (phylolm, p=0.014). This means that species with smaller N_e had smaller right tails and therefore fewer substitutions at extremely conserved sites. To further parse potential fitness impacts of mutations in protein-coding regions, we examined genes with associated viability phenotypes in single-gene knockout mouse lines classified by the International Mouse Phenotyping Consortium (IMPC), assuming that, when aggregated across many genes, viability classifications are correlated to their fitness impacts in other species(23). Species with smaller N_e had proportionally more missense mutations relative to coding mutations in nearly all categories (phylolm, p<3.00e-5; Fig. 2; figs. S5–S6). We observed proportionally fewer missense mutations in IMPC lethal genes relative to IMPC viable genes (ANOVA, p<4.42e-9; fig. S7), reflecting stronger purifying selection in the lethal gene class, but the negative correlation was nonetheless consistent for both lethal and viable categories (Fig. 2). This relationship supports both theoretical predictions that smaller populations experiencing strong drift accumulate and fix weakly and moderately deleterious alleles (drift load)(12, 24) and empirical studies involving fewer or single taxa(25–27).

The correlations between N_e and conservation status and between N_e and drift load suggests that historical demography may influence contemporary extinction risk by shaping genome-wide diversity and genetic load. We found inconsistent relationships, however, between a species' proportional genetic load and its odds of being threatened. Species with proportionally more missense substitutions were more likely to be threatened when considering all genes (phyloglm, p=0.002; fig. S4D), as well as genes in lethal and viable IMPC categories (phyloglm, p<0.023; fig. S6), as observed in other taxa(28). Drift load estimated from evolutionary constraint across the genome, however, showed the opposite pattern: species with proportionally fewer substitutions at evolutionarily conserved sites were more likely to be threatened (phyloglm, p=1.38e-05; fig. S4C). This latter result contrasts with expectations, given that threatened species have smaller N_e on average (Fig 1) and smaller N_e is associated with proportionally more substitutions at conserved sites (phylolm, p=9.6e-3; fig. S4A). Interestingly, a previous study of 100 mammal genomes also found that threatened species had lower mean conservation scores across mutations(29). They suggested that the pattern may reflect fewer recessive deleterious alleles due to purging or the loss of these rare alleles to drift. The conflicting relationships between conservation status and metrics of drift load thus do not provide strong support for a mechanistic link between fixed drift load as measured in this study and species' resilience against extinction.

Genomic information can help predict extinction risk

Historical N_e was the most consistent genomic predictor of conservation status across regression models, while the predictive value of genetic load metrics varied with phylogenetic context (Fig. 3, tables S2–S3). Ordinal and logistic regression models

incorporating genomic variables with taxonomic order and dietary trophic level showed that the effect of N_e varied by ecological context. For example, an herbivore with a given N_e was more likely to be threatened than a carnivore or omnivore with the same N_e (Fig. 3B), supporting findings of elevated extinction risk in herbivores despite larger populations(30). Similarly, Carnivora and Primates both had increased risk with lower levels of severely deleterious genetic load. However, the specific metric of load that predicted conservation status differed among taxonomic orders, perhaps reflecting differences in natural history or ecological flexibility (figs. S8–S10). Principal components (PC) regression of demographic and genetic load variables showed that, overall, threatened species tended to have proportionally more deleterious mutations in coding regions, lower heterozygosity, and smaller N_e (PC1; p=0.0038), as well as proportionally more missense substitutions (PC3; p=5.6e-4; Fig. 3A, table S3). Although no single genomic variable unambiguously discriminated threatened from non-threatened species (fig. S2), many have predictive value, which will be particularly relevant for species lacking adequate ecological or census data.

Although ecological data were more powerful than genomic data to predict extinction risk in our predictive models, models using only information from single genomes nonetheless identified species at risk of being threatened. We generated random forest models to predict conservation status from ecological traits(31, 32) and genomic features, using area under the receiver operating characteristic (AUROC) to evaluate performance. A model with AUROC of 0.5 has no predictive ability, whereas a model with AUROC of 1.0 has perfect predictive performance. We selected predictive variables from among 13 genome-wide summary statistics including demographic history, genetic diversity, and genetic load variables, ~57,000 window-based metrics per genome, and 39 ecological variables from PanTHERIA(15) including physiological, life-history, and behavioral variables (table S4). Models including only genomic features and no ecological variables (17 models; AUROC ranged from 0.69–0.82) performed worse than models including only ecological variables (1 model; AUROC 0.88) and similarly to models including both genomic and ecological variables (17 models; AUROC range 0.68–0.83; table S5). Models with only genomic features were, however, consistently better able to distinguish threatened from non-threatened species (tables S5–S6; fig. S11–13) compared to random chance (i.e. AUROC of 0.5). Models including only genomic variables performed similarly to other studies that predicted IUCN status from ecological or morphological data with comparable sample sizes (e.g. AUC ranging from 0.67–0.90 for n=171–430 species) (33–35).

The number of species with values for ecological, genome-wide summary statistics, and window-based metrics differed, which may affect model performance. To compare the predictive value of genomic and ecological features directly, we next tested models in a set of 210 species for which both data types were available (tables S4 and S6). Again, the model with genome-wide summary statistics alone was predictive of threatened status (AUROC 0.71), but performed more poorly than the model with ecological variables (AUROC 0.83). Combining genomic summary statistics with ecological variables led to a modest improvement in distinguishing threatened from non-threatened species (AUROC=0.85) compared to genomic variables alone, with N_e as the fourth most important predictor in the model after weaning age, age at first birth, and age of sexual maturity (fig. S14). Models including genomic window-based features never outperformed models with

ecological variables alone (table S6), suggesting that complementary information provided by genomic versus ecological data may be better captured by summary or transformed variables (e.g. principal components) than by numerous weakly informative window features that may overwhelm the predictive models. Overall, our evaluation suggests that while genomic information from a single individual is not better than ecological data for predicting threatened status, these data do have predictive value, especially when ecological variables are unavailable.

As a demonstration of their utility, we applied our regression and random forest models to predict the status of three species considered "Data Deficient" by the IUCN (Fig. 3D). The models suggest the Upper Galilee Mountains blind mole rat (*Nannospalax galili*), which lacks ecological data, is least likely to be threatened (11–44% probability), whereas the killer whale (*Orcinus orca*), for which both ecological and genomic data are available, is more likely to be threatened (35–68% probability), consistent with the identification of some at-risk populations(36). Predictions for the Java lesser chevrotain (*Tragulus javanicus*) depend on model specifications, with the highest threat prediction from the within-order regression model (67% probability), and other models suggesting it is less likely to be threatened (24–49% probability). The results indicate that, among the three species, the killer whale should be prioritized for further study, and demonstrate how genomic data can provide a rapid and inexpensive initial conservation assessment.

Discussion

Our results provide empirical support for theoretical predictions that small populations accumulate and fix weakly and moderately deleterious alleles, and demonstrate a correlation between historical effective population size and contemporary extinction risk. We found little evidence, however, that species with historically small effective population sizes have higher risks of extinction because of elevated drift load. Alternatively, historically small populations may have elevated extinction risk simply because these populations are small and thus more vulnerable to other threats such as habitat loss or change, the introduction of infectious disease, competition with invasive species, and new hunting or predation pressures.

Despite the limitations of assuming that a single genome is representative of the diversity within a species, our comparative genomics approach allowed us to maximize the number of species analyzed to explore the power to detect genomic correlates of endangerment. Empirical studies suggest a single individual can represent a species for characteristics shaped by long-term evolutionary history; variation in the proportion of deleterious mutations is typically smaller within species than between(37, 29), and historical N_e estimates are consistent across conspecifics(38, 39). The analysis of multiple resequenced individuals per species, however, will increase accuracy and resolution by capturing intraspecific variation in genetic diversity, heterozygosity, and inbreeding (especially for species with strong population structure), enabling estimation of allele frequencies, improving inference of more recent demographic history, and allowing better detection of rare and segregating variants(e.g. inbreeding load; *22*). The latter may be particularly important for estimating extinction risk, as segregating variants tend to be enriched

for deleterious alleles(40, 41) and may disproportionately impact extinction risk from population bottlenecks(12). In the future, larger data sets comprising multiple individuals per species may shed light on long-standing questions about the relative impact on fitness of many weakly deleterious alleles versus a few strongly deleterious alleles(22, 25, 37, 42, 43).

Inferring real-world fitness from genomic data includes caveats. Evolutionary constraint may, for example, reflect past selection on loci that no longer impact fitness(44). Loci that seem functionally important in model species may be irrelevant to the species of interest, compensatory mutations may ameliorate the impact of deleterious mutations, and factors such as dominance, epistasis, pleiotropy, and purging may also complicate the relationship between genetic load and fitness. Finally, local differences in habitat may mean that the impact of deleterious mutations differs among individuals or populations(25, 45, 46). For these reasons, the impact of the observed proportionally higher load in smaller populations will be challenging to know in the absence of direct fitness data, such as reproductive success and the frequencies of genetic diseases and congenital abnormalities(26, 43, 47).

As additional genomes and population resequencing data become available(48), the power and accuracy of predictions of extinction risk from genomes will improve(8). Our analyses of the genomes of single individuals, which can be generated rapidly and inexpensively(49), demonstrate the potential for using genomic estimates of demography, diversity, and genetic load to triage species in need of immediate management intervention, and we join in the calls for including genomics into conservation status assessments(50–53).

Materials and Methods

We provide a summary of our materials and methods below; refer to the Supplemental Materials and Methods for further detail.

Mammal genomes and metadata

We examined genomic variation in 240 species represented by 241 reference genomes in the Zoonomia multispecies alignment. The genome assemblies varied in quality, with contig N50 values ranging from 1 KB to 56 MB (table S1). Short-read sequence data, usually from the reference individual, were used to estimate metrics related to historical demography, heterozygosity, and heterozygous deleterious variants from single genomes. Homozygous deleterious genetic load was estimated relative to reconstructed ancestral sequences from the multispecies alignment (fig. S1). We tested correlations between all genomic metrics, and between genomic metrics and extinction risk, using a statistical framework that accounts for phylogenetic relationships across species. Using regression and machine learning models, we tested the potential for genomic data to predict the conservation status of species.

For all species, we compiled metadata on conservation status, diet, and generation time (table S1). We assigned a conservation status (Least Concern (LC), Near Threatened (NT), Vulnerable (VU), Endangered (EN) or Critically Endangered (CR)) to the lowest known taxonomic level of the sequenced sample, using the IUCN Red List of Threatened Species (IUCN Red List API v. 3) as a proxy for extinction risk. We classified each species as carnivore, herbivore, or omnivore based on(54), using information for the genus when

species-specific information was unavailable. From available metadata, we categorized the sample used for both the reference genome and short-read data as a wild, captive, or domesticated individual.

Tests for correlations between variables were conducted with phylogenetic linear regression or phylogenetic logistic regression in the R package *phylolm(55)*, incorporating the phylogenetic tree with branch lengths(56) to account for non-independence.

Estimating historical effective population sizes and genome-wide heterozygosity

We called heterozygous positions in all genomes with short-read data using the GATK best practices pipeline as described previously(7). Briefly, we mapped paired-end sequencing data to the respective genome assemblies using BWA mem (version 0.7.15)(57), marked and removed optical duplicates, and called heterozygous variants using the HaplotypeCaller module of the GATK software suite (version 3.6)(58).

We inferred the history of effective population sizes (N_e) for each species using PSMC (version 0.6.5-r67)(59). We called variants in each genome from scaffolds >50KB in length, filtered for sequence read coverage and base quality score, and used these as input for PSMC. We rescaled the PSMC output using species-specific generation times(60) and a mammalian mutation rate(21) and calculated the harmonic mean across temporal estimates from periods >10 kya. To compare contemporary population sizes to historical N_e, we obtained census population estimates (N_c) for 89 species from the PanTHERIA database(15), estimating N_c as the product of population density and geographic area from census data(15, 61).

To identify runs of homozygosity (RoH), we used our previously described method(7). For every assembly, we calculated the ratio of heterozygous to callable positions in non-overlapping, 50-kb windows, and fit a 2-component Gaussian Mixture Model to the joint distribution, which is expected to be bimodal with a peak at the lower tail of the distribution corresponding to runs of homozygosity (fig. S1B). Windows were then assigned as RoH or non-RoH and used to calculate the proportion of the genome in RoH (fRoH), genome-wide heterozygosity, and outbred heterozygosity (i.e. heterozygosity in non-RoH regions; figs. S2 and S15).

Deleterious genetic load

We called heterozygous variants from single sample, short-read data mapped to the reference genome of each species. Homozygous substitutions were estimated from each reference genome relative to the closest reconstructed ancestral sequence in the phylogeny using the *halBranchMutations* tool in the Comparative Genomics Toolkit(62). Because new alleles become fixed or lost on the order of $<4N_e$ generations(63), most homozygous substitutions between species are likely fixed. We assessed the potential functional impact of mutations by 1) evolutionary conservation of the site (phyloP), and 2) the estimated impact of the mutation on protein-coding genes. Mutations at evolutionarily conserved sites (phyloP>2.27; (9)), and those that cause nonsynonymous changes in protein-coding genes, were assumed to be predominantly harmful(19). Variant sites in each genome were assigned human-based phyloP scores estimated from the multispecies alignment(9). To infer functional impacts

on protein-coding genes, each genome was annotated with human orthologs by lifting over human exon intervals to the target species. Synonymous, missense and loss-of-function variants were then estimated in the program *SnpEff* v.5.0e(64). We also examined mutations in single-copy genes with associated viability phenotypic data in knockout mice as classified by the International Mouse Phenotyping Consortium (IMPC)(23), using IMPC categories (e.g. lethal or viable) as a proxies for gene essentiality and the potential fitness impacts of mutations in these genes(23).

Predicting threat from genomic variables

To predict whether a species is threatened (NT, VU, EN, and CR categories) or non-threatened (LC category), we modeled conservation status across species from genomic variables using both regression and machine learning models.

We took two main approaches in our regression models of conservation status across species, using 1) phylogenetic logistic regression to model threatened versus non-threatened status, which allowed us to test the significance of predictor variables, but not make predictions for species with unknown threat status, and 2) ordinal regression models of specific IUCN categories, which allowed us to test significance and make predictions for species with unknown threat status. Unlike logistic regression, ordinal regression did not inherently incorporate the phylogeny, so we included taxonomic order as a factor in the models. We tested 13 genomic variables (table S2), modeled individually and as principal components, and included taxonomic order and dietary trophic level, a previously described correlate of extinction risk(65). We estimated model error by fitting parameters on 80% of the data and testing the remaining 20% of the data across 100 runs with different data subsets.

We used random-forest based classification to estimate the likelihood that a species is threatened from 13 genome-wide summary statistics of heterozygosity, demographic history, and genetic load, and from 5 genomic metrics within homologous 50KB windows (table S4). We trained models using the two genomic data types (windows-based and genome-wide) separately and combined, and incorporated 39 ecological variables from the PanTHERIA database (table S4). We used the scikit-learn 1.0.2 package for fitting all the models(66).

We first split our dataset into a 75% training set and a 25% test set. For each model, we performed preprocessing and imputation steps using only the training data, then trained the model on the training set and evaluated it on the test set. We ran 5-fold cross validation on the training set to determine the optimal set of hyperparameters, tuning the number of decision trees, the maximum depth of the trees, and the number of features used at each decision to optimize a performance metric. We used AUROC to estimate how well a model predicts the correct output class. AUROC is designed to be more robust to class imbalance in comparison to a metric such as accuracy.

To leverage all available data, we first ran models using all species with data for a given data type (table S5). The number of species with values for ecological, genome-wide summary statistics, and window-based metrics differed however, which may impact the results. To

compare the performance of ecological and genomic variables and their combination across the same set of species, we also trained and tested models in the set of species for which both data types were available (table S6).

The Zoonomia alignment included three species classified as "Data Deficient" by the IUCN, the Upper Galilee Mountains blind mole rat (*Nannospalax galili*), the Java lesser chevrotain (*Tragulus javanicus*), and the killer whale (*Orcinus orca*). The blind mole rat lacked ecological data on PanTHERIA. We used the within-order and across-order ordinal regression models and all random forest models to predict the probability that these species are threatened.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

Data and materials availability:

The data presented in this paper are detailed in supplementary materials. Summary data and analysis scripts are available at https://github.com/LaMariposa/zoonomia_biodiversity. NCBI accession numbers for sequence data used in analyses are given in table S1.

References and Notes

1. Barnosky AD, Matzke N, Tomiya S, Wogan GOU, Swartz B, Quental TB, Marshall C, McGuire JL, Lindsey EL, Maguire KC, Mersey B, Ferrer EA, Has the Earth's sixth mass extinction already arrived? Nature. 471, 51–57 (2011). [PubMed: 21368823]

2. Ceballos G, Ehrlich AH, Ehrlich PR, The Annihilation of Nature: Human Extinction of Birds and Mammals (JHU Press, 2015).

3. Supple MA, Shapiro B, Conservation of biodiversity in the genomics era. Genome Biol. 19, 131 (2018). [PubMed: 30205843]

4. Hansson B, Morales HE, van Oosterhout C, Comment on "Individual heterozygosity predicts translocation success in threatened desert tortoises." Science. 372 (2021), p. eabh1105. [PubMed: 34083458]

5. Hansson B, Westerberg L, On the correlation between heterozygosity and fitness in natural populations. Molecular Ecology. 11 (2002), pp. 2467–2474. [PubMed: 12453232]

6. DeWoody JA, Harder AM, Mathur S, Willoughby JR, The long-standing significance of genetic diversity in conservation. Mol. Ecol. 30 (2021), pp. 4147–4154. [PubMed: 34191374]

7. Zoonomia Consortium A comparative genomics multitool for scientific discovery and conservation. Nature. 587, 240–245 (2020). [PubMed: 33177664]

8. van Oosterhout C, Mutation load is the spectre of species conservation. Nat Ecol Evol. 4, 1004–1006 (2020). [PubMed: 32367032]

9. Christmas MJ, Kaplow IM, Zoonomia consortium, Evolutionary constraint and innovation across hundreds of placental mammals. Science.

10. Ripple WJ, Estes JA, Beschta RL, Wilmers CC, Ritchie EG, Hebblewhite M, Berger J, Elmhagen B, Letnic M, Nelson MP, Schmitz OJ, Smith DW, Wallach AD, Wirsing AJ, Status and Ecological Effects of the World's Largest Carnivores. Science. 343 (2014), doi:10.1126/science.1241484.

11. Ripple WJ, Newsome TM, Wolf C, Dirzo R, Everatt KT, Galetti M, Hayward MW, Kerley GIH, Levi T, Lindsey PA, Macdonald DW, Malhi Y, Painter LE, Sandom CJ, Terborgh J, Van Valkenburgh B, Collapse of the world's largest herbivores. Science Advances. 1 (2015), doi:10.1126/sciadv.1400103.

12. Kardos M, Armstrong EE, Fitzpatrick SW, Hauser S, Hedrick PW, Miller JM, Tallmon DA, Funk WC, The crucial role of genome-wide genetic variation in conservation. Proc. Natl. Acad. Sci. U. S. A. 118 (2021), doi:10.1073/pnas.2104642118.

13. Teixeira JC, Huber CD, The inflated significance of neutral genetic diversity in conservation genetics. Proc. Natl. Acad. Sci. U. S. A. 118 (2021), doi:10.1073/pnas.2015096118.

14. Peart CR, Tusso S, Pophaly SD, Botero-Castro F, Wu C-C, Aurioles-Gamboa D, Baird AB, Bickham JW, Forcada J, Galimberti F, Gemmell NJ, Hoffman JI, Kovacs KM, Kunnasranta M, Lydersen C, Nyman T, de Oliveira LR, Orr AJ, Sanvito S, Valtonen M, Shafer ABA, Wolf JBW, Determinants of genetic variation across eco-evolutionary scales in pinnipeds. Nat Ecol Evol. 4, 1095–1104 (2020). [PubMed: 32514167]

15. Jones KE, Bielby J, Cardillo M, Fritz SA, O'Dell J, Orme CDL, Safi K, Sechrest W, Boakes EH, Carbone C, Connolly C, Cutts MJ, Foster JK, Grenyer R, Habib M, Plaster CA, Price SA, Rigby EA, Rist J, Teacher A, Bininda-Emonds ORP, Gittleman JL, Mace GM, Purvis A, PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. Ecology. 90 (2009), pp. 2648–2648.

16. IUCN SSC Antelope Specialist Group, "Beatragus hunteri: IUCN SSC Antelope Specialist Group" (2017), doi:10.2305/IUCN.UK.2017-2.RLTS.T6234A50185297.en.

17. Zhao S, Zheng P, Dong S, Zhan X, Wu Q, Guo X, Hu Y, He W, Zhang S, Fan W, Zhu L, Li D, Zhang X, Chen Q, Zhang H, Zhang Z, Jin X, Zhang J, Yang H, Wang J, Wang J, Wei F, Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. Nat. Genet. 45, 67–71 (2013). [PubMed: 23242367]

18. Hedrick PW, Conservation genetics and North American bison (Bison bison). J. Hered. 100, 411–420 (2009). [PubMed: 19414501]

19. Henn BM, Botigué LR, Bustamante CD, Clark AG, Gravel S, Estimating the mutation load in human genomes. Nat. Rev. Genet. 16, 333–343 (2015). [PubMed: 25963372]

20. Kimura M, Evolutionary Rate at the Molecular Level. Nature. 217 (1968), pp. 624–626. [PubMed: 5637732]

21. Kumar S, Subramanian S, Mutation rates in mammalian genomes. Proc. Natl. Acad. Sci. U. S. A. 99, 803–808 (2002). [PubMed: 11792858]

22. Hedrick PW, Garcia-Dorado A, Understanding Inbreeding Depression, Purging, and Genetic Rescue. Trends Ecol. Evol. 31, 940–952 (2016). [PubMed: 27743611]

23. Muñoz-Fuentes V, Cacheiro P, Meehan TF, Aguilar-Pimentel JA, Brown SDM, Flenniken AM, Flicek P, Galli A, Mashhadi HH, Hrab de Angelis M, Kim JK, Lloyd KCK, McKerlie C, Morgan H, Murray SA, Nutter LMJ, Reilly PT, Seavitt JR, Seong JK, Simon M, Wardle-Jones H, Mallon A-M, Smedley D, Parkinson HE, IMPC consortium, The International Mouse Phenotyping Consortium (IMPC): a functional catalogue of the mammalian genome that informs conservation. Conserv. Genet. 19, 995–1005 (2018). [PubMed: 30100824]

24. Kimura M, Maruyama T, Crow JF, THE MUTATION LOAD IN SMALL POPULATIONS. Genetics. 48 (1963), pp. 1303–1312. [PubMed: 14071753]

25. Grossen C, Guillaume F, Keller LF, Croll D, Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. Nat. Commun. 11, 1–12 (2020). [PubMed: 31911652]

26. Robinson JA, Räikkönen J, Vucetich LM, Vucetich JA, Peterson RO, Lohmueller KE, Wayne RK, Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. Sci Adv. 5, eaau0757 (2019). [PubMed: 31149628]

27. Yoshida K, Ravinet M, Makino T, Toyoda A, Kokita T, Mori S, Kitano J, Accumulation of Deleterious Mutations in Landlocked Threespine Stickleback Populations. Genome Biol. Evol. 12, 479–492 (2020). [PubMed: 32232440]

28. Rolland J, Schluter D, Romiguier J, Vulnerability to Fishing and Life History Traits Correlate with the Load of Deleterious Mutations in Teleosts. Mol. Biol. Evol. 37, 2192–2196 (2020). [PubMed: 32163146]

29. van der Valk T, de Manuel M, Marques-Bonet T, Guschanski K, Estimates of genetic load suggest frequent purging of deleterious alleles in small populations. bioRxiv (2021), doi:10.1101/696831.

30. Atwood TB, Valentine SA, Hammill E, McCauley DJ, Madin EMP, Beard KH, Pearse WD, Herbivores at the highest risk of extinction among mammals, birds, and reptiles. Science Advances. 6 (2020), p. eabb8458. [PubMed: 32923612]

31. Bland LM, Collen B, Orme CDL, Bielby J, Predicting the conservation status of data-deficient species. Conserv. Biol. 29, 250–259 (2015). [PubMed: 25124400]

32. Davidson AD, Hamilton MJ, Boyer AG, Brown JH, Ceballos G, Multiple ecological pathways to extinction in mammals. Proc. Natl. Acad. Sci. U. S. A. 106, 10702–10705 (2009). [PubMed: 19528635]

33. Walls RHL, Dulvy NK, Eliminating the dark matter of data deficiency by predicting the conservation status of Northeast Atlantic and Mediterranean Sea sharks and rays. Biological Conservation. 246 (2020), p. 108459.

34. Miles DB, Can Morphology Predict the Conservation Status of Iguanian Lizards? Integr. Comp. Biol. 60, 535–548 (2020). [PubMed: 32559284]

35. Kopf RK, Shaw C, Humphries P, Trait-based prediction of extinction risk of small-bodied freshwater fishes. Conserv. Biol. 31, 581–591 (2017). [PubMed: 27976421]

36. Jourdain E, Ugarte F, Víkingsson GA, Samarra FIP, Ferguson SH, Lawson J, Vongraven D, Desportes G, North Atlantic killer whale Orcinus orca populations: a review of current knowledge and threats to conservation. Mammal Review. 49 (2019), pp. 384–400.

37. Robinson JA, Kyriazis CC, Nigenda-Morales SF, Beichman AC, Rojas-Bracho L, Robertson KM, Fontaine MC, Wayne RK, Lohmueller KE, Taylor BL, Morin PA, The critically endangered vaquita is not doomed to extinction by inbreeding depression. Science. 376, 635–639 (2022). [PubMed: 35511971]

38. Saremi NF, Supple MA, Byrne A, Cahill JA, Coutinho LL, Dalén L, Figueiró HV, Johnson WE, Milne HJ, O'Brien SJ, O'Connell B, Onorato DP, Riley SPD, Sikich JA, Stahler DR, Villela PMS, Vollmers C, Wayne RK, Eizirik E, Corbett-Detig RB, Green RE, Wilmers CC, Shapiro B, Puma genomes from North and South America provide insights into the genomic consequences of inbreeding. Nat. Commun. 10, 4769 (2019). [PubMed: 31628318]

39. Meyer WK, Venkat A, Kermany AR, van de Geijn B, Zhang S, Przeworski M, Evolutionary history inferred from the de novo assembly of a non-model organism, the blue-eyed black lemur. Mol. Ecol. 24, 4392–4405 (2015). [PubMed: 26198179]

40. Bertorelle G, Raffini F, Bosse M, Bortoluzzi C, Iannucci A, Trucchi E, Morales HE, van Oosterhout C, Genetic load: genomic estimates and applications in non-model animals. Nat. Rev. Genet. 23, 492–503 (2022). [PubMed: 35136196]

41. Wolf JBW, Künstner A, Nam K, Jakobsson M, Ellegren H, Nonlinear dynamics of nonsynonymous (dN) and synonymous (dS) substitution rates affects inference of selection. Genome Biol. Evol. 1, 308–319 (2009). [PubMed: 20333200]

42. Khan A, Patel K, Shukla H, Viswanathan A, van der Valk T, Borthakur U, Nigam P, Zachariah A, Jhala YV, Kardos M, Ramakrishnan U, Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. Proc. Natl. Acad. Sci. U. S. A. 118 (2021), doi:10.1073/pnas.2023018118.

43. Smeds L, Ellegren H, From high masked to high realized genetic load in inbred Scandinavian wolves. Mol. Ecol. (2022), doi:10.1111/mec.16802.

44. Huber CD, Kim BY, Lohmueller KE, Population genetic models of GERP scores suggest pervasive turnover of constrained sites across mammalian evolution. PLoS Genet. 16, e1008827 (2020). [PubMed: 32469868]

45. Mee JA, Yeaman S, Unpacking Conditional Neutrality: Genomic Signatures of Selection on Conditionally Beneficial and Conditionally Deleterious Mutations. Am. Nat. 194, 529–540 (2019). [PubMed: 31490722]

46. Zhang Y, Stern AJ, Nielsen R, Evolution of the genetic architecture of local adaptations under genetic rescue is determined by mutational load and polygenicity, doi:10.1101/2020.11.09.374413.

47. Robinson JA, Brown C, Kim BY, Lohmueller KE, Wayne RK, Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. Curr. Biol. 28, 3487–3494.e4 (2018). [PubMed: 30415705]

48. Shaffer HB, Toffelmier E, California Conservation Genomics Project First Year Annual Report (2020) (available at https://escholarship.org/content/qt2sc7s29z/qt2sc7s29z.pdf).

49. Dudchenko O, Shamim MS, Batra SS, Durand NC, Musial NT, Mostofa R, Pham M, St Hilaire BG, Yao W, Stamenova E, Hoeger M, Nyquist SK, Korchina V, Pletch K, Flanagan JP, Tomaszewicz A, McAloose D, Estrada CP, Novak BJ, Omer AD, Aiden EL, The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000, doi:10.1101/254797.

50. Allendorf FW, Hohenlohe PA, Luikart G, Genomics and the future of conservation genetics. Nat. Rev. Genet. 11, 697–709 (2010). [PubMed: 20847747]

51. Mcmahon BJ, Teeling EC, Höglund J, How and why should we implement genomics into conservation? Evol. Appl. 7, 999–1007 (2014). [PubMed: 25553063]

52. Brandies P, Peel E, Hogg CJ, Belov K, The Value of Reference Genomes in the Conservation of Threatened Species. Genes. 10 (2019), doi:10.3390/genes10110846.

53. van Oosterhout C, Speak SA, Birley T, Bortoluzzi C, Percival-Alwyn L, Urban LH, Groombridge JJ, Segelbacher G, Morales HE, Genomic erosion in the assessment of species extinction risk and recovery potential, doi:10.1101/2022.09.13.507768.

54. Nowak RM, Walker EP, Walker's Mammals of the World (JHU Press, 1999).

55. si T L Ho C. Ané, A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. Syst. Biol. 63, 397–408 (2014). [PubMed: 24500037]

56. Foley NM, Mason VC, Harris AJ, Bredemeyer KR, Damas J, Lewin HA, Eizirik E, Gates J, Zoonomia Consortium MS Springer WJ. Murphy, A genomic timescale for placental mammal evolution. Science.

57. Li H, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM (2013), (available at http://arxiv.org/abs/1303.3997).

58. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010). [PubMed: 20644199]

59. Li H, Durbin R, Inference of human population history from individual whole-genome sequences. Nature. 475, 493–496 (2011). [PubMed: 21753753]

60. Pacifici M, Santini L, Marco MD, Baisero D, Francucci L, Marasini GG, Visconti P, Rondinini C, Generation length for mammals. Nature Conservation. 5, 89–94 (2013).

61. Roddy AB, Alvarez-Ponce D, Roy SW, Mammals with Small Populations Do Not Exhibit Larger Genomes. Molecular Biology and Evolution. 38 (2021), pp. 3737–3741. [PubMed: 33956142]

62. Hickey G, Paten B, Earl D, Zerbino D, Haussler D, HAL: a hierarchical format for storing and analyzing multiple genome alignments. Bioinformatics. 29 (2013), pp. 1341–1342. [PubMed: 23505295]

63. Otto SP, Whitlock MC, Fixation Probabilities and Times. Encyclopedia of Life Sciences (2006), doi:10.1038/npg.els.0005464.

64. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly. 6 (2012), pp. 80–92. [PubMed: 22728672]

65. Purvis A, Gittleman JL, Cowlishaw G, Mace GM, Predicting extinction risk in declining species. Proc. Biol. Sci. 267, 1947–1952 (2000). [PubMed: 11075706]

66. Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, Gramfort A, Thirion B, Varoquaux G, Machine learning for neuroimaging with scikit-learn. Front. Neuroinform. 8, 14 (2014). [PubMed: 24600388]

67. Martin M, Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 17 (2011), p. 10.

68. Institute Broad, Toolkit Picard (2019), (available at http://broadinstitute.github.io/picard/).

69. Gower G, Tuke J, Rohrlach AB, Soubrier J, Llamas B, Bean N, Cooper A, Population size history from short genomic scaffolds: how short is too short? bioRxiv (2018), doi:10.1101/382036.

70. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25, 2078–2079 (2009). [PubMed: 19505943]

71. Li H, Handsaker B, Danecek P, McCarthy S, Marshall J, BCFtools.

72. Li H, Implementation of the Pairwise Sequentially Markovian Coalescent (PSMC) model (2016), (available at https://github.com/lh3/psmc).

73. Freedman AH, Gronau I, Schweizer RM, Ortega-Del Vecchyo D, Han E, Silva PM, Galaverni M, Fan Z, Marx P, Lorente-Galdos B, Beale H, Ramirez O, Hormozdiari F, Alkan C, Vilà C, Squire K, Geffen E, Kusak J, Boyko AR, Parker HG, Lee C, Tadigotla V, Wilton A, Siepel A, Bustamante CD, Harkins TT, Nelson SF, Ostrander EA, Marques-Bonet T, Wayne RK, Novembre J, Genome sequencing highlights the dynamic early history of dogs. PLoS Genet. 10, e1004016 (2014). [PubMed: 24453982]

74. Patton AH, Margres MJ, Stahlke AR, Hendricks S, Lewallen K, Hamede RK, Ruiz-Aravena M, Ryder O, McCallum HI, Jones ME, Hohenlohe PA, Storfer A, Contemporary Demographic Reconstruction Methods Are Robust to Genome Assembly Quality: A Case Study in Tasmanian Devils. Mol. Biol. Evol. 36, 2906–2921 (2019). [PubMed: 31424552]

75. Hawkins MTR, Culligan RR, Frasier CL, Dikow RB, Hagenson R, Lei R, Louis EE Jr, Genome sequence and population declines in the critically endangered greater bamboo lemur (Prolemur simus) and implications for conservation. BMC Genomics. 19, 445 (2018). [PubMed: 29884119]

76. Frankham R, Effective population size/adult population size ratios in wildlife: a review. Genet. Res. 89, 491–503 (2007). [PubMed: 18976539]

77. Schreiber J, Pomegranate: fast and flexible probabilistic modeling in python. J. Mach. Learn. Res. 18, 5992–5997 (2018).

78. Armstrong J, Hickey G, Diekhans M, Fiddes IT, Novak AM, Deran A, Fang Q, Xie D, Feng S, Stiller J, Genereux D, Johnson J, Marinescu VD, Alföldi J, Harris RS, Lindblad-Toh K, Haussler D, Karlsson E, Jarvis ED, Zhang G, Paten B, Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature. 587, 246–251 (2020). [PubMed: 33177663]

79. Siepel A, Pollard KS, Haussler D, New Methods for Detecting Lineage-Specific Selection. Lecture Notes in Computer Science (2006), pp. 190–205.

80. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, Ward LD, Lowe CB, Holloway AK, Clamp M, Gnerre S, Alföldi J, Beal K, Chang J, Clawson H, Cuff J, Di Palma F, Fitzgerald S, Flicek P, Guttman M, Hubisz MJ, Jaffe

DB, Jungreis I, Kent WJ, Kostka D, Lara M, Martins AL, Massingham T, Moltke I, Raney BJ, Rasmussen MD, Robinson J, Stark A, Vilella AJ, Wen J, Xie X, Zody MC, Institute Sequencing Platform and Whole Genome Assembly Team Broad, Baldwin J, Bloom T, Chin CW, Heiman D, Nicol R, Nusbaum C, Young S, Wilkinson J, Worley KC, Kovar CL, Muzny DM, Gibbs RA, College of Medicine Human Genome Sequencing Center Sequencing Team Baylor, Cree A, Dihn HH, Fowler G, Jhangiani S, Joshi V, Lee S, Lewis LR, Nazareth LV, Okwuonu G, Santibanez J, Warren WC, Mardis ER, Weinstock GM, Wilson RK, Institute at Washington University Genome, Delehaunty K, Dooling D, Fronik C, Fulton L, Fulton B, Graves T, Minx P, Sodergren E, Birney E, Margulies EH, Herrero J, Green ED, Haussler D, Siepel A, Goldman N, Pollard KS, Pedersen JS, Lander ES, Kellis M, A high-resolution map of human evolutionary constraint using 29 mammals. Nature. 478, 476–482 (2011). [PubMed: 21993624]

81. Christmas MJ, Kaplow IM, Zoonomia consortium, Evolutionary constraint and innovation across hundreds of placental mammals. Science.

82. Exonerate, (available at https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate-manual).

83. Robinson JA, Kyriazis CC, Nigenda-Morales SF, Beichman AC, Rojas-Bracho L, Robertson KM, Fontaine MC, Wayne RK, Lohmueller KE, Taylor BL, Morin PA, The critically endangered vaquita is not doomed to extinction by inbreeding depression. Science. 376 (2022), pp. 635–639. [PubMed: 35511971]

84. Sánchez-Barreiro F, Gopalakrishnan S, Ramos-Madrigal J, Westbury MV, de Manuel M, Margaryan A, Ciucani MM, Vieira FG, Patramanis Y, Kalthoff DC, Timmons Z, Sicheritz-Pontén T, Dalén L, Ryder OA, Zhang G, Marquès-Bonet T, Moodley Y, Gilbert MTP, Historical population declines prompted significant genomic erosion in the northern and southern white rhinoceros (Ceratotherium simum). Mol. Ecol. 30, 6355–6369 (2021). [PubMed: 34176179]

85. Venables WN, Ripley BD, Modern Applied Statistics with S (Springer, Fourth, 2002).

86. R Core Team R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2015).

87. Foote AD, Liu Y, Thomas GWC, Vina T, Alföldi J, Deng J, Dugan S, van Elk CE, Hunter ME, Joshi V, Khan Z, Kovar C, Lee SL, Lindblad-Toh K, Mancia A, Nielsen R, Qin X, Qu J, Raney BJ, Vijay N, Wolf JBW, Hahn MW, Muzny DM, Worley KC, Gilbert MTP, Gibbs RA, Convergent evolution of the genomes of marine mammals. Nat. Genet. 47, 272–275 (2015). [PubMed: 25621460]
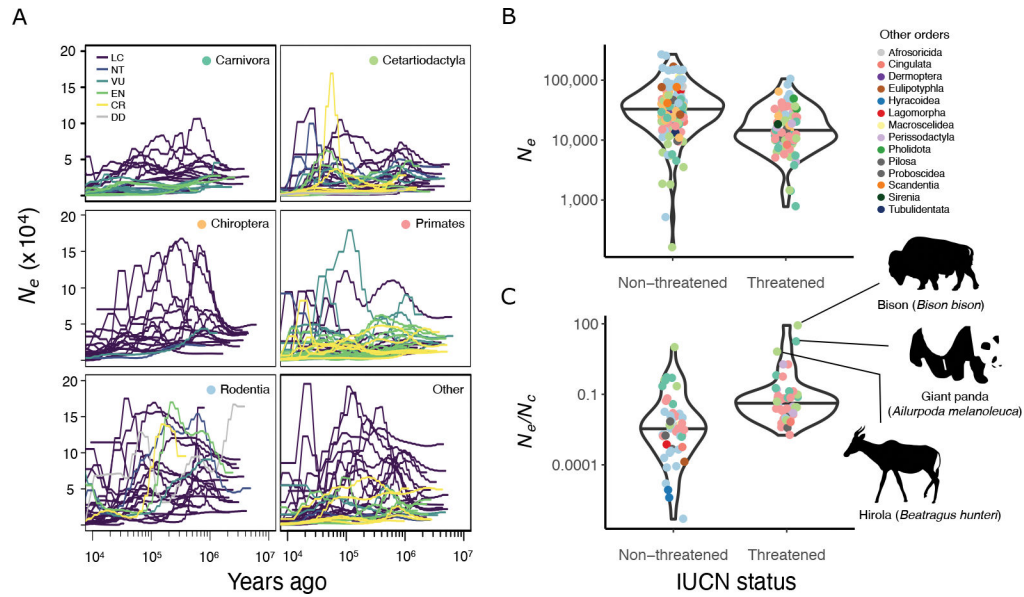
Fig. 1. Demographic history across mammalian orders and IUCN Red List categories.
(**A**) Estimates of effective population sizes (N_e) over time displayed by taxonomic order. Lines represent individual species, colored by IUCN status (LC= Least Concern, NT=Near Threatened, VU=Vulnerable, EN=Endangered, CR=Critically Endangered, DD=Data Deficient). Colored dots correspond to the taxonomic order of species depicted in (**B**) and (**C**). For visualization, only species with N_e estimates under 200,000 for every time point are shown. (**B**) Harmonic mean N_e was significantly lower in threatened IUCN categories relative to non-threatened (phylolm, p<3.3e-5). (**C**) The ratio of historical N_e to contemporary census population size (N_e/N_c) can identify species with smaller N_c than expected from historical N_e (phylolm, p=0.012). Points in (**B**) and (**C**) show individual species, colored by taxonomic order.
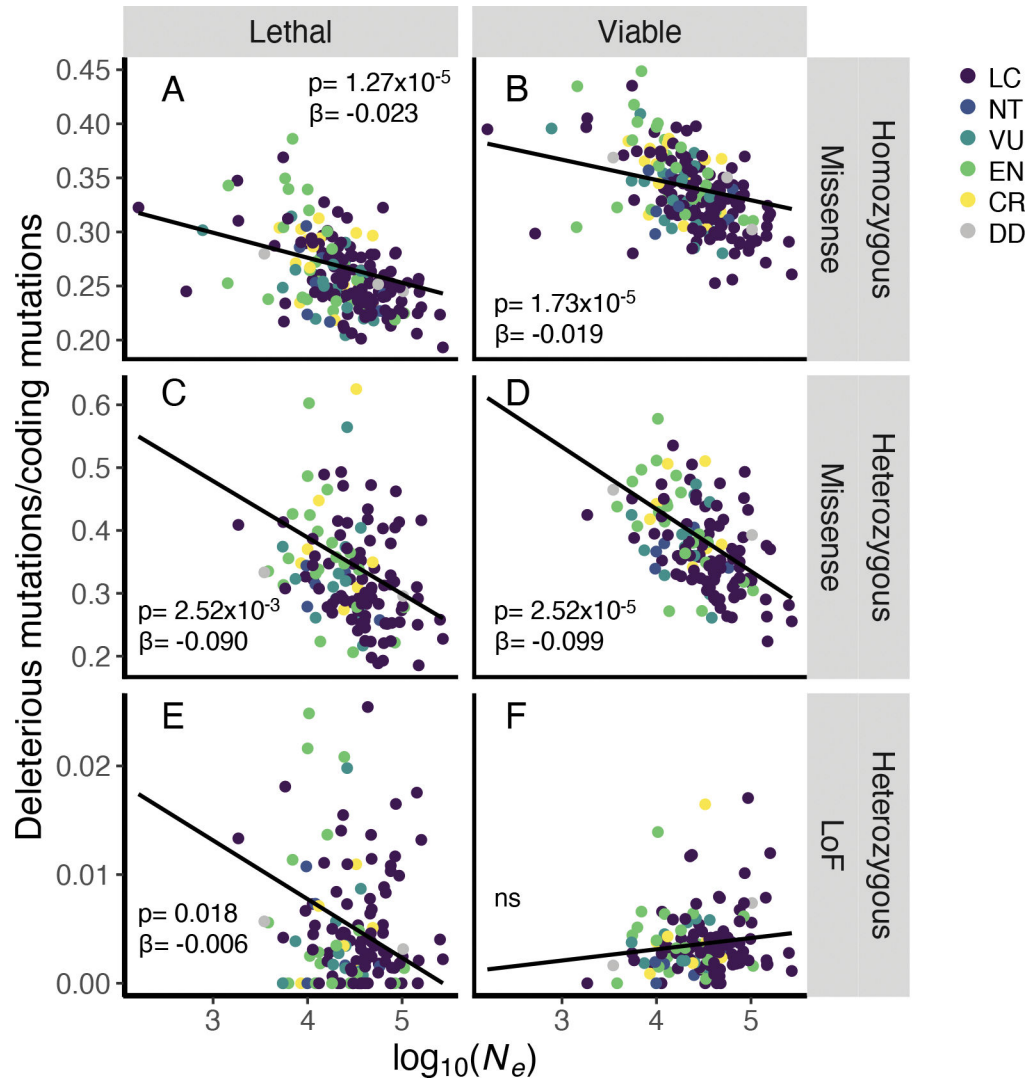
Fig. 2. Historically small populations have more deleterious genetic load in protein-coding genes.
Proportion of homozygous missense substitutions **(A-B)**, heterozygous missense variants
(C-D) and heterozygous loss-of-function variants **(E-F)** in genes as a function of historical
N_e across species. Genes were classified by associated lethal or viable phenotypes
in knockout mice. Proportions of heterozygous and homozygous missense mutations
were negatively correlated with N_e (all $p<0.052$), whereas heterozygous loss-of-function
alleles were not consistently correlated with N_e. Phylogenetically corrected p-values and
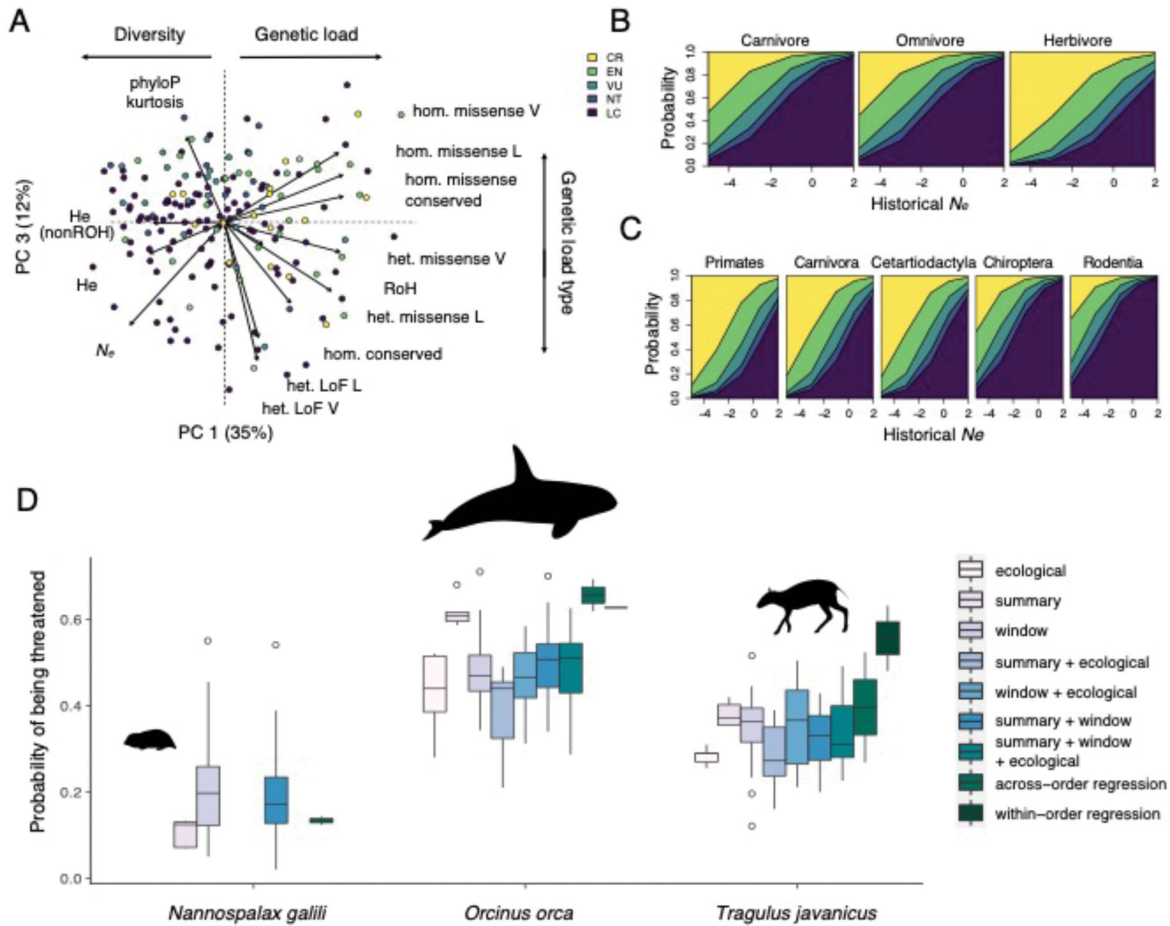coefficients (phylolm) are reported.

Fig. 3. Prediction of conservation status of species using genomic information.
(A) Principal components (PCs) that significantly predict threatened status. PC1 describes heterozygosity, N_e and deleterious variation, and PC3 distinguishes types of deleterious variation. Loadings of genomic variables (arrows; table S3) are labeled as described in table S2 (L=IMPC lethal genes; V=IMPC viable genes). Points indicate species, colored by IUCN status as shown in (B). **(B-C)** Probability of assignment to IUCN categories by diet and scaled values of historical N_e **(B)**, and by taxonomic order and historical N_e of species **(C)**. Decreased historical N_e is consistently associated with increased risk, but the magnitude varies by diet and taxonomic order. **(D)** Conservation status predictions for three data deficient species using random forest models with window-based metrics (windows), ecological variables (ecological), and/or genome-wide summary variables (summary), and predictions from regression models within and across taxonomic orders. *Nannospalax galili* lacked ecological data and adequate within-order data, so only predictions from across-order regression and windows models are shown for this species.