Estimates of genetic load in small populations suggest extensive purging of deleterious alleles

Tom van der Valk[1*], Marc de Manuel[2], Tomas Marques-Bonet[2,3,4,5], Katerina Guschanski[1*]

[1]Department of Ecology and Genetics, Animal Ecology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 752 36, Uppsala, Sweden

[2]Institute of Evolutionary Biology (UPF-CSIC), PRBB, Dr. Aiguader 88, 08003 Barcelona, Spain

[3]Catalan Institution of Research and Advanced Studies (ICREA), Passeig de Lluís Companys, 23, 08010, Barcelona, Spain

[4]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 4, 08028 Barcelona, Spain

[5]Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Edifici ICTA-ICP, c/ Columnes s/n, 08193 Cerdanyola del Vallès, Barcelona, Spain

[*]Corresponding authors: tom.vandervalk@ebc.uu.se (T.v.d.V.); katerina.guschanski@ebc.uu.se (K.G.)

Abstract: Declining populations are expected to experience negative genetic consequences of inbreeding, which over time can drive them to extinction. Yet, many species have survived in small populations for thousands of generations without apparent fitness effects, possibly due to genetic purging of partially deleterious recessive alleles in inbred populations. We estimate the abundance of deleterious alleles in a range of mammals and find that conversely to current conservation thinking species with historically small population size and low genetic diversity generally have lower genetic load compared to species with large population sizes. Rapid population declines will thus dis-proportionally affect species with high diversity, as they carry many deleterious alleles that can reach fixation before being removed by genetic purging.

Main Text: Small inbred populations of wild animals frequently show lower survival, less efficient mating and lower reproduction than large outbred populations (*1*), as consequence of high levels of genome-wide homozygosity, including at loci with partially recessive deleterious alleles (*2*). The negative fitness consequences of inbreeding have also been directly shown on the genomic level (*3*). Nevertheless, many animals have survived in small populations for thousands of generations without apparent strong negative fitness effects. A suggested explanation for this phenomenon is genetic purging — the increased efficiency of purifying selection at removing partially recessive deleterious alleles in inbred populations (*4*). Whereas in large populations partially recessive deleterious alleles are mostly found at low frequency, these alleles can drift to high frequency in small populations (*5*). Mating between related individuals subsequently brings recessive alleles in a homozygous state, exposing them to purifying selection and thus leading to their more efficient removal from small populations over time (*5*). Although genetic purging has been shown in several animal populations (*6*–*9*), it remains largely unknown to what extent it represents a central evolutionary force. As wild animal populations across the globe experience rapid human-caused declines (*10*), inbreeding and the resulting genetic consequences can directly contribute to their extinction (*11*). Understanding under what circumstances genetic purging acts and how common it is among endangered populations could therefore help to identify species facing the most severe genetic consequences of population declines.

To address this issue, we used genomic data to estimate the strength of genetic purging experienced by wild mammalian populations, as mammals are among the most affected by human-induced population declines (*10*). To identify deleterious alleles, we analysed evolutionary genomic constraints, which are accurate predictors for the fitness consequences of mutations (*12*). Genomic sites that remained conserved during millions of years of evolution are expected to be functionally important, and therefore mutations at such sites can serve as a proxy for genetic load – the reduction of population mean fitness due to genetic factors (*13*, *14*). Using a panel of 100 mammalian reference genomes, comprising all major mammalian lineages, we calculated the genomic evolutionary rate profiling (GERP) scores as the number of rejected substitutions, i.e. substitutions that would have occurred if the focal genomic element was neutral but did not occur because it has been under functional constrains (*15*) (Supplementary material). Mutations at highly conserved genomic sites (high GERP-scores) are likely deleterious, whereas those at low GERP-scores are expected to be mostly neutral. We then estimated individual relative genetic load in 670 individuals belonging to 42 mammalian species, using publicly available whole genome re-sequencing data, as the genome-wide average GERP-score for the derived alleles (Figure 1) (Supplementary material).

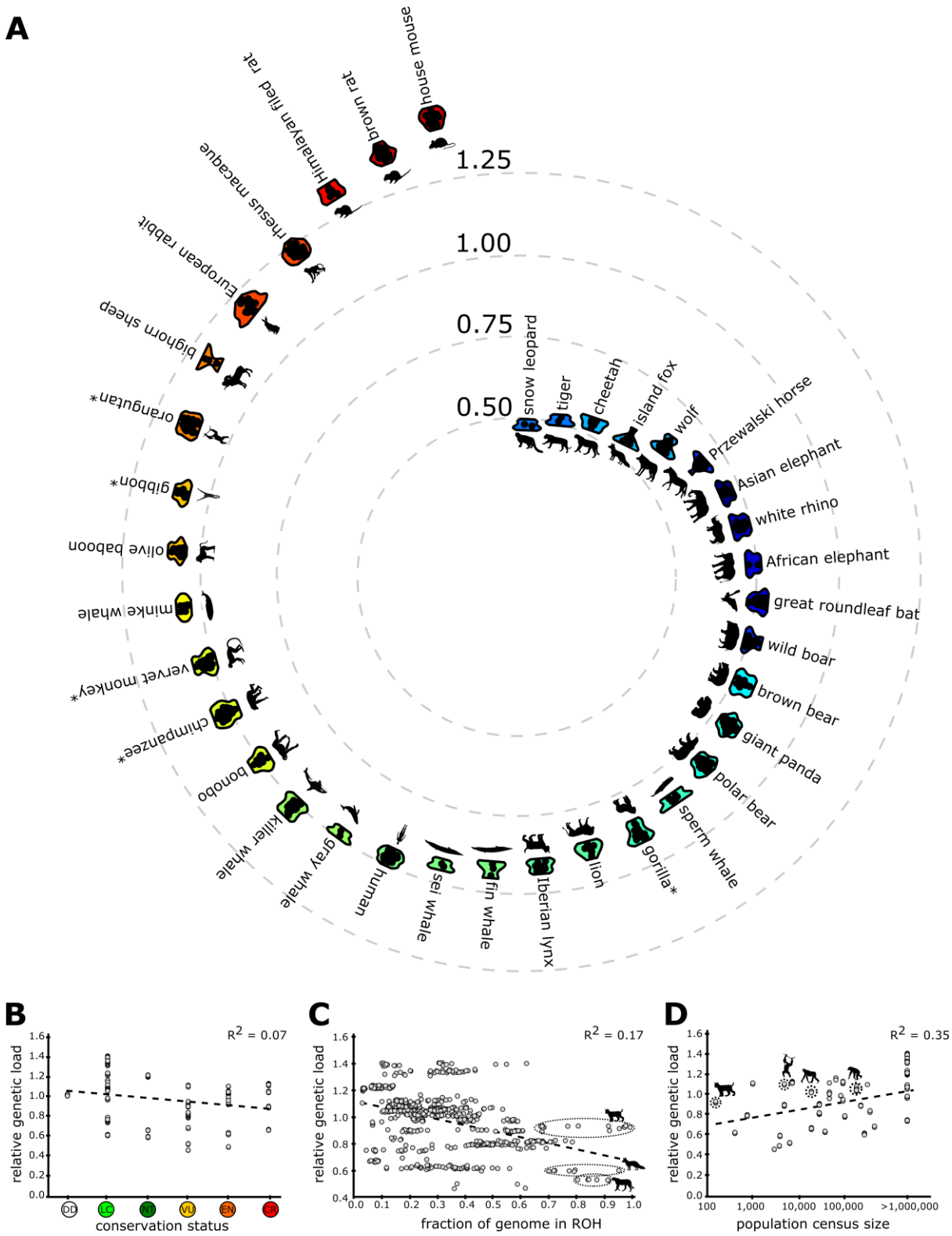

Fig. 1. Relative genetic load in mammals. *(**A**) Genetic load is depicted as the average GERP-score of the derived allele for each individual within a species. Several closely related species (Sumatran and Bornean orang-utans, gibbons, vervet monkeys, and eastern and western gorillas) are grouped together for clarity (depicted by the asterisks). (**B**) Relative genetic load is not explained by conservation status. DD: data deficient, LC: least concern, NT: near threatened, VU: vulnerable, EN:*

endangered, CR: critically endangered. ***(C)*** *Relative genetic load is negatively correlated with genetic diversity. Species with recurrent bottlenecks and/or small population size show low load despite high inbreeding (e.g. cheetah and island foxes). In contrast, some highly inbred species, which experienced recent dramatic population decline show disproportionally high genetic load (e.g. Iberian lynx).* ***(D)*** *Genetic load is generally higher in species with large census population size (species with population size above 1 million are grouped together for clarity). However, some species with historically large population sizes and recent strong declines (e.g. Sumatran orangutan, Iberian Lynx) show relatively high genetic load. Each grey dot represents a species in (B) and (D) and an individual genome in (C). Dotted lines depict the best fitting linear intercept.*

As the genome sequences of individuals belonging to the same species are highly similar, especially at the conserved sites, within-species differences in genetic load are generally based on few divergent alleles. We indeed observed few intraspecific differences, suggesting that our measure of genetic load reflects long-term evolutionary processes (e.g. over hundreds of generations) (Table S1, Fig. 1A). We found that estimates of genetic load differ strongly among the studied mammals (Fig. 1A) and do not correlate with species conservation status (Fig. 1B). We also did not detect a strong phylogenetic signal in genetic load, as closely related species (e.g. orang-utan and human, ~14-16 My divergence) differ strongly in their estimates of genetic load, whereas some highly divergent species (e.g. African elephant and great roundleaf bat, ~99-109 My divergence) (*16*) show comparable genetic load scores (Fig. 1A).

We observed a weak inverse relationship between genetic load and inbreeding (Fig. 1C). Species with low genetic load, i.e. relatively few derived alleles at putatively deleterious sites, have high proportion of their genome in runs of homozygosity (e.g. snow leopard, tiger, island fox, wolves, cheetah, Fig. 1C, Table S1). Conversely, species with high genetic load frequently have a low genome-wide rate of homozygosity (e.g. house mouse, brown rat, Himalayan rat, European rabbit, vervet monkey, olive baboon, rhesus macaque, Table S1). Large changes in levels of inbreeding can occur within only a few generations, which is also exemplified by the high degree of intra-species variation in inbreeding (±SD 27%) compared to genetic load (±SD 1.3%), a processes that takes place over hundreds of generations (Table S1). Individual measures of genetic load are therefore overall only weakly correlated with individual measures of inbreeding (R=0.17, Fig. 1C).

Contrary to the prevailing notion that small populations have high genetic load (*17*), we observe a positive relationship between relative genetic load and population size (Fig. 1D, Table S1). Generally, species with small population size have lower genetic load than species with large population sizes

(Fig. 1D, Table S1), suggesting that purging of deleterious alleles can be an important evolutionary force. However, we observe relatively high genetic load in several species with historically large population sizes that have experienced dramatic recent population declines (e.g. chimpanzees, orangutans, bonobos and Iberian lynx, Fig. 1D) (*18*). This corroborates recent findings from genetic simulations, which demonstrated that strong declines in population size disproportionally affect ancestrally large populations (*19*).

As selection can only act on variation, deleterious alleles that are fixed within a population are especially problematic for long-term population viability. We thus estimated the fraction of fixed derived alleles stratified by GERP-score for all species with at least five individuals in our dataset (Fig. 2). Generally, species with low genetic load carry few derived alleles at high GERP-scores (e.g. cheetah, island fox, Przewalski horse) (Figs. 1, 2), however, these alleles frequently appear to be fixed in the population (Fig. 2). In contrast, although some populations with high genetic load (e.g. house mouse, brown rat, Himalayan field rat, European rabbit, vervet monkey, olive baboon, rhesus macaque) carry relatively many putatively deleterious alleles, the majority of these are at low frequency and unlikely to appear in the homozygous state in any given individual (Fig. 2). Thus, while purging removes deleterious alleles in highly inbred species, some deleterious alleles nonetheless reach fixation, which can subsequently lead to negative fitness consequences without the opportunity for additional genetic purging. This could also explain why inbreeding depression has been reported in the cheetah and (Swedish) wolves despite the relatively low overall genetic load (*20*, *21*) . Taken together, these observation are especially worrying for genetically diverse populations that experience rapid population declines, as we show that species with high genetic diversity generally carry relatively many deleterious alleles and thus a high proportion of these could reach fixation before genetic purging can act (see also *19*).

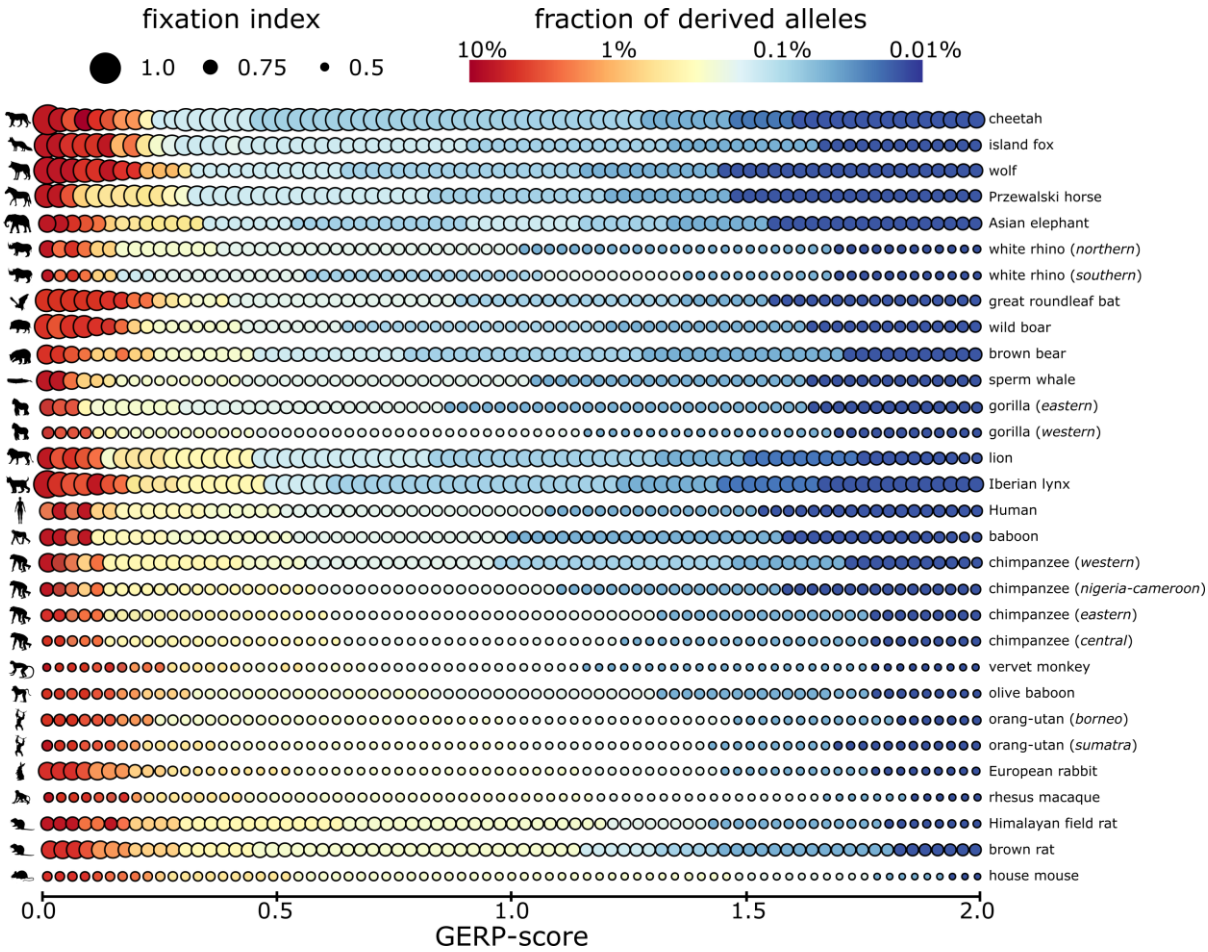

Fig. 2. Fixation of derived alleles. *Species are ranked by average GERP-score from the lowest (at the top of the graph) to highest (bottom of the graph). Circle sizes represent the fraction of derived alleles that are fixed within the population for a given GERP-score bin. The colour depicts the percentage of derived alleles within a given GERP-score bin out of all derived alleles in the population. The majority of derived alleles and most fixed derived alleles are found at low GERP-scores and hence in regions of low selective constraint (Figs. S5, S6). Species with a low genetic load carry low proportion of derived alleles at high GERP-scores many of which are fixed, whereas species with high genetic load (at the bottom of the graph) show many derived alleles at high GERP-scores that are however less often fixed in the population.*

Higher genetic load of individuals from large populations calls into question the commonly employed conservation strategy of genetic rescue, the increase of genetic diversity in inbred populations through introduction of outbred individuals. Although genetic rescue can increase population fitness on the short-term (*22*), the long-term effects can be dramatic. This is exemplified by the collapse of the Isle Royale wolves, a population that maintained good population viability for decades. After interbreeding with a mainland wolf migrant, the Isle Royale wolves initially showed higher

reproductive success. However, subsequent inbreeding in this population eventually resulted in the increase in frequency of deleterious alleles, most likely introduced by the immigrant, and eventual marked decline of the population (*23*). The translocation of an outbred individual with high absolute number of deleterious alleles (even though these alleles segregate at low frequency within the population) into an inbred population that experienced genetic purging will thus often have strongly negative consequences. We thus warn against genetic rescue strategies if they are not followed by a clearly delineated long-term plan to reduce inbreeding, for instance through repeated introductions or through use of pre-screened individuals with low genetic load.

References:

1. L. F. Keller, D. M. Waller, Inbreeding effects in wild populations. *Trends Ecol. Evol.* **17** (2002), pp. 230–241.

2. B. Charlesworth, D. Charlesworth, The genetic basis of inbreeding depression. *Genet. Res.* **74**, 329–40 (1999).

3. R. L. Rogers, M. Slatkin, Excess of genomic defects in a woolly mammoth on Wrangel island. *PLOS Genet.* **13**, e1006601 (2017).

4. P. W. Hedrick, Purging inbreeding depression and the probability of extinction: full-sib mating. *Heredity (Edinb).* **73**, 363–372 (1994).

5. D. Charlesworth, J. H. Willis, The genetics of inbreeding depression. *Nat. Rev. Genet.* **10**, 783–796 (2009).

6. J. A. Robinson, C. Brown, B. Y. Kim, K. E. Lohmueller, R. K. Wayne, Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. *Curr. Biol.* **28**, 3487-3494.e4 (2018).

7. Y. Xue, J. Prado-Martinez, P. H. Sudmant, V. Narasimhan, Q. Ayub, M. Szpak, P. Frandsen, Y. Chen, B. Yngvadottir, D. N. Cooper, M. de Manuel, J. Hernandez-Rodriguez, I. Lobon, H. R. Siegismund, L. Pagani, M. A. Quail, C. Hvilsom, A. Mudakikwa, E. E. Eichler, M. R. Cranfield, T. Marques-Bonet, C. Tyler-Smith, A. Scally, Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science*. **348**, 242–5 (2015).

8. A. Benazzo, E. Trucchi, J. A. Cahill, P. Maisano Delser, S. Mona, M. Fumagalli, L. Bunnefeld, L. Cornetti, S. Ghirotto, M. Girardi, L. Ometto, A. Panziera, O. Rota-Stabelli, E. Zanetti, A. Karamanlidis, C. Groff, L. Paule, L. Gentile, C. Vilà, S. Vicario, L. Boitani, L. Orlando, S. Fuselli, C. Vernesi, B. Shapiro, P. Ciucci, G. Bertorelle, Survival and divergence in a small group: The extraordinary genomic history of the endangered Apennine brown bear stragglers. *Proc. Natl. Acad. Sci.* **114**, E9589–E9597 (2017).

9. T. van der Valk, D. Díez-del-Molino, T. Marques-Bonet, K. Guschanski, L. Dalén, Historical Genomes Reveal the Genomic Consequences of Recent Population Decline in Eastern Gorillas. *Curr. Biol.* **29**, 165-170.e6 (2019).

10. S. L. Pimm, C. N. Jenkins, R. Abell, T. M. Brooks, J. L. Gittleman, L. N. Joppa, P. H. Raven, C. M. Roberts, J. O. Sexton, The biodiversity of species and their rates of extinction, distribution, and protection. *Science (80-.).* **344** (2014), pp. 1246752–1246752.

176 11. D. Spielman, B. W. Brook, R. Frankham, Most species are not driven to extinction before
177 genetic factors impact them. *Proc. Natl. Acad. Sci.* **101**, 15261–15264 (2004).

178 12. G. M. Cooper, J. Shendure, Needles in stacks of needles: finding disease-causal variants in a
179 wealth of genomic data. *Nat. Rev. Genet.* **12**, 628–640 (2011).

180 13. P. Librado, C. Gamba, C. Gaunitz, C. Der Sarkissian, M. Pruvost, A. Albrechtsen, A. Fages, N.
181 Khan, M. Schubert, V. Jagannathan, A. Serres-Armero, L. F. K. Kuderna, I. S. Povolotskaya, A.
182 Seguin-Orlando, S. Lepetz, M. Neuditschko, C. Thèves, S. Alquraishi, A. H. Alfarhan, K. Al-
183 Rasheid, S. Rieder, Z. Samashev, H.-P. Francfort, N. Benecke, M. Hofreiter, A. Ludwig, C.
184 Keyser, T. Marques-Bonet, B. Ludes, E. Crubézy, T. Leeb, E. Willerslev, L. Orlando, Ancient
185 genomic changes associated with domestication of the horse. *Science (80-.).* **356**, 442–445
186 (2017).

187 14. B. M. Henn, L. R. Botigué, S. Peischl, I. Dupanloup, M. Lipatov, B. K. Maples, A. R. Martin, S.
188 Musharoff, H. Cann, M. P. Snyder, L. Excoffier, J. M. Kidd, C. D. Bustamante, Distance from
189 sub-Saharan Africa predicts mutational load in diverse human genomes. *Proc. Natl. Acad. Sci.*
190 *U. S. A.* **113**, E440-9 (2016).

191 15. E. V Davydov, D. L. Goode, M. Sirota, G. M. Cooper, A. Sidow, Identifying a High Fraction of
192 the Human Genome to be under Selective Constraint Using GERP++. *PLoS Comput Biol*. **6**,
193 1001025 (2010).

194 16. S. Kumar, G. Stecher, M. Suleski, S. B. Hedges, TimeTree: A Resource for Timelines, Timetrees,
195 and Divergence Times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).

196 17. D. H. Reed, R. Frankham, Correlation between fitness and genetic diversity. *Conserv. Biol.* **17**,
197 230–237 (2003).

198 18. IUCN, The IUCN Red List of Threatened Species. Version 2019.1 (2019).

199 19. C. C. Kyriazis, R. K. Wayne, K. E. Lohmueller, High genetic diversity can contribute to
200 extinction in small populations. *bioRxiv*, 678524 (2019).

201 20. D. E. Wildt, M. Bush, J. G. Howard, S. J. O'Brien, D. Meltzer, A. Van Dyk, H. Ebedes, D. J. Brand,
202 Unique seminal quality in the South African cheetah and a comparative evaluation in the
203 domestic cat. *Biol. Reprod.* **29**, 1019–25 (1983).

204 21. M. Åkesson, O. Liberg, H. Sand, P. Wabakken, S. Bensch, Ø. Flagstad, Genetic rescue in a
205 severely inbred wolf population. *Mol. Ecol.* **25**, 4745–4756 (2016).

22. R. Frankham, Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. *Mol. Ecol.* **24**, 2610–8 (2015).

23. J. A. Robinson, J. Räikkönen, L. M. Vucetich, J. A. Vucetich, R. O. Peterson, K. E. Lohmueller, R. K. Wayne, "Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction" (2019).

Funding: TvdV is supported by a scholarship from the Foundation for Zoological Research. TMB is supported by BFU2017-86471-P (MINECO/FEDER, UE), U01 MH106874 grant, Howard Hughes International Early Career, Obra Social "La Caixa" and Secretaria d'Universitats i Recerca and CERCA Programme del Departament d'Economia i Coneixement de la Generalitat de Catalunya (GRC 2017 SGR 880). KG is supported by a Formas grant (2016-00835). The computations for this study were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project SNIC SNIC 2017/7-255.

Author contributions: TvdV and MdM analysed the data. All authos were involved in the study design and interpetration of the results. KG and TvdV wrote the manuscript with input from all authors.

Data and materials availability: All scripts used in this study will be available on github/tvdvalk.

Supplementary Materials:

Materials and Methods

Fig S1 – S8

References (1 – 18)

Supplementary Data S1 – S2

Supplementary Materials for:

Estimates of genetic load in small populations suggest extensive purging of deleterious alleles

Tom van der Valk, Marc de Manuel, Tomas Marques-Bonet, Katerina Guschanski

Materials and Methods

Fig S1 – S8

References (1 – 18)

Supplementary Data S1 – S2

Materials and Methods

Single nucleotide variant calling

We obtained published re-sequencing data for 670 mammalian genomes from 42 species and mapped these to the phylogenetically closest available reference genome for each species (Table S1) using bwa mem v0.7.17 (*1*). In total, 27 reference genomes were used for this task (Table S1). We then obtained and filtered variant calls for each individual using GATK HaplotypeCaller v3.8 following the "short variant discovery best-practices guidelines" including "hard filtering" (*2*). Additionally, we only kept within-species bi-allelic sites and removed all indels and sites below one third and above three times the genome-wide autosomal coverage (*3*).

Genomic Evolutionary Rate Profiling

We used the software GERP++ (Genomic Evolutionary Rate Profiling) to calculate the number of "rejected substitutions" (a proxy for evolutionary constrains) for each site in the same 27 reference genomes that were used in mapping of the re-sequencing data (Table S1, Fig. S1) (*4*). GERP++ estimates the number of substitutions that would have occurred if the site was neutral given a multi-species sequence alignment and the divergence time estimates between the aligned species as provided in (*5*). A GERP-score, the number of rejected substitutions at a genomic site, is thus a measure of constraint that reflects the strength of past purifying selection at a particular locus. To calculate GERP-scores for a given focal reference genome, we used 100 non-domesticated mammalian de-novo assembled genomes (Table S2, Fig. S2), as domesticated species might give a biased estimate of purifying selection. Each individual genome sequence was converted into short FASTQ reads by sliding across the genome in non-overlapping windows of 50 base pairs and transforming each window into a separate FASTQ read. The

resulting FASTQ reads from the 100 mammalian genomes were then mapped to each respective focal reference genome with bwa mem v0.7.17, slightly lowering the mismatch penalty (-B 3) and removing reads that mapped to multiple regions. Mapped reads were realigned around indels using GATK IndelRealigner (*6*, *7*). Next, we converted the mapped reads into a haploid FASTA consensus sequence (i.e. 100 times for each reference genome), excluding all sites with depth above one (as such sites contain at least one mismapped read). GERP++ was then used to calculate the number of rejected substitutions at all sites in the reference using the concatenated FASTA files and the species divergence time estimates from (*5*) (Fig. S2), excluding the focal reference from the calculation. Missing bases within the concatenated alignment were treated as non-conserved (i.e. sites for which only few reads mapped obtain low GERP scores). We excluded all sites for which the focal reference FASTQ reads did not map to themselves and sites with negative GERP-scores (as these most likely represent errors) and subsequently scaled all scores to a range from 0 to 2. Sites that are identical between species and have thus been preserved over long evolutionary time result in high GERP-scores (Fig. S1). Thus, high GERP-scores are only obtained for regions, where the majority of the 99 mammalian genomes (100 minus the focal reference) map to the respective reference.

Validating GERP scores

Inferring alleles with deleterious effects from genomic data of non-model organisms is hampered by the lack of functional information. For species with abundant medical data (e.g. human and mice), the deleterious effects for many (disease) variants are known, and thus the screening of individuals for such mutations can provide an estimate of genetic load (*8*). For non-model species, tools have been developed to assist in identifying mutations affecting regulatory sequences or those altering protein structure (*9–11*). However, such tools rely on accurate genome annotations, which are only available for a limited number of species (*12*). Here, we use an estimate of genome conservation across evolutionary time, measured by GERP-scores, as a proxy for the deleteriousness of a given genomic variant. Although, this method is limited with respect to identifying the likely fitness consequences of each individual variant, genome-wide measures can provide an indication of the relative genetic load within an individual without relying on curated databases or genome annotations (*13*). More importantly for our study, it allows for between-species comparison as long as a reference genome for the species of interest (or a closely related species) is available. GERP-scores have previously been calculated for the human reference (hg19) based on the whole genome alignment of 44 vertebrate genomes (*4*). As alignment algorithms are designed to obtain matches between highly divergent sequences, these alignments allow for the identification of highly conserved regions as well as those regions evolving faster than expected under neutrality (*4*). Obtaining such alignments requires considerable computational resources (*14*), is error prone (*15*) and non-scalable (e.g. the analysis is limited to those species that are part of the alignment set). In this study we used a short-read mapping based approach for the GERP-score calculations (Fig. S1). This pipeline is flexible, as it requires considerably less computational resources than whole-genome alignment approaches and thus GERP-scores can be

readily calculated for a broad set of study organisms. In addition, we doubled the number of genomes used for the GERP-score calculation in comparison to the previously published scores (*4*), possibly improving the accuracy. Although our used method is not suitable for the identification of fast evolving regions (as the reads do not map to highly distinct sequences), it performs well for genomic regions that are conserved among species. We validated our method using four independent approaches. First, for the human genome, we obtained a high correlation between the (positive) GERP-scores previously calculated based on the 44 whole-genome vertebrate alignment and those obtained with our pipeline (Pearson correlation $r = 0.944$) (Fig. S3). Second, our calculated GERP-scores are 4-6 times higher within exonic regions, known to be highly conserved and under purifying selection in vertebrates (*14*), than in intronic regions ($P < 2.2 \cdot 10^{-16}$) (Fig. S4). Third, the majority of within-population variable sites (88% ±SE 5% across all species) are found at the lowest 10% of GERP-scores, suggesting that low GERP-scores reflect neutrally evolving, variable regions, whereas variants at high GERP-scores are mostly removed from the populations by selection (Fig. S5). Finally, we observe that derived alleles are found in heterozygous state more often at high GERP-scores compared to low GERP-scores (where they often appear in the homozygous state, Fig. S6), suggesting that many derived alleles at high GERP-scores are likely to be recessive deleterious.

Ancestral allele inference

We called the ancestral allele at each site as the variant present in the phylogenetically closest outgroup. By using only one outgroup we retained the highest number of sites to be analysed (the more outgroups are added, the fewer sites will be mapped across all outgroups). We estimated the effect of using one or multiple outgroups for the ancestral allele inference by calling the majority allele among the mapped reads for 1 to 4 outgroups (a random base was choses if the allele frequency was equal) and show that this does not significantly change the estimates of genetic load (Fig. S7, S8), as genomic sites with high GERP-scores are generally conserved and thus identical among all outgroup species (Fig. S7). The derived alleles in each individual from the study dataset was then inferred against the called ancestral allele.

Relative genetic load

We estimated relative genetic load for each of the 670 study genomes as the average GERP-score of all derived alleles:

$$\frac{\sum_{i=0}^{n} D_i \cdot gerp_i}{\sum_{i=0}^{n} D_i}$$

Where D_i represents the ith derived allele and $gerp_i$ the GERP-score for the ith allele. Under the assumption that new mutations occur randomly with respect to the genomic region, we expect that in species that experienced strong purifying selection, derived alleles are found mostly at non-conserved sites (low GERP-scores), whereas accumulation of deleterious variants should result in a higher fraction of derived alleles at high GERP-scores.

Fixation of deleterious alleles

The fraction of fixed derived alleles was estimated for all species for which at least five individuals (e.g. 10 alleles) were present in our dataset. For species with more than five sequenced individuals, we randomly sampled 10 alleles at each site to exclude sample size bias. In both cases, we calculated the fraction of fixed derived alleles stratified by GERP-score.

Individual inbreeding estimates

We used PLINK1.9 (*16*) to identify the fraction of the genome in runs of homozygosity longer than 100kb, a measure of inbreeding (F_{ROH}), for all individuals with average genome coverage > 3X as in (*17*, *18*). To this end, we ran sliding windows of 50 SNPs on the VCF files, requiring at least one SNP per 50kb. In each individual genome, we allowed for a maximum of one heterozygous and five missing calls per window before we considered the ROH to be broken. To account for differences in genome assembly qualities we restricted our analysis to contigs of at least 1 megabase.

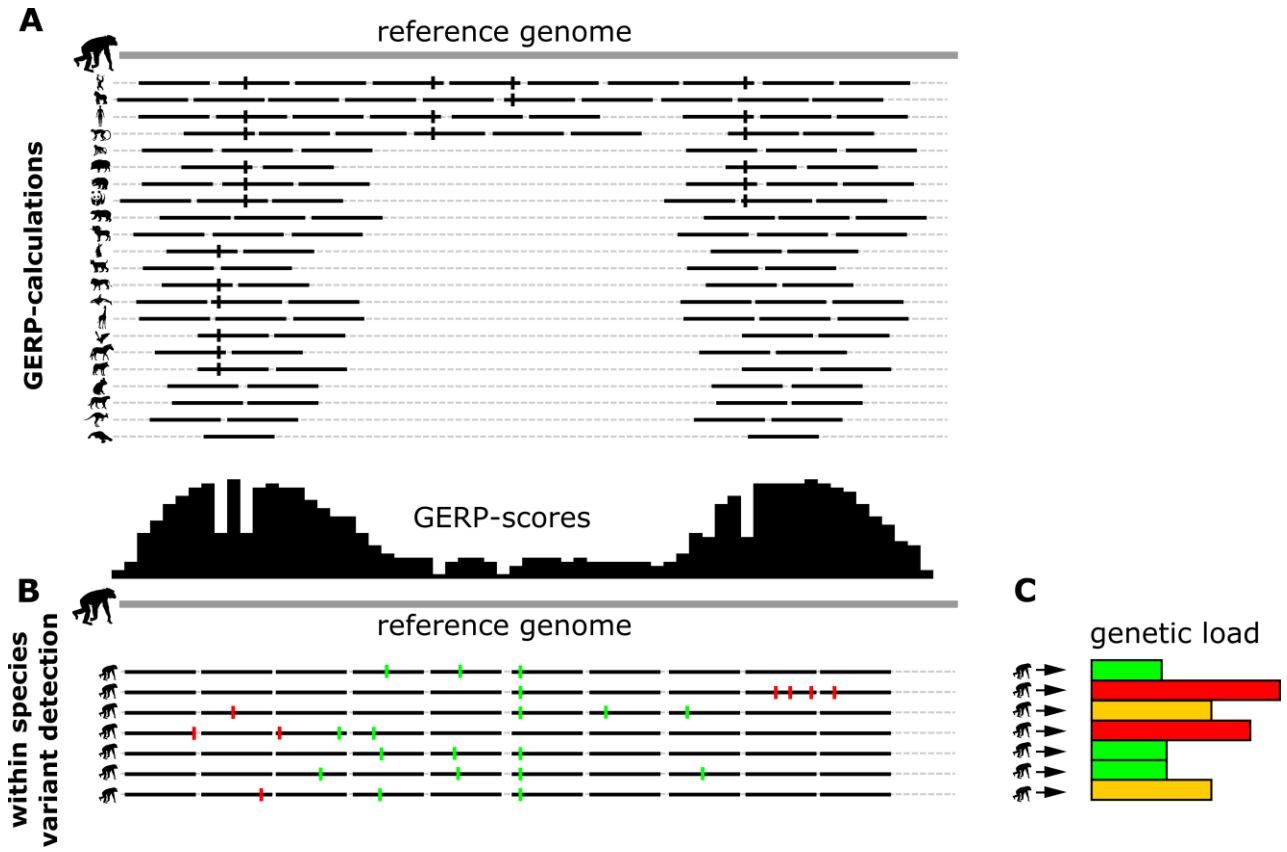

Fig. S1.

Schematic representation of the GERP-score pipeline. (A) A set of 100 de-novo assembled genomes is sliced into non-overlapping 50 base pair FASTQ read and aligned to the same reference as used for the within-species variance detection (SNP calling in (B)). A consensus sequence is then obtained for each of these 100 mapped genomes and GERP scores are subsequently calculated using the GERP++ software (excluding the focal reference from the calculation). Sites with few mapped reads or with a large proportion of variable alleles (depicted with vertical black bars on the individual reads) obtain low GERP scores, whereas sites identical among the majority of the mapped genomes obtain high GERP-scores. **(B)** Individual re-sequenced genomes from a population of a given study species are mapped to the reference

genome (chimpanzee in this example) and SNPs are subsequently identified for each individual within a population following the GATK "short variant discovery best practise" guidelines. **(C)** The genetic load of the derived alleles identified in (B) can now be estimated. Derived alleles at highly conserved sites are more likely to have a negative fitness effect (depicted with the red vertical bars in B) compared to derived alleles at less conserved sites (green vertical bars in B). The average GERP-score of the derived alleles is a measure of the relative genetic load carried by each individual (red=high genetic load, orange=intermediate genetic load, green=low genetic load).

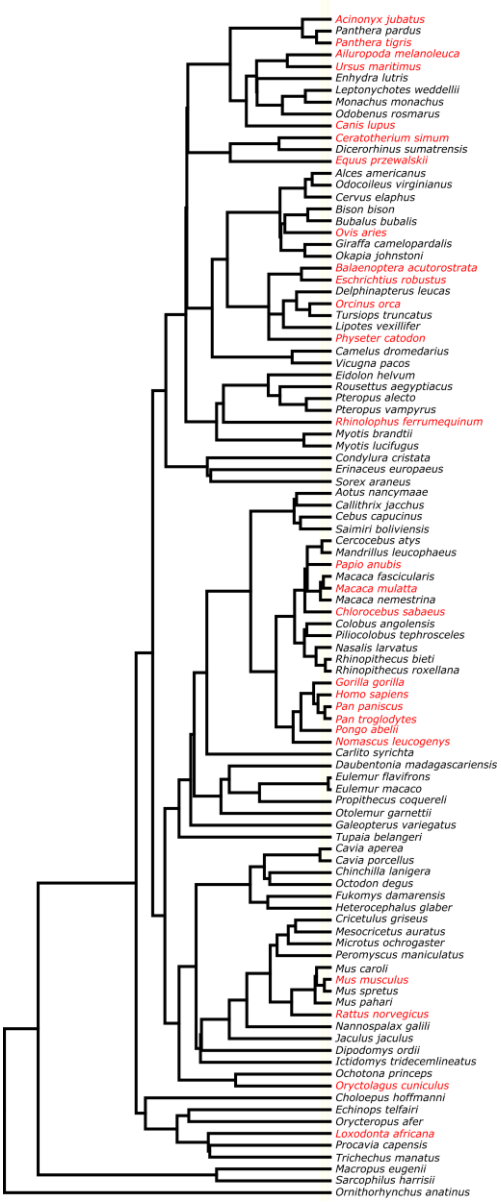

Fig. S2.

The 100 mammalian genomes and their divergence times estimates used for the GERP-score calculations. The divergence times between the species were obtained using the online software *TimeTree* which gives a dated phylogeny from a list of species through automated literature searches (*5*). The genomes depicted in red were also used for the mapping of re-sequencing data.

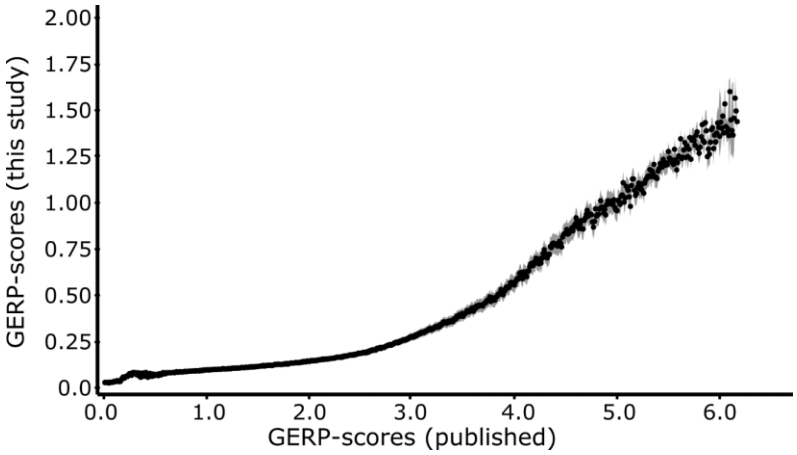

Fig. S3.

Correlation between previously published alignment-based GERP-scores and the GERP-scores calculated with the mapping-based approach in this study for the human genome (hg19). We binned all sites in the human genome by their published GERP-scores and calculated the average GERP-score for each bin of size 10 (black dots). Grey shaded area depicts ±1SD. Pearson correlation = 0.944, Spearman's rank correlation = 0.997. Note that we transformed our GERP-scores on a scale from 0 to 2, whereas the published scores are on the scale from 0 to 6.

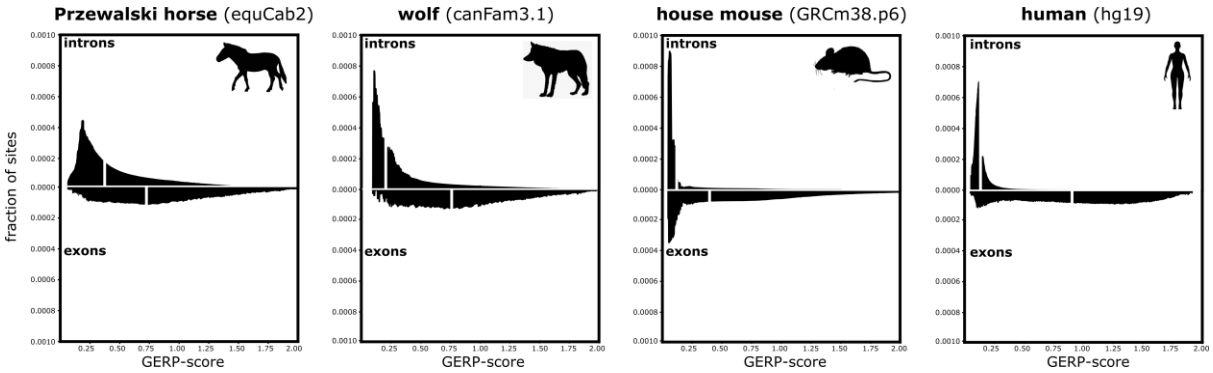

Fig. S4.

GERP-score of genomic partitions. The distribution of GERP-scores within introns (top panel) and exons (bottom panel) for 4 species with available high-quality reference genome annotations (used references between brackets). White lines within the plots depict the average GERP-score for a given genomic category. The highest GERP-scores are primarily found within exonic regions, with the average GERP-score in exons 4-6 times higher than within introns ($P < 2.2 \cdot 10^{-16}$ for all four species).

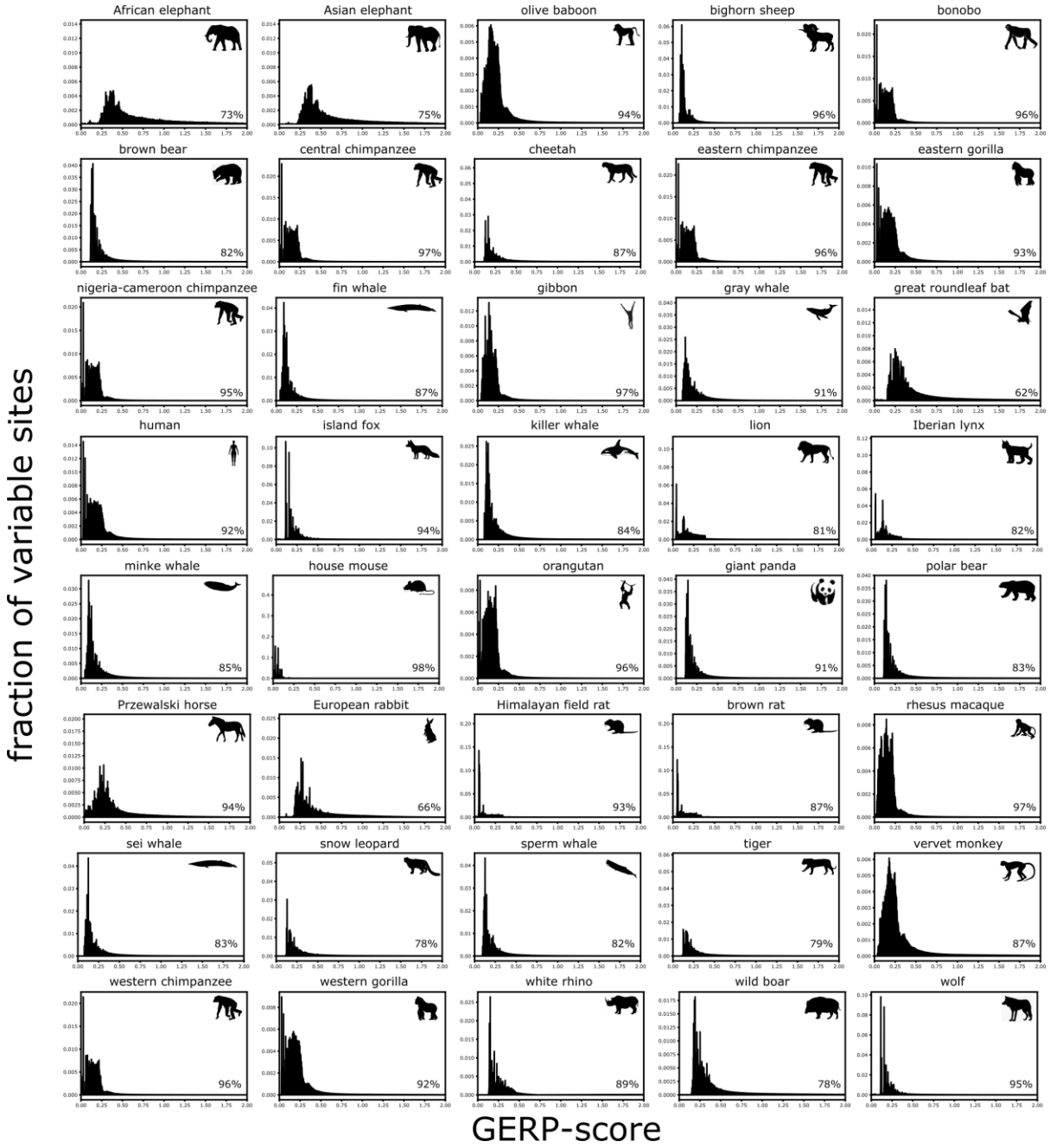

Fig. S5.

Distribution of within-population variable sites by GERP-score. The proportion of variable sites found within the genomic regions with the 10% lowest GERP-scores is depicted in the bottom right corner.

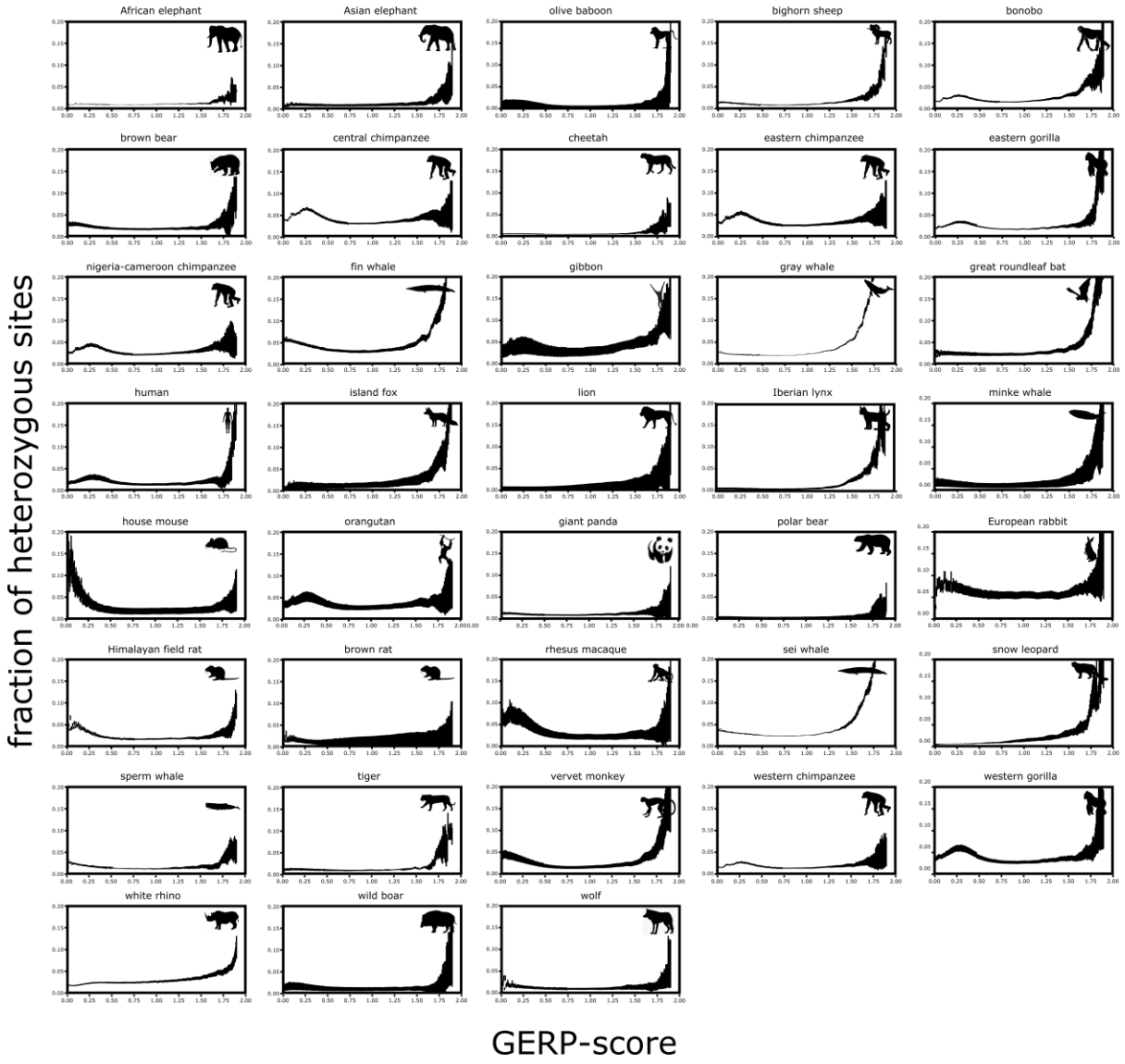

Fig. S6.

Proportion of heterozygous alleles out of total sites stratified by GERP-score. We included only samples with average genome wide coverage > 10X. Y-axis is scaled form 0 to 0.2 for clarity.

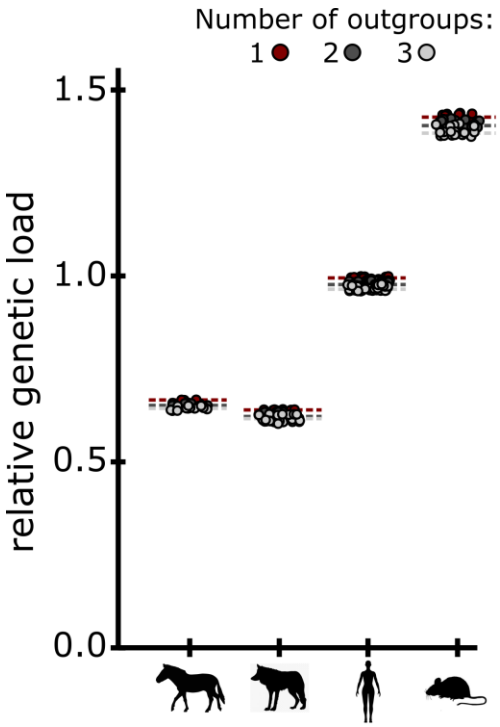

Fig. S7.

Relative genetic load calculated as the average GERP-score of the derived alleles using different number of outgroups. We inferred the derived state by either using one outgroup or the majority allele among 2 or 3 outgroups. We then re-calculated the genetic load for a phylogenetically diverse group of species (the Przewalski's horse, wolf, human and house mouse). Circles represent individual estimates and dotted lines depict the population averages for different number of used outgroups.

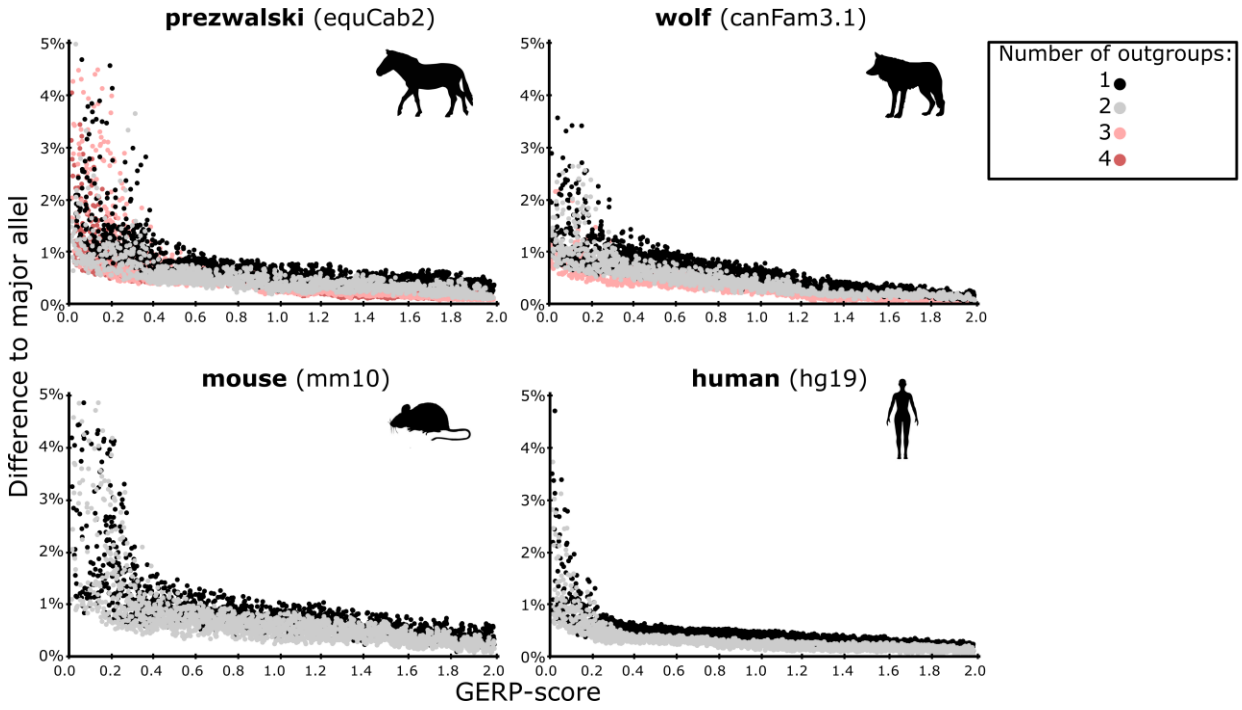

Fig. S8.

Ancestral allele inference depending on the number of used outgroups. Plots show the percentage of nucleotide differences to the major allele (among the complete phylogeny, e.g. all species that mapped to the site) by GERP-score depending on the number of outgroups used ot inter the ancestral allele. Increasing the number of outgroups only slightly increases the likelihood of calling the correct ancestral allele and comes at the cost of having fewer sites in total.

Table S1.

Individual genome re-sequencing data used to estimate genetic load and FROH
(Provided as a separate file)

Table S2.

Reference genomes used to calculate GERP-scores
(Provided as a separate file)

References

1. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM (2013), doi:arXiv:1303.3997.
2. G. A. Van der Auwera, M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel, M. A. DePristo, in *Current Protocols in Bioinformatics* (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2013), vol. 43, pp. 11.10.1-11.10.33.
3. H. Li, J. Wren, Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*. **30** (2014), pp. 2843–2851.
4. E. V Davydov, D. L. Goode, M. Sirota, G. M. Cooper, A. Sidow, Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. *PLoS Comput Biol*. **6**, 1001025 (2010).
5. S. Kumar, G. Stecher, M. Suleski, S. B. Hedges, TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
6. M. A. DePristo, E. Banks, R. Poplin, K. V Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
7. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–303 (2010).
8. B. M. Henn, L. R. Botigué, C. D. Bustamante, A. G. Clark, S. Gravel, Estimating the mutation load in human genomes. *Nat. Rev. Genet.* **16**, 333–343 (2015).
9. W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek, F. Cunningham, M. Eisenstein, M. Weil, A. Chen, P. Visscher, M. Brown, M. McCarthy, J. Yang, A. Saint Pierre, E. Génin, O. Zuk, S. Schaffner, K. Samocha, R. Do, E. Hechter, S. Kathiresan, L. Hindorff, P. Sethupathy, H. Junkins, E. Ramos, J. Mehta, F. Collins, X. Puente, S. Beà, R. Valdés-Mas, N. Villamor, J. Gutiérrez-Abril, J. Martín-Subero, D. Gudbjartsson, H. Helgason, S. Gudjonsson, F. Zink, A. Oddson, A. Gylfason, F. Collins, H. Varmus, K.-P. Koepfli, B. Paten, S. O'Brien, J. Cao, K. Schneeberger, S. Ossowski, T. Günther, S. Bender, J. Fitz, H. Daetwyler, A. Capitan, H. Pausch, P. Stothard, R. Binsbergen, R. Brøndum, C. Gonzaga-Jauregui, J. Lupski, R. Gibbs, J. Harrow, A. Frankish, J. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, K. Pruitt, G. Brown, S. Hiatt, F. Thibaud-Nissen, A. Astashyn, O. Ermolaeva, R. Dalgleish, P. Flicek, F. Cunningham, A. Astashyn, R. Tully, G. Proctor, F. Cunningham, M. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, S. Pabinger, A. Dander, M. Fischer, R. Snajder, M. Sperk, M. Efremova, W. McLaren, B. Pritchard, D. Rios, Y. Chen, P. Flicek, F. Cunningham, J. Höglund, G. Sahana, R. Brøndum, B. Guldbrandtsen, B. Buitenhuis, M. Lund, T. Godoy, G. Moreira, C. Boschiero, A. Gheyas, G. Gasparin, M. Paduan, E. Leslie, M. Taub, H. Liu, K. Steinberg, D. Koboldt, Q. Zhang, L. Hou, H. Zhao, C. Saunders, N. Miller, S. Soden, D. Dinwiddie, A. Noll, N. Alnadi, C. Wright, T. Fitzgerald, W. Jones, S. Clayton, J. McRae, M. Kogelenberg, G. McVean, A. Auton, L. Brooks, M. DePristo, R. Durbin, R. Handsaker, U. Paila, B. Chapman, R. Kirchner, A. Quinlan, P. Kersey, J. Allen, M.

Christensen, P. Davis, L. Falin, C. Grabmueller, A. Frankish, B. Uszczynska, G. Ritchie, J. Gonzalez, D. Pervouchine, R. Petryszak, J. Rodriguez, P. Maietta, I. Ezkurdia, A. Pietrelli, J.-J. Wesselink, G. Lopez, R. Petryszak, T. Burdett, B. Fiorelli, N. Fonseca, M. Gonzalez-Porta, E. Hastings, P. Kumar, S. Henikoff, P. Ng, I. Adzhubei, S. Schmidt, L. Peshkin, V. Ramensky, A. Gerasimova, P. Bork, A. Gonzalez-Perez, J. Deu-Pons, N. Lopez-Bigas, H. Shihab, J. Gough, D. Cooper, P. Stenson, G. Barker, K. Edwards, J. Schwarz, D. Cooper, M. Schuelke, D. Seelow, L. Ward, M. Kellis, D. Zerbino, S. Wilder, N. Johnson, T. Juettemann, P. Flicek, D. Adams, L. Altucci, S. Antonarakis, J. Ballesteros, S. Beck, A. Bird, C. Romanoski, C. Glass, H. Stunnenberg, L. Wilson, G. Almouzni, G. Cooper, E. Stone, G. Asimenos, E. Green, S. Batzoglou, A. Sidow, G. Ritchie, I. Dunham, E. Zeggini, P. Flicek, M. Kircher, D. Witten, P. Jain, B. O'Roak, G. Cooper, J. Shendure, H. Shihab, J. Gough, M. Mort, D. Cooper, I. Day, T. Gaunt, Y. Chen, F. Cunningham, D. Rios, W. McLaren, J. Smith, B. Pritchard, D. Rios, W. McLaren, Y. Chen, E. Birney, A. Stabenau, P. Flicek, S. Sherry, M. Ward, M. Kholodov, J. Baker, L. Phan, E. Smigielski, S. Forbes, N. Bindal, S. Bamford, C. Cole, C. Kok, D. Beare, I. Lappalainen, J. Lopez, L. Skipper, T. Hefferon, J. Spalding, J. Garner, D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, M. Landrum, J. Lee, G. Riley, W. Jang, W. Rubinstein, D. Church, P. Cingolani, A. Platts, L. Wang, M. Coon, T. Nguyen, L. Wang, L. Clarke, X. Zheng-Bradley, R. Smith, E. Kulesha, C. Xiao, I. Toneva, E. Bragin, E. Chatzimichali, C. Wright, M. Hurles, H. Firth, A. Bevan, D. McCarthy, P. Humburg, A. Kanapin, M. Rivas, K. Gaulton, J.-B. Cazier, K. Ardlie, D. Deluca, A. Segrè, T. Sullivan, T. Young, J. Stajich, D. Block, K. Boulez, S. Brenner, S. Chervitz, C. Dagdigian, A. Tan, G. Abecasis, H. Kang, J. Dunnen, S. Antonarakis, K. Wang, M. Li, H. Hakonarson, X. Liu, X. Jian, E. Boerwinkle, X. Jian, E. Boerwinkle, X. Liu, The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

10. I. Adzhubei, D. M. Jordan, S. R. Sunyaev, Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **76** (2013).
11. P. Cingolani, A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, D. M. Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin).* **6**, 80–92 (2012).
12. M. Yandell, D. Ence, A beginner's guide to eukaryotic genome annotation. *Nat. Rev. Genet.* **13**, 329–342 (2012).
13. G. M. Cooper, J. Shendure, Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat. Rev. Genet.* **12**, 628–640 (2011).
14. K. Lindblad-Toh, M. Garber, O. Zuk, M. F. Lin, B. J. Parker, S. Washietl, P. Kheradpour, J. Ernst, G. Jordan, E. Mauceli, L. D. Ward, C. B. Lowe, A. K. Holloway, M. Clamp, S. Gnerre, J. Alföldi, K. Beal, J. Chang, H. Clawson, J. Cuff, F. Di Palma, S. Fitzgerald, P. Flicek, M. Guttman, M. J. Hubisz, D. B. Jaffe, I. Jungreis, W. J. Kent, D. Kostka, M. Lara, A. L. Martins, T. Massingham, I. Moltke, B. J. Raney, M. D. Rasmussen, J. Robinson, A. Stark, A. J. Vilella, J. Wen, X. Xie, M. C. Zody, K. C. Worley, C. L. Kovar, D. M. Muzny, R. A. Gibbs, W. C. Warren, E. R. Mardis, G. M. Weinstock, R. K. Wilson, E. Birney, E. H. Margulies, J. Herrero, E. D. Green, D. Haussler, A. Siepel, N. Goldman, K. S. Pollard, J. S. Pedersen, E. S. Lander, M. Kellis, S. Lee, L. R. Lewis, L. V Nazareth, G. Okwuonu, J. Santibanez, W. C. Warren, E. R. Mardis, G. M. Weinstock, R. K. Wilson, Genome Institute at Washington University, K. Delehaunty, D. Dooling, C. Fronik, L. Fulton, B. Fulton, T. Graves, P. Minx, E. Sodergren, E. Birney, E. H. Margulies, J.

Herrero, E. D. Green, D. Haussler, A. Siepel, N. Goldman, K. S. Pollard, J. S. Pedersen, E. S. Lander, M. Kellis, A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*. **478**, 476–482 (2011).

15. K. M. Wong, M. A. Suchard, J. P. Huelsenbeck, Alignment Uncertainty and Genomic Analysis. *Science (80-.).* **319**, 473–476 (2008).
16. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
17. M. Bosse, H.-J. Megens, O. Madsen, Y. Paudel, L. A. F. Frantz, L. B. Schook, R. P. M. A. Crooijmans, M. A. M. Groenen, Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genet.* **8**, e1003100 (2012).
18. P. Dobrynin, S. Liu, G. Tamazian, Z. Xiong, A. A. Yurchenko, K. Krasheninnikova, S. Kliver, A. Schmidt-Küntzel, K.-P. Koepfli, W. Johnson, L. F. K. Kuderna, R. García-Pérez, M. de Manuel, R. Godinez, A. Komissarov, A. Makunin, V. Brukhin, W. Qiu, L. Zhou, F. Li, J. Yi, C. Driscoll, A. Antunes, T. K. Oleksyk, E. Eizirik, P. Perelman, M. Roelke, D. Wildt, M. Diekhans, T. Marques-Bonet, L. Marker, J. Bhak, J. Wang, G. Zhang, S. J. O'Brien, Genomic legacy of the African cheetah, Acinonyx jubatus. *Genome Biol.* **16**, 277 (2015).