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

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

Affiliations:

[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 author. Email: tom.vandervalk@ebc.uu.se (T.v.d.V.); katerina.guschanski@ebc.uu.se (K.G.)

Keywords: Genetic load, mutation load, endangered species, inbreeding, inbreeding depression, conservation, genetic rescue

Abstract: It is commonly thought that declining populations will experience negative genetic consequences as a result of increased inbreeding. Here we estimated the average deleteriousness of derived alleles in a range of mammals and found that species with historically small population size and low genetic diversity often have lower genetic load than species with large population sizes. This is likely the result of genetic purging – the more efficient removal of partially deleterious recessive alleles from inbred populations. Our findings suggest that genetic purging occurs over long evolutionary time frames, and therefore rapid population declines are likely to dis-proportionally increase mutational load in species with high diversity, as they carry many deleterious alleles that can reach fixation before genetic purging can remove them.

Main: Small inbred populations of wild animals frequently show lower survival, less efficient mating and lower reproductive success than large outbred populations (1) as a consequence of high levels of genome-wide homozygosity at loci with partially recessive deleterious alleles (2). Such negative fitness consequences of inbreeding have been directly confirmed on the genomic level (3–6). 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 (7). Whereas in large populations partially recessive deleterious alleles are mostly found at low frequency, these alleles can drift to high frequency in small populations (2). 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 over time (2). Although genetic purging has been shown to act in several animal populations (8–11), it remains largely unexplored to what extent this process represents a prominent evolutionary force. Wild animal populations across the globe experience rapid declines (12, 13), which are often followed by increased rates of inbreeding and the resulting genetic consequences can directly contribute to their extinction (14). 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 and prioritize them for conservation interventions.

Here, we used genomic data to estimate the prevalence of genetic purging in wild mammalian populations, as mammals are among the most affected by human-induced population declines (12). By calculating phylogenetic sequence conservation we aimed to identify genomic sites under strong evolutionary constraints, as mutations at these sites are expected to bear negative fitness consequences (15). 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 (16, 17). 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 selective constraints (18). Mutations at highly conserved genomic sites (those sites with high GERP-scores) are likely deleterious, whereas those at low GERP-scores are expected to be mostly neutral. We then used these GERP-scores to estimate individual relative genetic load in 655 individuals belonging to 41 different mammalian species, using publicly available whole genome re-sequencing data.

Results and Discussion

GERP scores indicate conserved genomic regions. We used a short-read mapping-based approach for the GERP-score calculations (see methods, SI Appendix, Fig. S1, S2) and validated its accuracy using five independent approaches. First, for the human genome, we obtained a high correlation between the GERP-scores previously calculated from a 44 whole-genome vertebrate alignment (18) and those obtained with our pipeline (Pearson correlation $r = 0.944$) (SI Appendix, Fig. S3). Second, our calculated GERP-scores are four to six times higher within exonic regions, known to be highly conserved and under purifying selection in vertebrates (19), than in intronic regions ($P < 2.2 \cdot 10^{-16}$) (SI Appendix, 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 mostly reflect neutrally evolving regions that are segregating within populations, whereas polymorphisms at high GERP-scores are mostly removed from the populations (SI Appendix, Fig. S5). Fourth, we observe that derived alleles at high GERP-scores are more often found in the heterozygous state, whereas at low GERP-scores they often appear in the homozygous state (SI Appendix, Fig. S6), suggesting that many derived alleles at high GERP-scores present in a population are likely to be recessive deleterious. Finally, GERP scores are on average higher at amino-acid changing sites that are expected to have a deleterious impact on the protein, as calculated with SIFT (20), compared to more tolerated amino-acid changing positions (SI Appendix, Fig. S7), providing an independent support for the likely negative effects of mutations in conserved parts of the genome. Taken together, these results are consistent with the expected dynamics of partially deleterious recessive alleles (21).

Population size and levels of inbreeding correlate with genetic load. Within species differences in genetic load are often estimated as the total number of putative deleterious alleles per genome (11, 17). However, when comparing different species, obtaining unbiased estimates of genetic load that are based on sums becomes challenging, as species differ from each other in genome size, mutation

rate and generation time and these factors lead to differences in the number of variable sites independent of genetic load. Thus, to allow comparisons across species, we calculated individual genetic load as the average GERP score of all derived alleles (Methods, Figure 1A). This measure is independent of the number of derived alleles identified in the genome, but rather reflects the distribution of conservation scores of the derived alleles. The rationale behind this approach is the following: As purifying selection acts on deleterious alleles, here identified as those at highly conserved sites, the distribution of derived alleles shifts towards lower GERP scores. The stronger the purifying selection, the stronger the shift of the distribution towards lower GERP scores. Thus the average GERP score of the derived alleles in a genome reflects the strength of purifying selection experienced by a given species over time. As we count all derived alleles in the genome, independent of zygosity, we implicitly assume co-dominance of these alleles (but see SI Appendix, Fig. S14 for different considerations of the dominance coefficient). Using this measure, we found that estimates of genetic load differ among the studied mammalian species, but are very similar for individuals belonging to the same species (Fig. 1A). The low inter-individual variation in genetic load (±SD 1.3%) suggests that this measure reflects long-term evolutionary processes (e.g. over hundreds of generations). We confirmed that our inferences are robust with respect to the outgroup species used to determine the ancestral state of alleles, as well as to the species used in the genome alignments for GERP-score calculations (SI Appendix, Fig. S8, S9, S10), suggesting that between-species differences are not driven by underlying technical artifacts.

We sought to explore whether measures of genetic load can be meaningfully applied in species conservation. Endangered species often have small population sizes, low genetic diversity and high inbreeding (14) and as a result are expected to carry high genetic load (1). Although we find that genetic load and species conservation status correlate, the relationship is weak (R=0.07, p=0.026, Fig. 1B) and goes in the opposite direction of the above expectation, with lower genetic load in species of greater conservation concern.

As conservation status does not correlate with genetic diversity and only weakly with population size (22), we investigated whether these measures may be explained by differences in genetic load among species. We observed a weak inverse relationship between genetic load and inbreeding (R=0.17, p=0.0099), calculated as proportion of the genome in runs of homozygosity (Fig. 1C). Many species with low genetic load, i.e. those with most derived alleles at only weakly conserved sites, have high proportions of their genome in runs of homozygosity (e.g. snow leopard, tiger, island fox, wolves, cheetah, Fig. 1C, SI Appendix, Dataset S1). Conversely, species with high genetic load frequently have

118 a low genome-wide rate of homozygosity (e.g. house mouse, brown rat, Himalayan rat, vervet
119 monkey, olive baboon, rhesus macaque, SI Appendix, Dataset S1). However, because in contrast to
120 intra-specific measures of genetic load, the level of inbreeding can change rapidly within only a few
121 generations, which is exemplified by the high degree of intra-specific variation in inbreeding (±SD 27%)
122 (SI Appendix, Dataset S1), the correlation between genetic load and inbreeding is only moderate (Fig.
123 1C).

124 The ratio of genetic load calculated for genomic regions outside and within runs of homozygosity also
125 differed across species (SI Appendix, Fig. S11). It was smaller in large outbred populations, signifying
126 higher genetic load within runs of homozygosity than outside (e.g. houses mouse, vervet monkey),
127 and larger in populations with a long history of inbreeding and long-term small population sizes,
128 showing lower genetic load within runs of homozygosity than outside (e.g. snow leopard) (SI Appendix,
129 Fig. S11). Overall, there was a significant correlation between this ratio and total genetic load in the
130 species, such that species with higher ratios showed lower genetic load (SI Appendix, Fig. S12). This
131 observation is consistent with the action of purging, where partially recessive deleterious alleles are
132 removed from the population through long-term inbreeding and exposure to purifying selection, since
133 within populations certain genomic regions are more frequently found in homozygosity than others
134 (23). In contrast, partially deleterious alleles persist in large populations as segregation load, as they
135 rarely and inconsistently appear in a homozygous state in any given individual.

136 Contrary to the notion that small populations have high genetic load (24), we observe a positive
137 relationship between genetic load and population size (Fig. 1D, SI Appendix, Dataset S1, R=0.35,
138 p=0.0085). Generally, species with small population size have lower genetic load than species with
139 large population sizes (Fig. 1D, SI Appendix, Dataset S1), suggesting that purging of deleterious alleles
140 can be an important evolutionary force. This relationship was only slightly affected by the dominance
141 coefficient of the derived alleles, with larger populations showing higher genetic load if the derived
142 alleles were considered dominant (SI Appendix, Fig. S13). In small populations, genetic load was
143 almost independent of the dominance coefficient, as most derived alleles are in a homozygous state
144 and thus exposed to selection. Despite generally lower load in small populations, we observe relatively
145 high genetic load in several species with historically large population sizes that have experienced
146 dramatic recent population declines (e.g. chimpanzees, orangutans, bonobos and Iberian lynx, Fig. 1D)
147 (25). This corroborates recent findings from genetic simulations, which demonstrate that strong
148 declines in population size disproportionally affect ancestrally large populations (26). Finally, after
149 controlling for the phylogenetic relationships among the studied species (27), the positive correlation
150 between population size and genetic load remains but is less pronounced (R=0.153, p = 0.023, SI

Appendix, Fig. S14). This is consistent with our earlier observation that changes in genetic load seem to occur over long evolutionary time-frames.

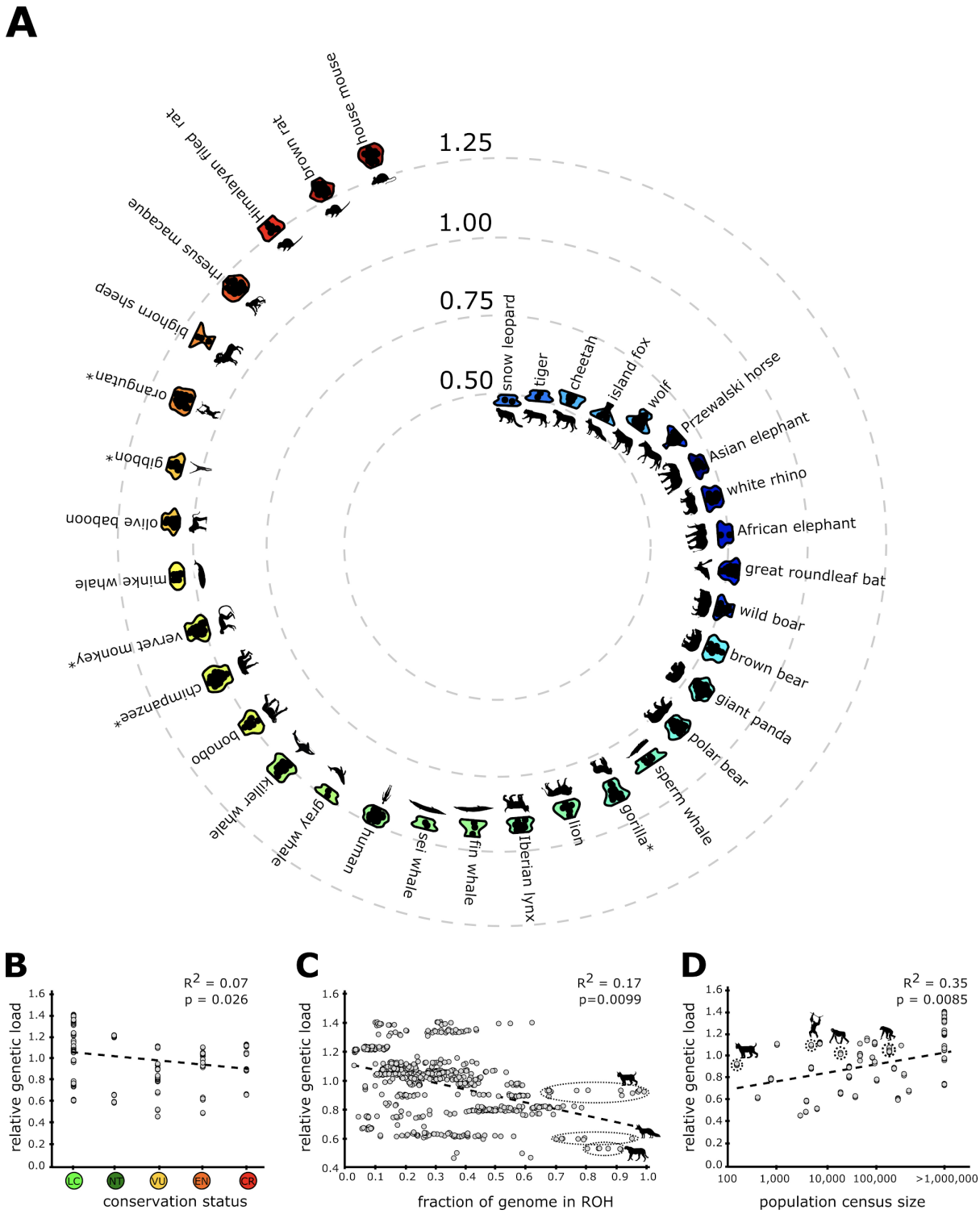

Fig. 1. Relative genetic load in mammals. *(**A**) Genetic load is depicted as the average GERP-score of the derived allele for each individual within a species. Several closely related species (Sumatran and Bornean orang-utans, gibbons, vervet monkeys, and eastern and western gorillas) are grouped together for clarity (marked by the asterisks). (**B**) Relative genetic load is not well explained by*

conservation status. LC: least concern, NT: near threatened, VU: vulnerable, EN: endangered, CR: critically endangered. ***(C)*** *Relative genetic load is negatively correlated with inbreeding. Species with recurrent bottlenecks and/or small population size show low load despite high inbreeding (e.g. cheetah and island fox). 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, chimpanzee, bonobo) show elevated genetic load given their population size. Each grey dot represents a species in (B) and (D) and an individual genome in (C). Dotted lines depict the best fitting linear intercept.*

Fixation of deleterious alleles. 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. 2a). Generally, we find that species with low genetic load carry few derived alleles at high GERP-scores (e.g. cheetah, island fox, Przewalski horse) (Fig. 2b), however, these alleles frequently appear to be fixed in the population (Fig. 2a). 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 a high proportion of putatively deleterious alleles at high GERP scores (Fig. 2b), the majority of these are at low frequency and less likely to appear in the homozygous state in any given individual (Fig. 2a). Thus, while purging removes strongly 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 (28, 29). Taken together, these observations are especially worrying for genetically diverse populations that experience rapid population declines, as we show that in species with high genetic diversity more derived alleles are found at high GERP scores and thus a high proportion of these putatively deleterious alleles could reach fixation before genetic purging can act (see also (24)).

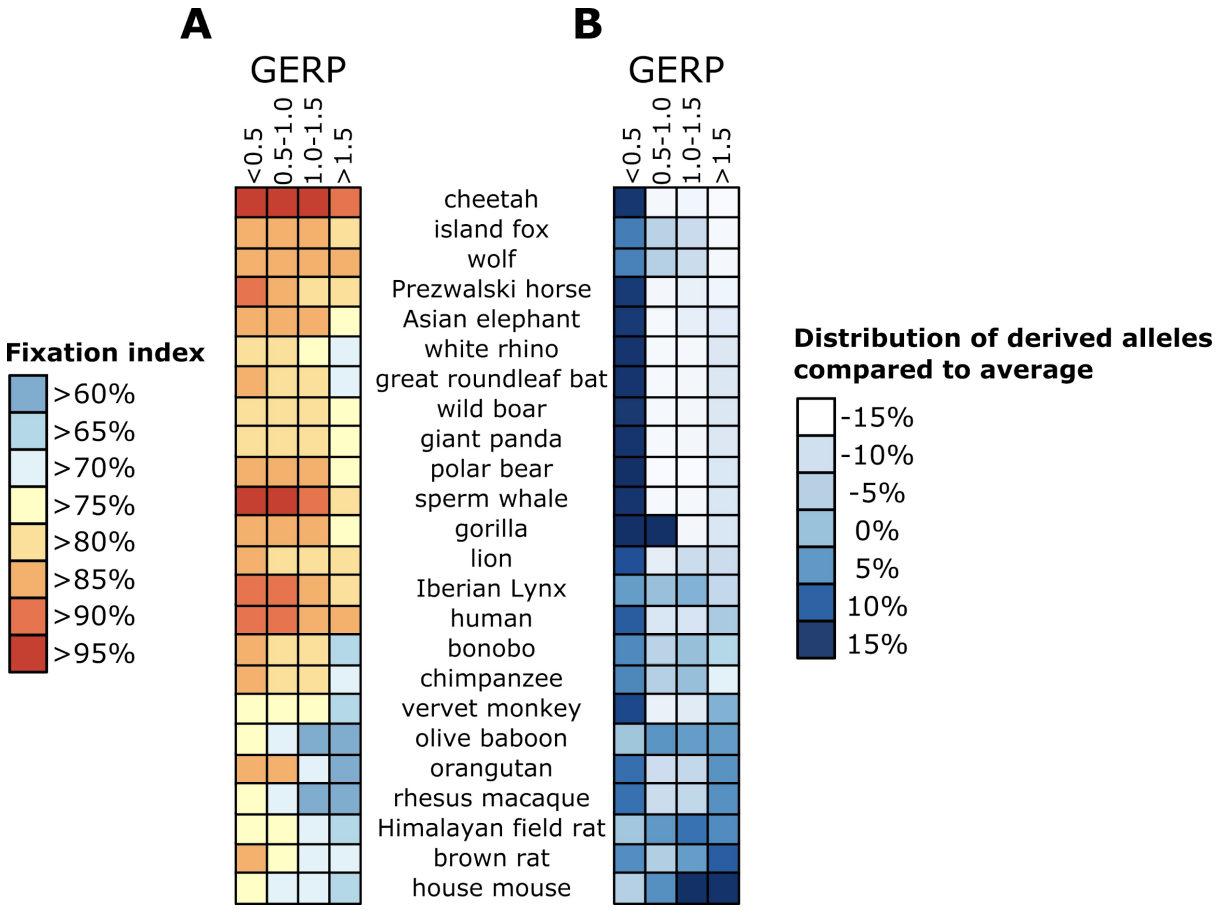

Fig. 2. Fixation of derived alleles. *Species are ranked by average GERP-score of the derived alleles (following Fig. 1A) from the lowest (at the top of the graph) to highest (bottom of the graph). (A) Percentage of derived alleles that are fixed within the species for a given GERP-score bin. E.g. most derived alleles in cheetah are fixed, even at high GERP-score, whereas in house mouse many derived alleles are segregating in the population. (B) Distribution of derived alleles in each species for each GERP-score bin compared to the average over all species. Darker colors represent relatively more derived alleles within the given GERP-score class. E.g. cheetahs have more derived alleles at low GERP-scores, whereas derived alleles in the house mouse are more frequent at high GERP-scores.*

Genetic implications for conservation of small populations. Our study describes a general trend of increasing genetic load with population size and outbreeding (Fig. 1c-d). This pattern allows the identification of populations and species that could represent potential targets for conservation interventions as those falling above the trendline and thus showing higher genetic load than would be expected given their level of inbreeding or their population size. For instance, individuals of the Iberian lynx show elevated genetic load given their level of inbreeding, which is in accordance with their history of strong recent bottlenecks. Similarly, multiple great ape species (orangutans, bonobos,

204 chimpanzees) show higher genetic load than would be expected given their current population size,
205 which is in line with their rapid population declines. These species and populations could thus
206 represent priorities for conservation interventions, as they appear to have a larger genetic load than
207 other species of comparable population size and might therefore be vulnerable to inbreeding
208 depression.

209 Deliberate purging by inducing inbreeding has previously been suggested as a management strategy
210 for reducing inbreeding depression (30). Current evidence for its efficiency is limited (21), but some
211 wild populations show evidence of successful purging. For instance, the Steward Island robins, birds
212 with a long-term small population size, show no correlation between inbreeding and juvenile survival
213 (31), whereas juvenile survival in the closely related Chatman Island robins, which went through a
214 drastic recent bottleneck, is negatively associated with levels of inbreeding (31). The long-term small
215 population size of the Stewart Island robin thus seems to have facilitated extensive purging, making
216 this species more "adapted" to small population size. Although this shows that deliberate purging
217 could potentially result in a positive outcome, several species with recent sharp population declines
218 show little signs of genetic purging (e.g. orangutan, chimpanzee, lions). Therefore purging seems to
219 be of limited efficacy during rapid population declines (and the consequently sharp rise in inbreeding).
220 Simulations also suggest that although inbreeding can be effective in purging lethal alleles within a
221 few generations, the removal of partially recessive deleterious alleles with smaller fitness effects
222 requires long-lasting and slow inbreeding (8, 32). As such, deliberate purging seems to be only an
223 effective strategy for short-term conservation if the aim is to remove (nearly) lethal alleles from a
224 population (33). In addition, experimental studies show that rapid serial bottlenecks can result in a
225 highly increased risk of extinction (34). Given all this evidence, it is conceivable that small inbred
226 populations that are alive today represent those species that have effectively purged deleterious
227 alleles, whereas rapid inbreeding has led to the extinction of many other populations. Finally,
228 deliberate inbreeding will result in loss of genetic diversity and thus limit future adaptive potential of
229 populations (35, 36). In management of captive breeding and restoration programmes, deliberate
230 inbreeding is thus a risky strategy with likely negative outcomes in most cases.

231 The higher genetic load of individuals from large populations represents an important aspect to be
232 considered for the commonly employed conservation strategy of genetic rescue. The main objective
233 of genetic rescue is to mitigate inbreeding depression by introducing genetic diversity and beneficial
234 variants through translocation of outbred individuals into an inbred population. Although genetic
235 rescue has been shown to increase population fitness in the short-term for a range of mammals (37,
236 38), the long-term effects can be negative. This is exemplified by the collapse of the Isle Royale wolves,
237 a population that maintained good population viability for decades. After interbreeding with a

238 mainland wolf migrant, the Isle Royale wolves initially showed higher reproductive success. However,
239 subsequent inbreeding in this population eventually resulted in the increase in frequency of
240 deleterious alleles, introduced by the immigrant, and eventual marked decline of the population (5).
241 A similar event occurred in the inbred Scandinavian wolf population following the arrival of a single
242 immigrant wolf, but subsequent additional immigration has resulted in additional outbreeding and a
243 period of recent population growth (39). Recurrent introductions of genetically diverse individuals into
244 the population was also proposed as an explanation as to why domestic chickens, in contrast to most
245 domesticated species, do not shower higher genetic load than their wild counterparts (40). Our
246 analyses demonstrate that individuals from large outbred populations often carry higher genetic load,
247 present as partially recessive segregating deleterious alleles. The introduction of an outbred individual
248 with an array of deleterious alleles into an inbred population that has experienced genetic purging
249 may thus have negative consequences. These consequences can possibly be avoided by selecting
250 individuals with low genetic load to be used for genetic rescue (26). However, if not followed by a
251 clearly delineated long-term plan to reduce inbreeding, for instance through repeated introductions,
252 the addition of strongly deleterious alleles into a small population with low capacity for long-term
253 purifying selection and purging may lead to fixation of these alleles with detrimental effects for
254 population survival.

255

256 **Methods**

257 **Single nucleotide variant calling.** We obtained published re-sequencing data for 655 mammalian
258 genomes from 41 species and mapped these to the phylogenetically closest available reference
259 genome for each species (SI Appendix, Dataset S1) using bwa mem v0.7.17 (41). In total, 27 reference
260 genomes were used for this task (SI Appendix, Dataset S1). We then obtained and filtered variant calls
261 for each individual using GATK HaplotypeCaller v3.8 following the "short variant discovery best-
262 practices guidelines" including "hard filtering" (42). Additionally, we only kept within-species bi-allelic
263 sites and removed all indels and sites below one third and above three times the genome-wide
264 autosomal coverage (43).
265 **Genomic Evolutionary Rate Profiling.** We used the software GERP++ (Genomic Evolutionary Rate
266 Profiling) to calculate the number of "rejected substitutions" (a proxy for evolutionary constraints) for
267 each site in the same 27 reference genomes that were used in mapping of the re-sequencing data (SI
268 Appendix, Dataset S1, SI Appendix, Fig. S1) (18). GERP++ estimates the number of substitutions that
269 would have occurred if the site was neutral given a multi-species sequence alignment and the
270 divergence time estimates between the aligned species as provided in (44). A GERP-score, the number

271 of rejected substitutions at a genomic site, is thus a measure of constraint that reflects the strength
272 of past purifying selection at a particular locus. To calculate GERP-scores for a given focal reference
273 genome, we used 100 non-domesticated mammalian de-novo assembled genomes (SI Appendix,
274 Dataset S2, SI Appendix, Fig. S2), as domesticated species might give a biased estimate of purifying
275 selection. Each individual genome sequence was converted into short FASTQ reads by sliding across
276 the genome in non-overlapping windows of 50 base pairs and transforming each window into a
277 separate FASTQ read. The resulting FASTQ reads from the 100 mammalian genomes were then
278 mapped to each respective focal reference genome with bwa mem v0.7.17, slightly lowering the
279 mismatch penalty (-B 3) and removing reads that mapped to multiple regions. Mapped reads were
280 realigned around indels using GATK IndelRealigner (45, 46). Next, we converted the mapped reads
281 into a haploid FASTA consensus sequence (i.e. 100 times for each reference genome), excluding all
282 sites with depth above one (as such sites contain at least one mismapped read). GERP++ was then
283 used to calculate the number of rejected substitutions at all sites in the reference using the
284 concatenated FASTA files and the species divergence time estimates from (44) (SI Appendix, Fig. S2),
285 excluding the focal reference from the calculation. Missing bases within the concatenated alignment
286 were treated as non-conserved (i.e. sites for which only few reads mapped obtain low GERP scores).
287 We excluded all sites for which the focal reference FASTQ reads did not map to themselves and sites
288 with negative GERP-scores (as these most likely represent errors) and subsequently scaled all scores
289 to a range from 0 to 2. Sites that are identical between species and have thus been preserved over
290 long evolutionary time result in high GERP-scores (SI Appendix, Fig. S1). Thus, high GERP-scores are
291 only obtained for regions, where the majority of the 99 mammalian genomes (100 minus the focal
292 reference) map to the respective reference.
293 **Effects of phylogeny on estimated GERP scores.** To evaluate whether GERP score estimates could be
294 affected by the availability or lack of sequenced genomes from closely related species (e.g. primates
295 with many available genomic sequences from close relatives versus Proboscidea for which most of the
296 sequenced species are evolutionary distant), we simulated a tree with the same number of species
297 (N=40) but different evolutionary distances between the focal species and all other species compared
298 (SI Appendix, Fig. S10). We calculated the GERP score for the human genome for three phylogenetic
299 trees that differed in the average evolutionary distance between the focal genome (human) and all
300 other species. Mean evolutionary distance ranged from 33.6 million years (SI Appendix, Fig. S10 Tree
301 1 containing only primates and two outgroups) to 57.3 million years (SI Appendix, Fig. S10 Tree 2) and
302 72 million years (SI Appendix, Fig. S10 Tree 3). We observed a close similarity and strong correlation
303 between estimated GERP scores independent of the evolutionary distance separating the focal species
304 from other species in the phylogenetic tree, suggesting that any biases arising during the calculation

305 of GERP scores between densely and sparsely represented parts of the mammalian tree of life are
306 likely to be minimal.
307 **Ancestral allele inference.** We called the ancestral allele at each site as the variant present in the
308 phylogenetically closest outgroup. The derived alleles in each individual from the study dataset was
309 then inferred against the called ancestral allele. By using only one outgroup we retained the highest
310 number of sites to be analysed (the more outgroups are added, the fewer sites will be mapped across
311 all outgroups). We estimated the effect of using one or multiple outgroups for the ancestral allele
312 inference by calling the majority allele among the mapped reads for 1, 2 and 3 outgroups (a random
313 base was choses if the allele frequency was equal) and show that this does not significantly change
314 the estimates of genetic load (SI Appendix, Fig. S8, S9), as genomic sites with high GERP-scores are
315 generally conserved and thus identical among all outgroup species (SI Appendix, Fig. S8).
316 **Relative genetic load.** We estimated relative genetic load for each of the genomes as the average
317 GERP-score of all derived alleles:

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

318
319 Where D_i represents the ith derived allele and $gerp_i$ the GERP-score for the ith allele. We calculated
320 averages to allow comparisons across species that differ in genome size and genome quality, as both
321 these factors will affect the more frequently used genetic load estimates that sum across all
322 deleterious mutations in the genome (11, 17). Whereas the score based on sums is highly appropriate
323 for comparisons within a single species or across closely related species where variation in genome
324 size and quality is negligible, the estimate based on means reflects the overall distribution of the
325 mutations without the need to identify every single one of them. Under the assumption that new
326 mutations occur randomly with respect to the genomic region, we expect that in species that
327 experienced strong purifying selection, derived alleles are found mostly at non-conserved sites (low
328 GERP-scores), whereas accumulation of deleterious variants should result in a higher fraction of
329 derived alleles at high GERP-scores.
330 **Calculation of SIFT scores.** SIFT-scores are a measure of the likelihood that an amino-acid change
331 causes disruption of the protein function (20). We used SIFT 4G to identify all amino-acid changes
332 classified as synonymous, deleterious and tolerated in the human, wolf and mouse genomes (as these
333 have high quality genome annotation available) and calculated the distribution of GERP-scores among
334 those three categories to test for correspondence between the two methods.

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 (47) 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 (48, 49). 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.

Phylogenetic independent contrast. We used the Phylogenetic independent contrast (PIC) (27) method from the R "ape" library to run a linear model on the correlations between population census size and genetic load, correcting for the fact that some species share a longer evolutionary history than others using the phylogenetic tree and species divergence times obtained from TimeTree (44)

Acknowledgments: TvdV acknowledges support of a scholarship from the Foundation for Zoological Research. TMB is supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 864203), BFU2017-86471-P (MINECO/FEDER, UE), "Unidad de Excelencia María de Maeztu", funded by the AEI (CEX2018-000792-M), 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). We acknowledge support from the Uppsala Multidisciplinary Centre for Advanced Computational Science for access to the UPPMAX computational infrastructure.

Author contributions: TvdV and MdM conceived the study and analysed the data. All authors interpreted the results. KG and TvdV wrote the manuscript with input from all authors.

Data and materials availability: All data used in this study is available on the European Nucleotide archive (accession numbers are given in SI Appendix, Dataset S1). Scripts used in this study are available on github/tvdvalk.

367 **References:**

368 1. L. F. Keller, D. M. Waller, Inbreeding effects in wild populations. *Trends Ecol. Evol.* **17**, 230–241
369 (2002).

370 2. D. Charlesworth, J. H. Willis, The genetics of inbreeding depression. *Nat. Rev. Genet.* **10**, 783–
371 796 (2009).

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

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

377 5. J. A. Robinson, *et al.*, Genomic signatures of extensive inbreeding in Isle Royale wolves, a
378 population on the threshold of extinction. *Sci Adv* **5**, eaau0757 (2019).

379 6. J. Prado-Martinez, *et al.*, The genome sequencing of an albino Western lowland gorilla reveals
380 inbreeding in the wild. *BMC Genomics* **14**, 363 (2013).

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

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

386 9. Y. Xue, *et al.*, Mountain gorilla genomes reveal the impact of long-term population decline and
387 inbreeding. *Science* **348**, 242–245 (2015).

388 10. A. Benazzo, *et al.*, Survival and divergence in a small group: The extraordinary genomic history
389 of the endangered Apennine brown bear stragglers. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E9589–
390 E9597 (2017).

391 11. C. Grossen, F. Guillaume, L. F. Keller, D. Croll, Purging of highly deleterious mutations through
392 severe bottlenecks in Alpine ibex. *Nat. Commun.* **11**, 1001 (2020).

393 12. S. L. Pimm, *et al.*, The biodiversity of species and their rates of extinction, distribution, and
394 protection. *Science* **344**, 1246752 (2014).

395 13. G. Ceballos, P. R. Ehrlich, P. H. Raven, Vertebrates on the brink as indicators of biological
396 annihilation and the sixth mass extinction. *Proc. Natl. Acad. Sci. U. S. A.* (2020)
397 https:/doi.org/10.1073/pnas.1922686117.

398 14. D. Spielman, B. W. Brook, R. Frankham, Most species are not driven to extinction before genetic
399 factors impact them. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15261–15264 (2004).

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

402 16. P. Librado, *et al.*, Ancient genomic changes associated with domestication of the horse. *Science*
403 **356**, 442–445 (2017).

17. B. M. Henn, *et al.*, Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E440–9 (2016).

18. E. V. Davydov, *et al.*, Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* **6**, e1001025 (2010).

19. K. Lindblad-Toh, *et al.*, A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).

20. R. Vaser, S. Adusumalli, S. N. Leng, M. Sikic, P. C. Ng, SIFT missense predictions for genomes. *Nat. Protoc.* **11**, 1–9 (2016).

21. P. W. Hedrick, A. Garcia-Dorado, Understanding Inbreeding Depression, Purging, and Genetic Rescue. *Trends Ecol. Evol.* **31**, 940–952 (2016).

22. D. Díez-Del-Molino, F. Sánchez-Barreiro, I. Barnes, M. T. P. Gilbert, L. Dalén, Quantifying Temporal Genomic Erosion in Endangered Species. *Trends Ecol. Evol.* **33**, 176–185 (2018).

23. T. J. Pemberton, *et al.*, Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.* **91**, 275–292 (2012).

24. D. H. Reed, R. Frankham, Correlation between Fitness and Genetic Diversity. *Conserv. Biol.* **17**, 230–237 (2003).

25. , The IUCN Red List of Threatened Species (2019).

26. C. Kyriazis, R. K. Wayne, K. E. Lohmueller, High genetic diversity can contribute to extinction in small populations. *bioRxiv* (2019) https:/doi.org/10.1101/678524.

27. J. Felsenstein, Phylogenies and the Comparative Method. *Am. Nat.* **125**, 1–15 (1985).

28. D. E. Wildt, *et al.*, Unique seminal quality in the South African cheetah and a comparative evaluation in the domestic cat. *Biol. Reprod.* **29**, 1019–1025 (1983).

29. M. Åkesson, *et al.*, Genetic rescue in a severely inbred wolf population. *Mol. Ecol.* **25**, 4745–4756 (2016).

30. M. Á. R. de Cara, B. Villanueva, M. Á. Toro, J. Fernández, Purging deleterious mutations in conservation programmes: combining optimal contributions with inbred matings. *Heredity* **110**, 530–537 (2013).

31. E. S. Kennedy, C. E. Grueber, R. P. Duncan, I. G. Jamieson, Severe inbreeding depression and no evidence of purging in an extremely inbred wild species—the Chatham Island black robin. *Evolution* **68**, 987–995 (2014).

32. A. García-Dorado, On the consequences of ignoring purging on genetic recommendations for minimum viable population rules. *Heredity* **115**, 185–187 (2015).

33. A. García-Dorado, Understanding and predicting the fitness decline of shrunk populations: inbreeding, purging, mutation, and standard selection. *Genetics* **190**, 1461–1476 (2012).

34. B. D. Griffen, J. M. Drake, A review of extinction in experimental populations. *Journal of Animal Ecology* **77**, 1274–1287 (2008).

35. R. Lande, S. Shannon, The Role of Genetic Variation in Adaptation and Population Persistence in

441 a Changing Environment. *Evolution* **50**, 434 (1996).

442 36. K. F. Smith, D. F. Sax, K. D. Lafferty, Evidence for the role of infectious disease in species
443 extinction and endangerment. *Conserv. Biol.* **20**, 1349–1357 (2006).

444 37. R. Frankham, Genetic rescue of small inbred populations: meta-analysis reveals large and
445 consistent benefits of gene flow. *Molecular Ecology* **24**, 2610–2618 (2015).

446 38. J. T. Hogg, S. H. Forbes, B. M. Steele, G. Luikart, Genetic rescue of an insular population of large
447 mammals. *Proceedings of the Royal Society B: Biological Sciences* **273**, 1491–1499 (2006).

448 39. M. Kardos, *et al.*, Genomic consequences of intensive inbreeding in an isolated wolf population.
449 *Nature Ecology & Evolution* **2**, 124–131 (2018).

450 40. S. Qanbari, *et al.*, Genetics of adaptation in modern chicken. *PLoS Genet.* **15**, e1007989 (2019).

451 41. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-*
452 *bio.GN]* (2013).

453 42. G. A. Van der Auwera, *et al.*, From FastQ data to high-confidence variant calls: the genome
454 analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11–10 (2013).

455 43. H. Li, Toward better understanding of artifacts in variant calling from high-coverage samples.
456 *Bioinformatics* **30**, 2843–2851 (2014).

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

459 45. M. A. DePristo, *et al.*, A framework for variation discovery and genotyping using next-
460 generation DNA sequencing data. *Nature Genetics* **43**, 491–498 (2011).

461 46. A. McKenna, *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-
462 generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

463 47. S. Purcell, *et al.*, PLINK: A Tool Set for Whole-Genome Association and Population-Based
464 Linkage Analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).

465 48. M. Bosse, *et al.*, Regions of homozygosity in the porcine genome: consequence of demography
466 and the recombination landscape. *PLoS Genet.* **8**, e1003100 (2012).

467 49. P. Dobrynin, *et al.*, Genomic legacy of the African cheetah, Acinonyx jubatus. *Genome Biology*
468 **16** (2015).

Supplementary Materials for

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

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

Correspondence to: tom.vandervalk@ebc.uu.se (T.v.d.V.); katerina.guschanski@ebc.uu.se (K.G.)

This PDF file includes:

Figs. S1 to S14

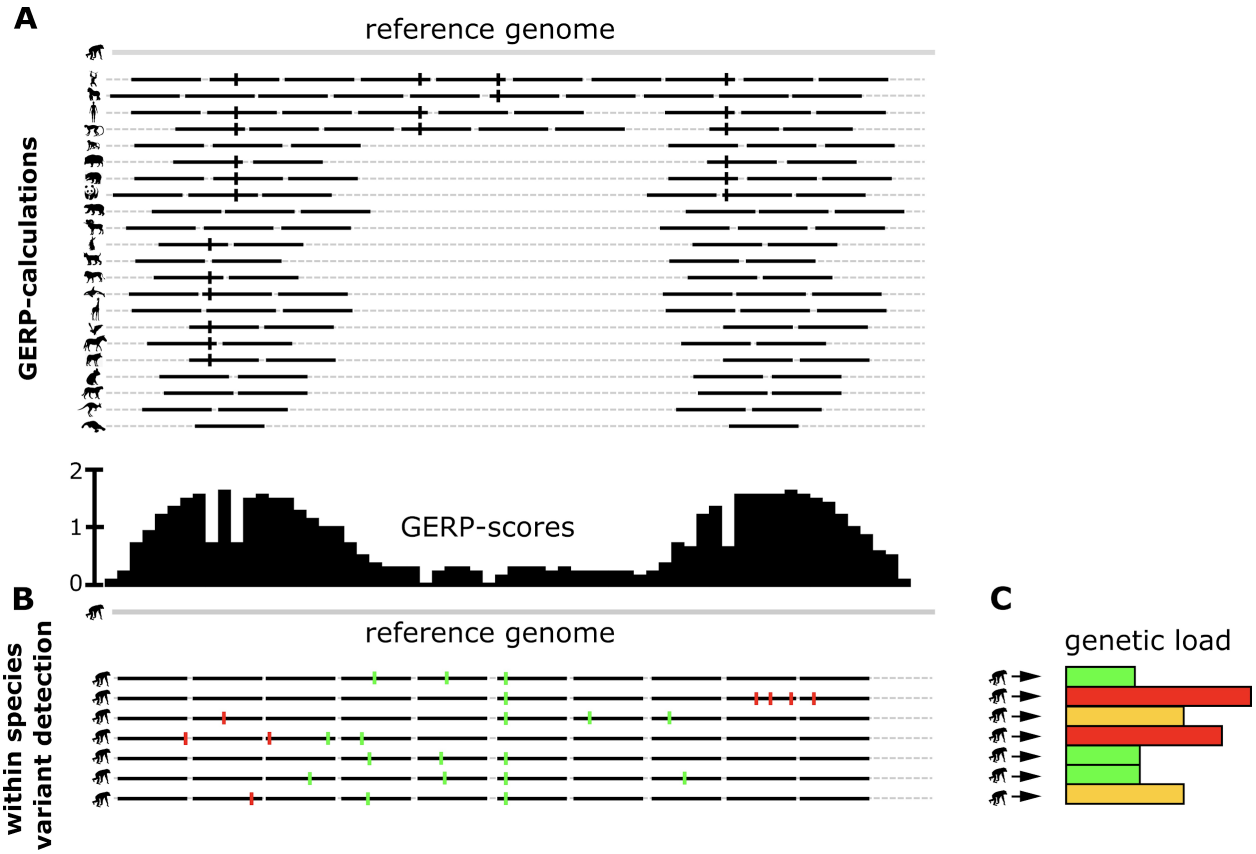

Fig. S1. Schematic representation of the GERP-score pipeline. (A) A set of 100 de-novo assembled genomes is sliced into non-overlapping 50 base pair FASTQ read and aligned to the same reference as used for the within-species variance detection (SNP calling in (B)). A consensus sequence is then obtained 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 the 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 used as a measure of the genetic load carried by each individual (red=high genetic load, orange=intermediate genetic load, green=low genetic load).

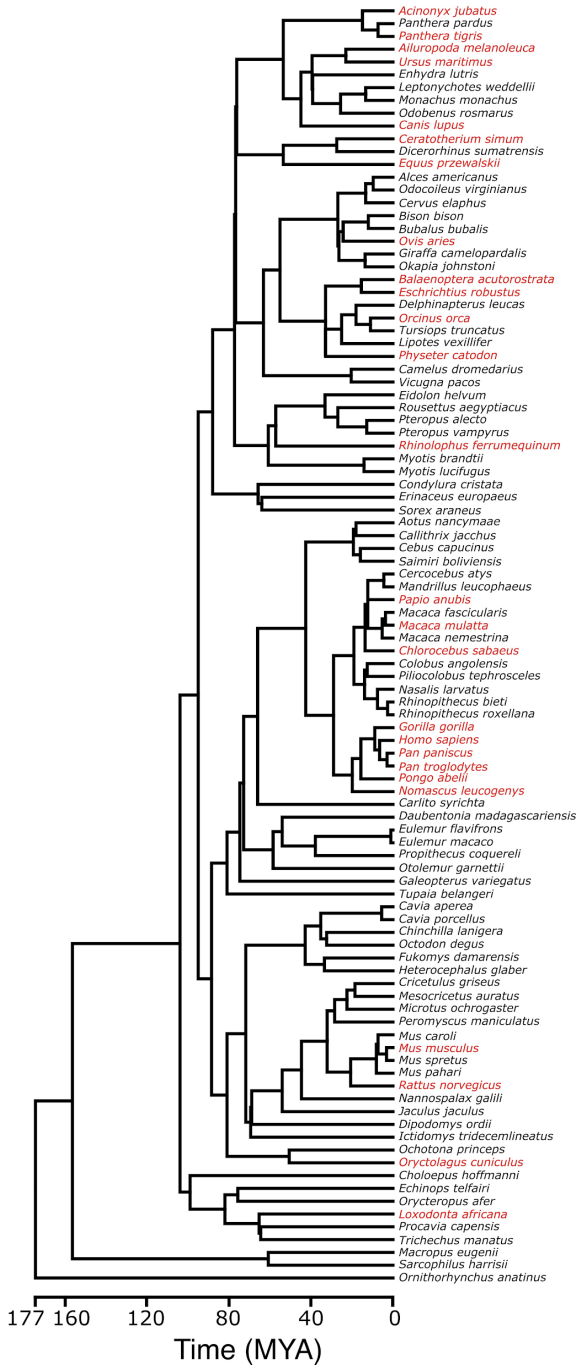

Fig. S2. The 100 mammalian genomes and their divergence times estimates used for the GERP-score calculations. The divergence times between the species were obtained using the online software TimeTree, which gives a dated phylogeny from a list of species through automated literature searches (5). The genomes depicted in red were also used for the mapping of re-sequencing population-level data. Note, there are fewer such genomes than the number of studied species, as for species without a reference genome we used the closest available relative for mapping.

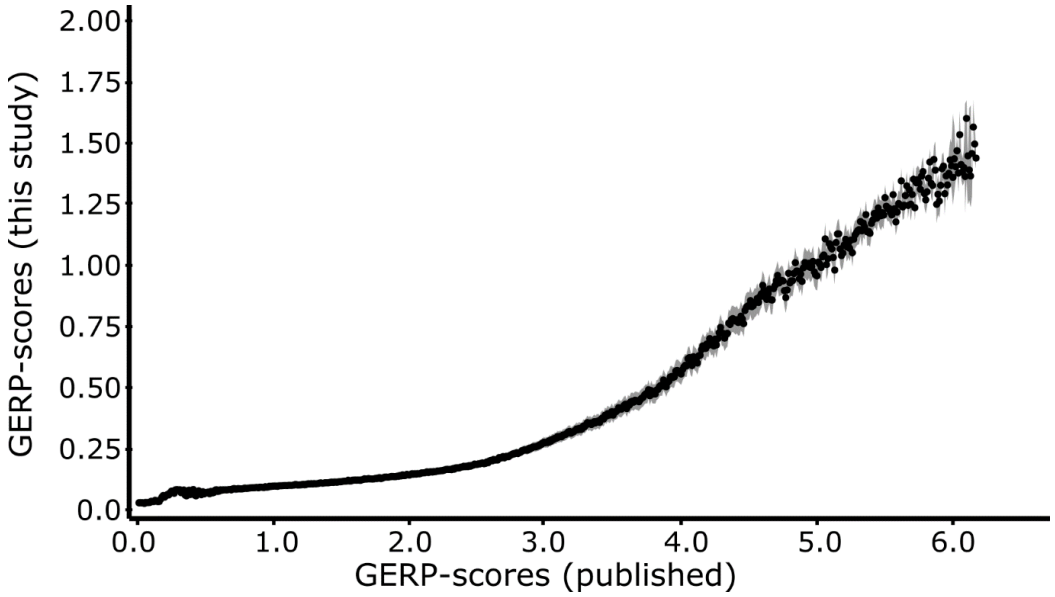

Fig. S3. Correlation between previously published alignment-based GERP-scores and the GERP-scores calculated with the mapping-based approach in this study for the human genome (hg19). We binned all sites in the human genome by their published GERP-scores and calculated the average GERP-score for each bin of size 10 (black dots). Grey shaded area depicts ±1SD. Pearson correlation = 0.944, Spearman's rank correlation = 0.997, p = $P < 2.2 \cdot 10^{-16}$. 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.

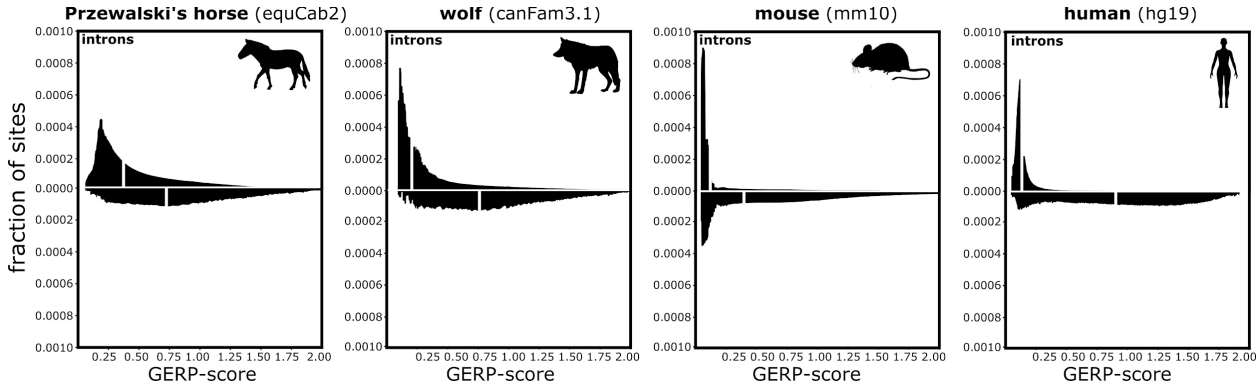

Fig. S4. GERP-score of genomic partitions. The distribution of GERP-scores within introns (top) and exons (bottom) for 4 species with available high-quality reference genome annotations (used references between brackets). White lines within the plots depict the average GERP-score for a given genomic 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), which is consistent with stronger sequence conservation and selective constraints in exons compared on introns.

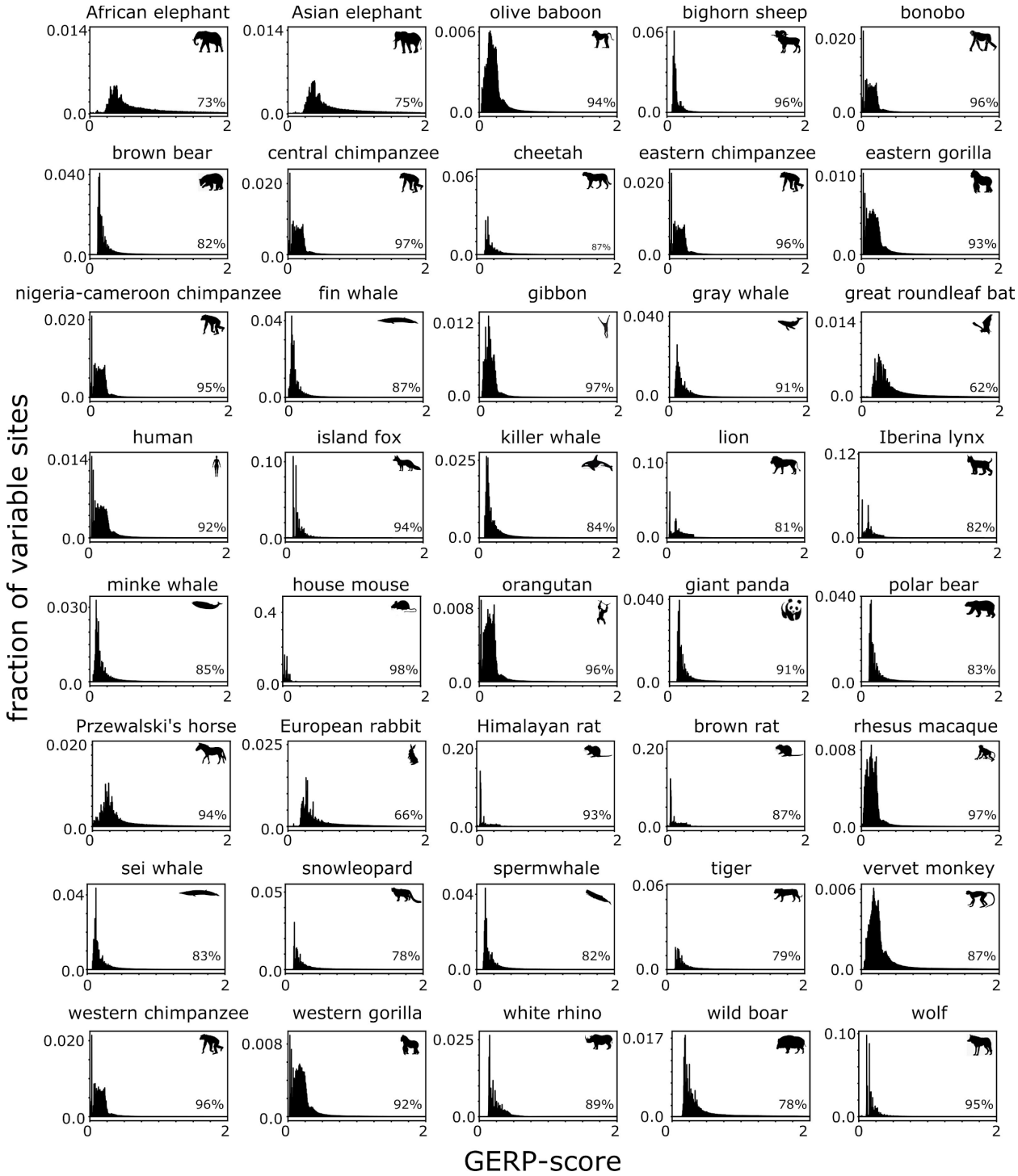

Fig. S5. Distribution of segregating sites by GERP-score. The proportion of variable sites plotted against GERP scores, with the percentage found in the 10% lowest GERP-scores depicted in the bottom right corner. A large majority of variable sites are found at low GERP-scores and thus within fast evolving, likely neutral, regions.

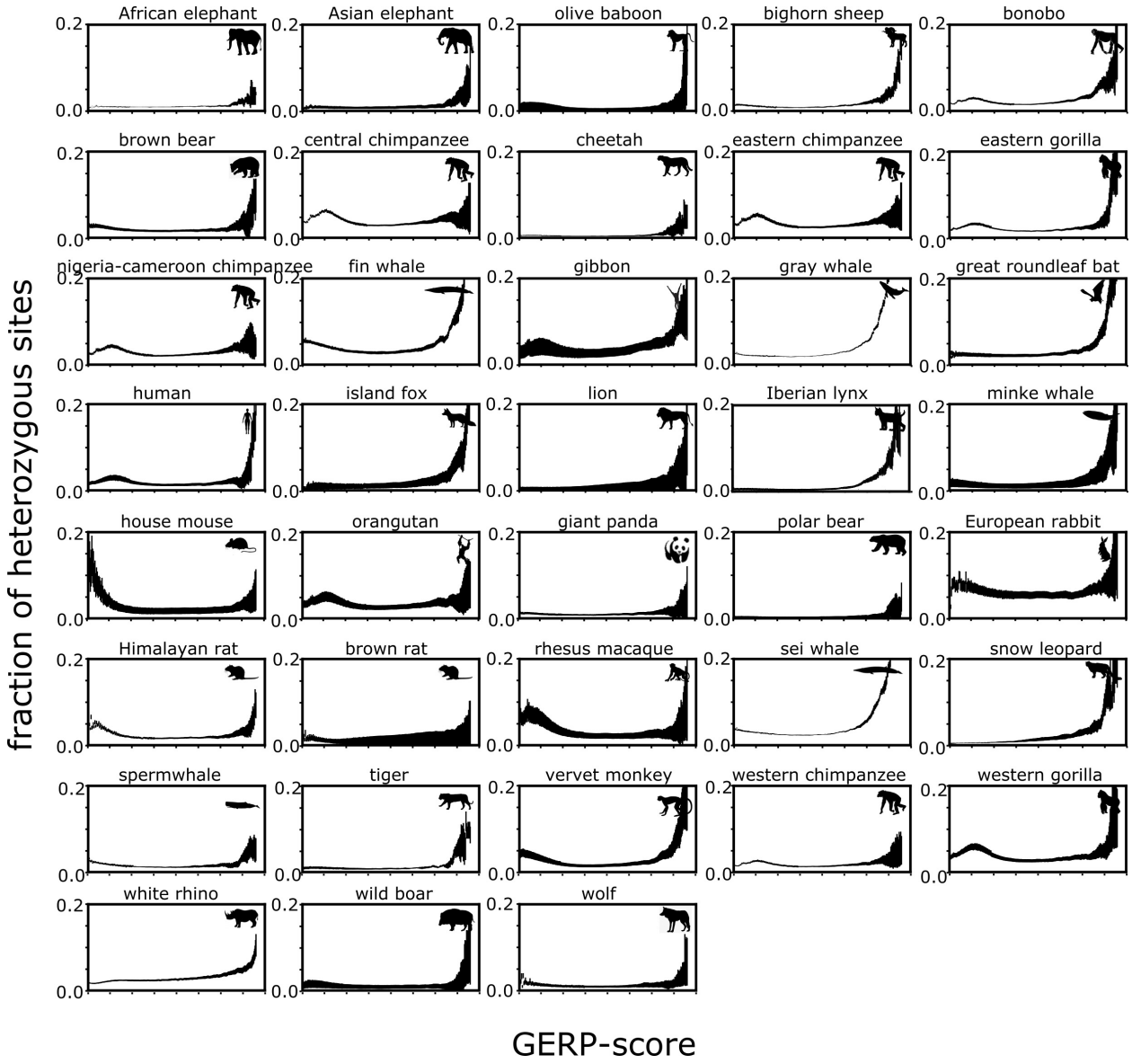

Fig. S6. Proportion of heterozygous alleles out of total sites stratified by GERP-score. We included only samples with average genome wide coverage > 10X. X-axis is scaled from 0 to 2 for all species. The proportion of heterozygous derived alleles increases at high GERP-scores, consistent with them representing segregation load.

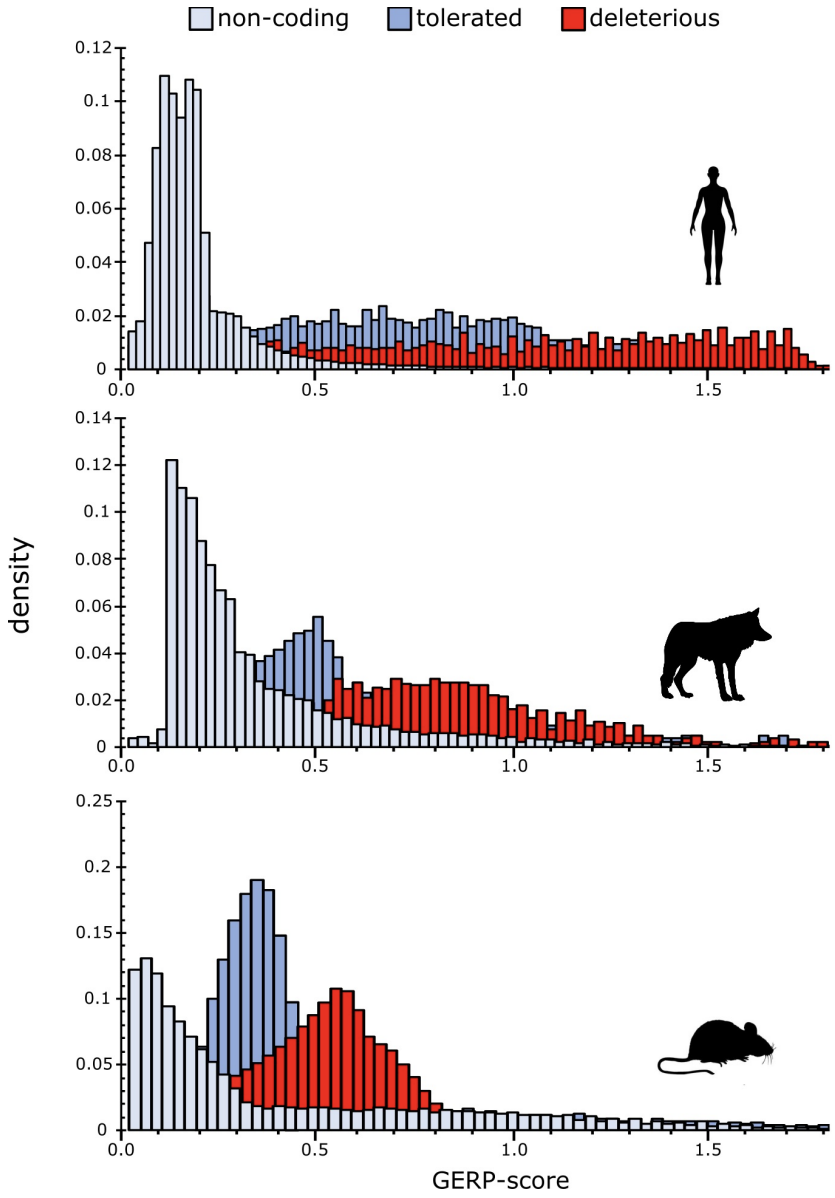

Fig. S7. Correlation between SIFT and GERP-scores. Derived alleles in three phylogenetically distinct species with well-annotated genomes were classified with SIFT into non-coding, tolerated and deleterious variants. Generally, deleterious alleles are in genomic regions with higher GERP-scores than tolerated and non-coding alleles, providing independent confirmation for putatively deleterious effects of variants with high GERP-scores across mammals.

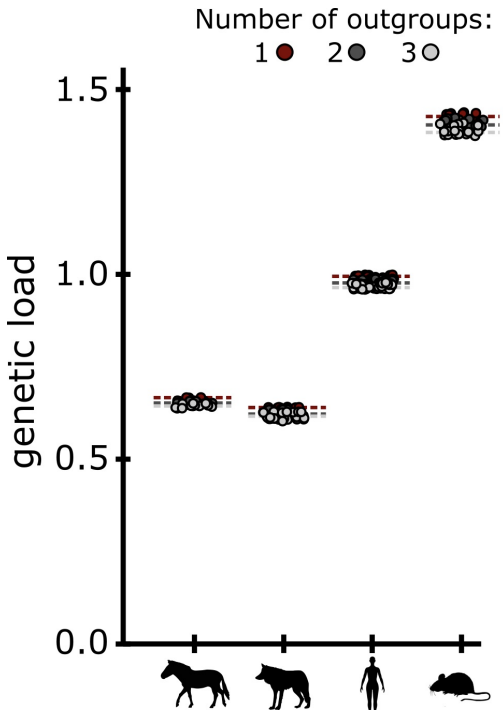

Fig. S8. Relative genetic load calculated as the average GERP-score of the derived alleles using a different number of outgroups. We inferred the derived state by either using one outgroup or the majority allele among 2 or 3 outgroups. We then re-calculated the genetic load for a phylogenetically diverse group of species (the Przewalski's horse, wolf, human and house mouse). Circles represent individual estimates and dotted lines depict the population averages for the different number of used outgroups. Although the population averages are higher when using only one outgroup, the differences are minimal and consistent across species.

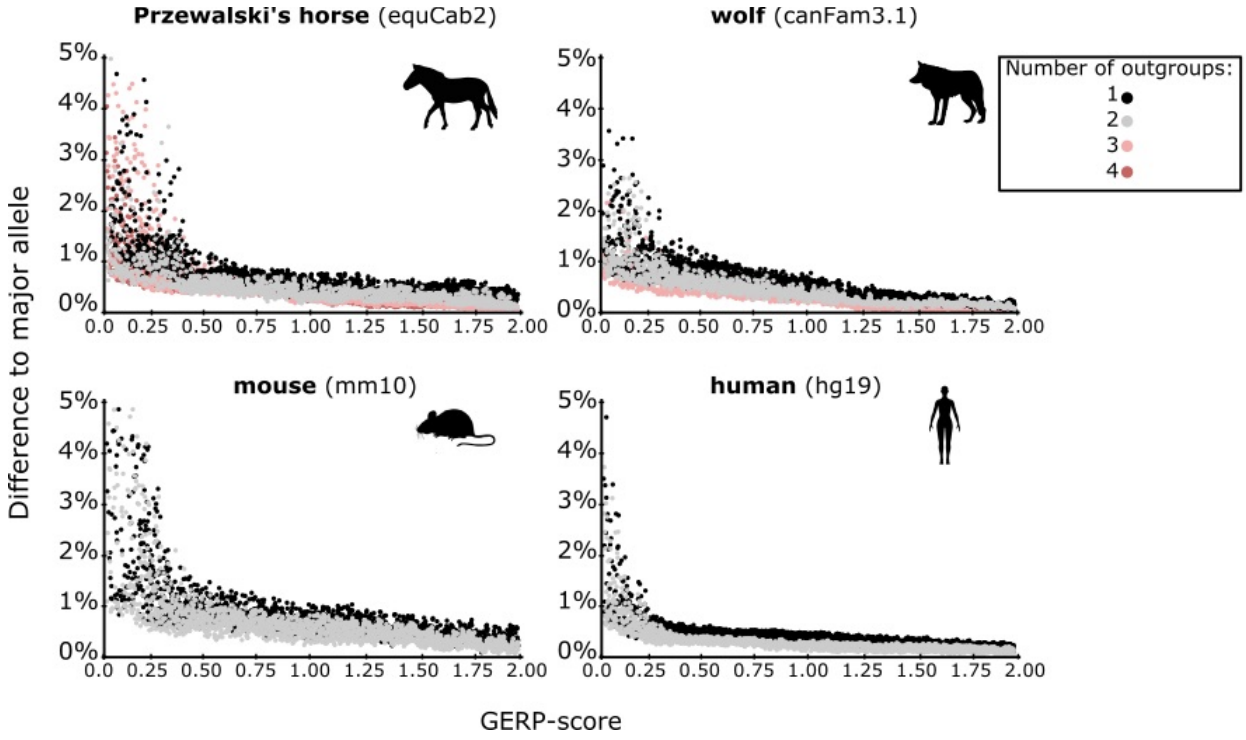

Fig. S9. Ancestral allele inference depending on the number of used outgroups. Plots show the percentage of nucleotide differences to the major allele (among the complete phylogeny, e.g. all species that mapped to the site) by GERP-score depending on the number of outgroups used to infer the ancestral 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.

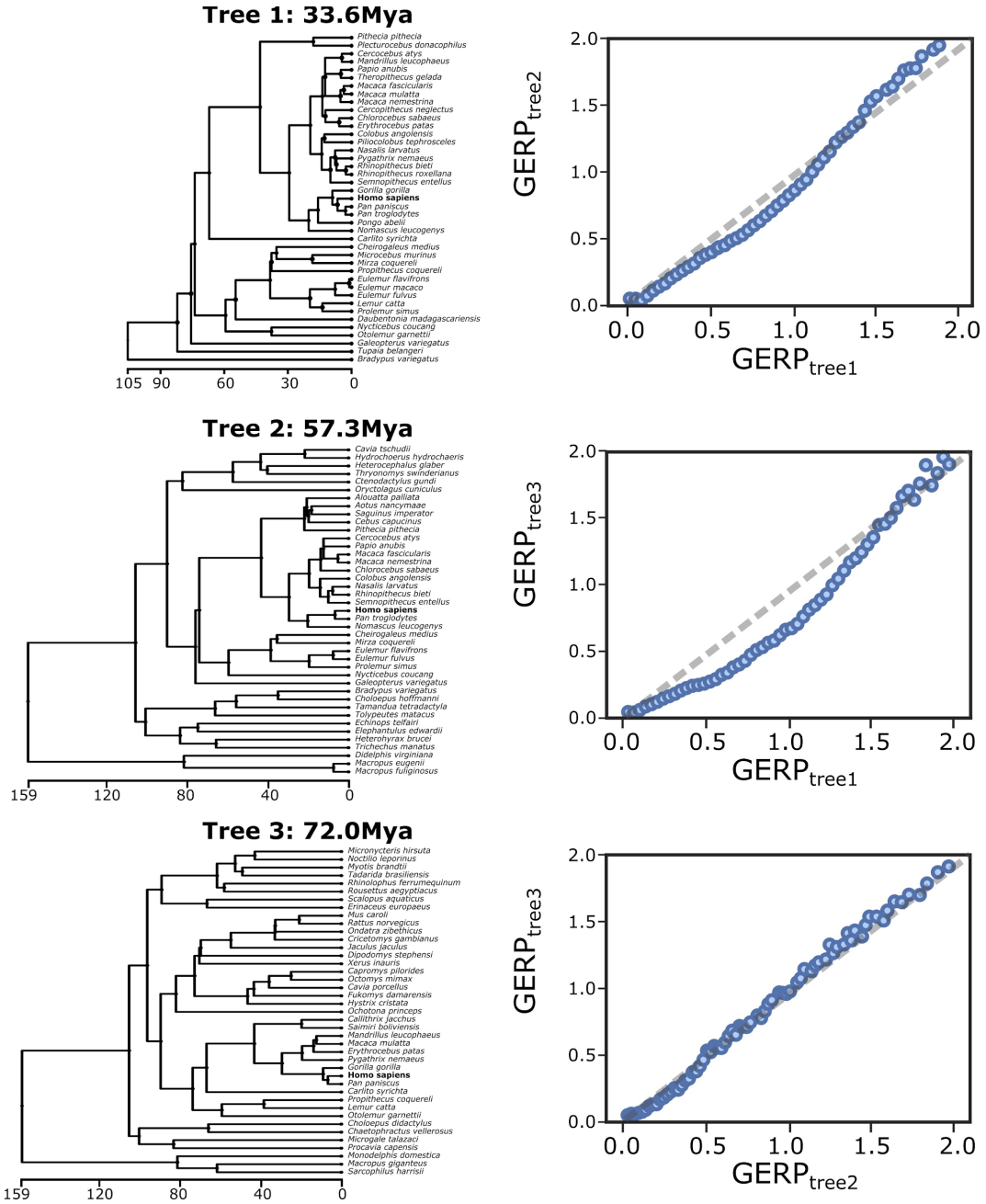

Fig. S10. Correlation between GERP-scores calculated for the human genome (highlighted in bold) based on different species trees. The average divergence of the reference species to the human genome is depicted above each tree. The dotted grey line in the right-hand graphs depicts parity. GERP-scores remain similar despite the use of different reference species.

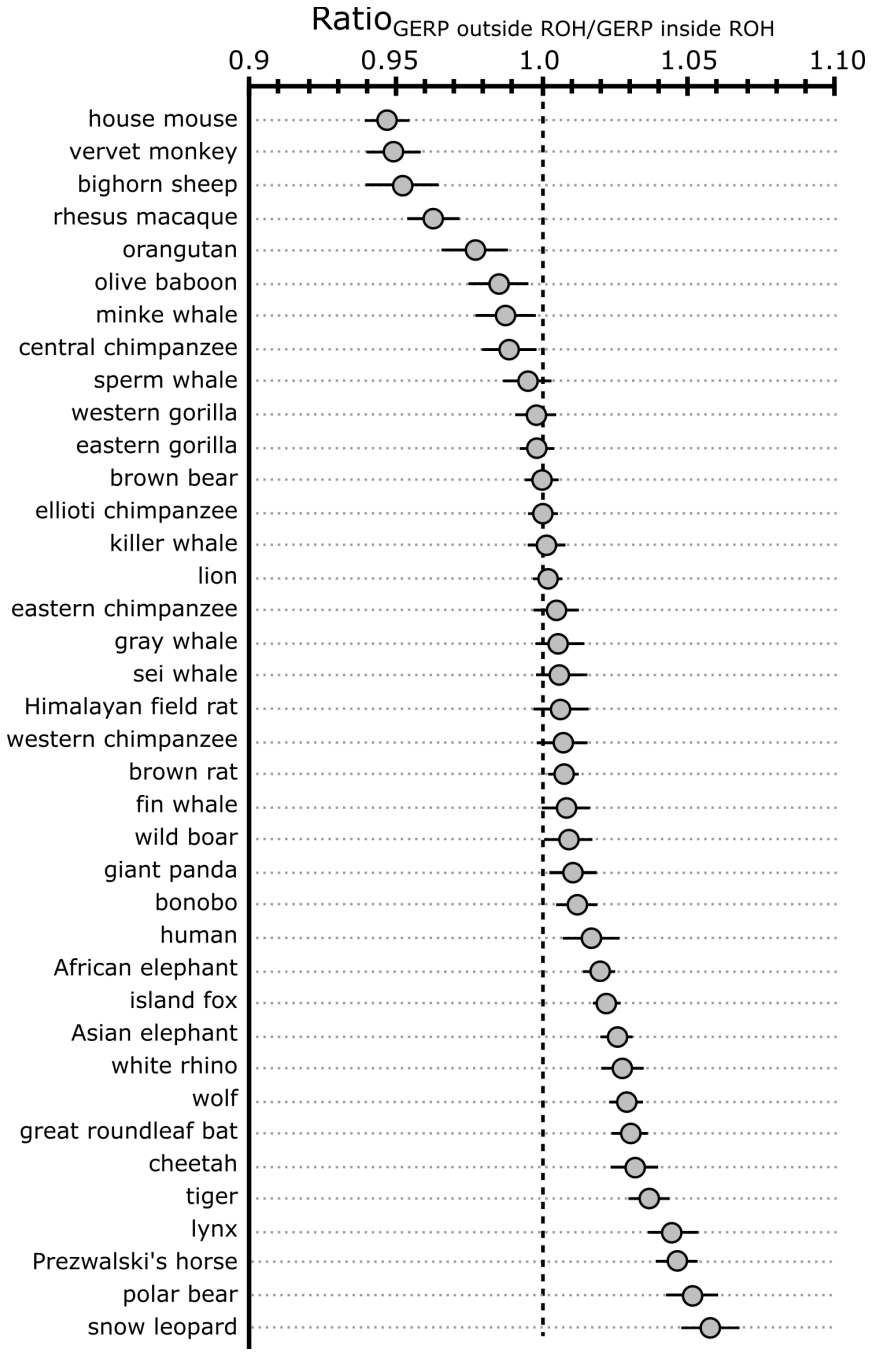

Fig. S11. Ratio between estimates of genetic load within and outside of runs of homozygosity. A ratio below 1 corresponds to higher genetic load within the runs of homozygosity than outside of them, ratio above 1 corresponds to higher genetic load outside the runs of homozygosity compared to within. Bars show inter-individual variation, the position of the grey circle corresponds to the mean value per species.

Species are ordered by increasing ratio. **Ratios were only calculated for species with high coverage genomes available in order to accurately infer ROHs.*

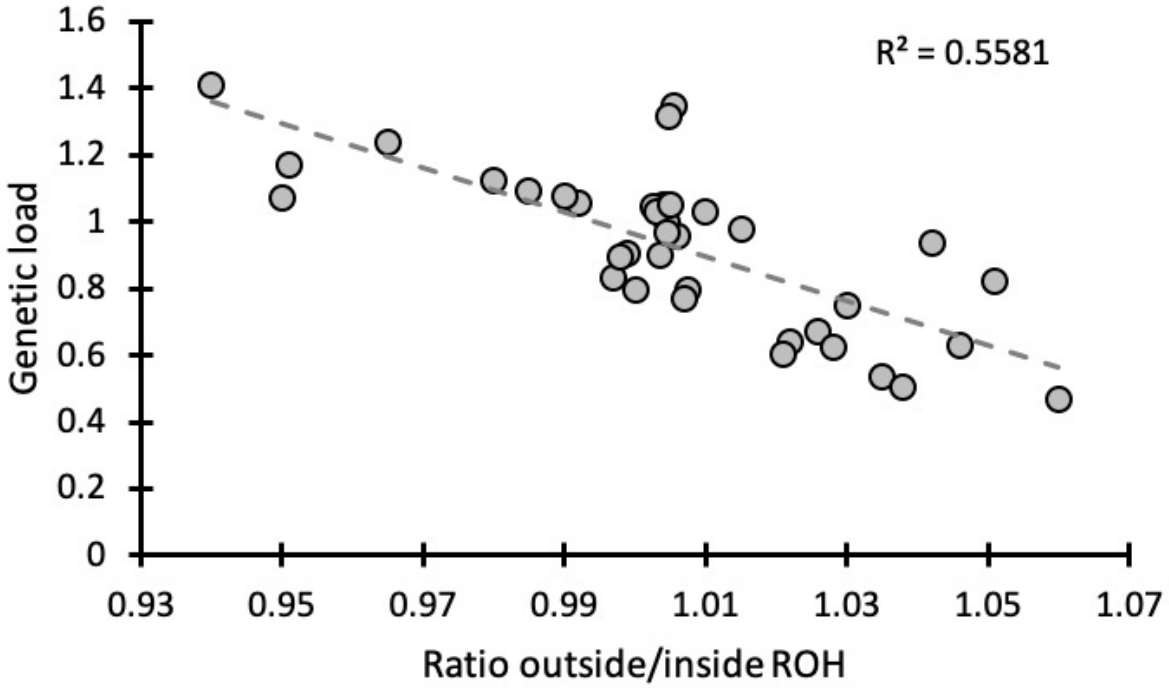

Fig. S12. Correlation between overall genetic load and the ratio of load inside and outside ROHs. Each dot represents the species average. Species with lower load generally have fewer deleterious alleles inside ROHs than outside, whereas species with high genetic load carry relatively more deleterious alleles inside ROHs.

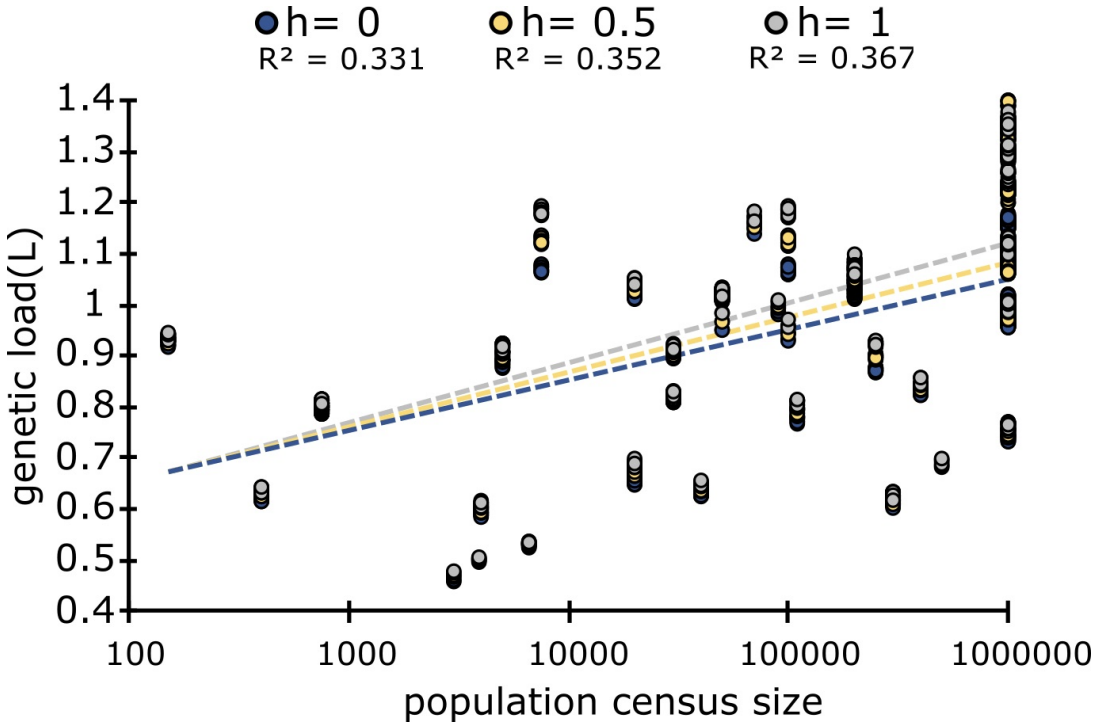

Fig. S13. Relationship between genetic load and population size for different dominance coefficients. Each circle corresponds to a study population or species, the colour of the circle corresponds to the dominance coefficient. Genetic load was calculated with the formula L = 1- W = 2shx(1-x) + sx^2, following (1), where s=GERP score of the derived allele used as a proxy for the selection coefficient, x=allele frequency and h=dominance coefficient that ranges from 0 (recessive) to 0.5 (additive) to 1 (dominant). Genetic load increases with population size for all values of h, but the correlation is stronger for dominant alleles. For small populations, dominance coefficient has no effect on the estimates of genetic load, as most derived alleles are homozygous and thus exposed to selection independent of the dominance coefficient. In large

populations the genetic load is higher if the derived alleles are dominant, but the effect of dominance is rather low in a cross-species comparison.

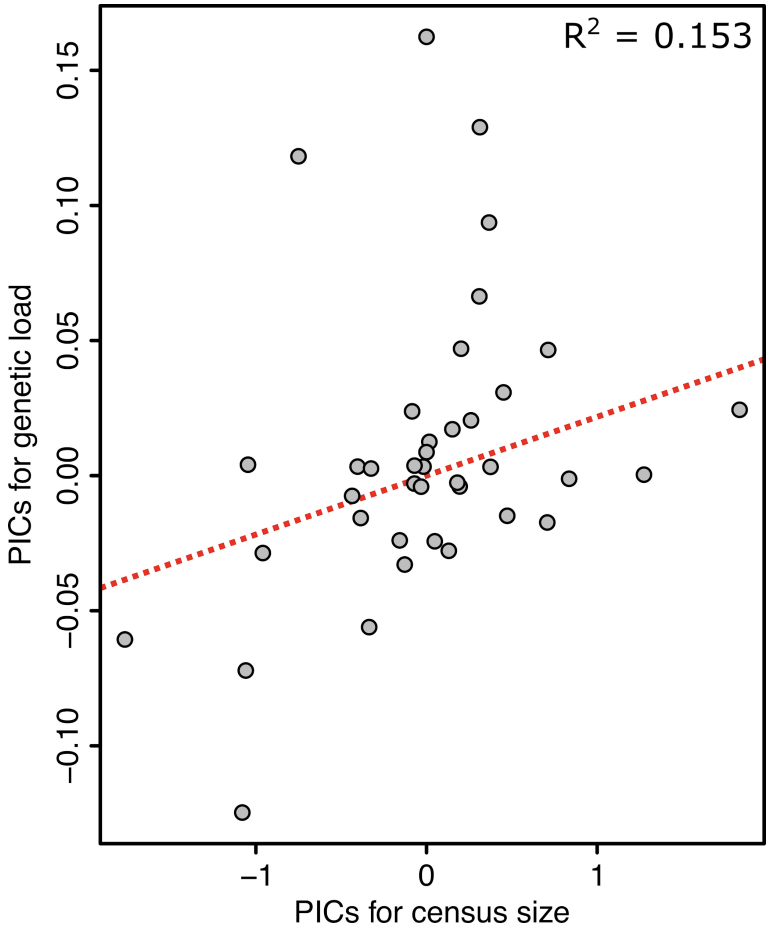

Fig. S14. Phylogenetic independent contrast (PIC). Census size and genetic load are correlated, corrected for phylogenetic relationship between the species (2) to account for ancestral traits shared between closely related species.

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)

1. M. Kimura, T. Maruyama, J. F. Crow, The mutation load in small populations. *Genetics* **48**, 1303–1312 (1963).

2. J. Felsenstein, Phylogenies and the Comparative Method. *Am. Nat.* **125**, 1–15 (1985).