Genetic load: genomic estimates and applications in non-model animals

Giorgio Bertorelle[1,11†], Francesca Raffini[1,11], Mirte Bosse[2,3], Chiara Bortoluzzi[4,5], Alessio Iannucci[6], Emiliano Trucchi[7], Hernán E. Morales[8,9,11] and Cock Van Oosterhout[10, 11†]

[1]*Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy*

[2]*Wageningen University and Research – Animal Breeding & Genomics, Wageningen, Netherlands*

[3] *Ecology and Evolution, VU University Amsterdam, The Netherlands*

[4]*Department of Genetics, University of Cambridge, United Kingdom*

[5]*Wellcome Sanger Institute, Hinxton, United Kingdom*

[6]*Department of Biology, University of Florence, Sesto Fiorentino, Italy*

[7]*Department of Life and Environmental Sciences, Marche Polytechnic University, Ancona, Italy*[8]*Section for Evolutionary Genomics, GLOBE institute, University of Copenhagen, Copenhagen, Denmark*

[9] *Centre for Evolutionary Hologenomics, GLOBE institute, University of Copenhagen, Copenhagen, Denmark*

[10]*School of Environmental Sciences, University of East Anglia, Norwich, United Kingdom*

[11]These authors contributed equally: Giorgio Bertorelle, Francesca Raffini, Hernán E. Morales, Cock Van Oosterhout

[†]e-mail: ggb@unife.it; c.van-oosterhout@uea.ac.uk

ABSTRACT

Genetic variation that is generated by mutation, recombination, and gene flow can reduce the mean fitness of a population, both now and in the future. This 'genetic load' has been estimated in a wide range of animal taxa using various approaches. Advances in genome sequencing and computational techniques now enable us to estimate the genetic load in populations and individuals without direct fitness estimates. Here, we review the classic and contemporary literature of genetic load. We describe contemporary approaches to quantify the genetic load in whole genome sequence data based on evolutionary conservation and annotations. We show that splitting the load into its two components — the realized load (or expressed load) and

masked load (or inbreeding load) — can improve our understanding of the population genetics of deleterious mutations.

[H1] Introduction

The evolutionary forces of mutation, recombination and gene flow introduce novel genetic variants into populations that form the substrate for natural selection and genetic drift, thereby driving evolutionary change. However, assuming that a population is adapted to its current environment, random changes to its genetic variation are more likely to be deleterious than beneficial[1]. Even for a novel beneficial variant, substitution of its ancestral variant creates a burden on the population, which is the cost of natural selection[2]. The resulting reduction in individual and mean population fitness is the genetic load, which could be considered as the price paid by a species for its capacity for further evolution[3].

Various 'load' definitions have been proposed to describe different processes that can lead to a reduction in fitness (**Box 1**). The term genetic load was introduced in 1950[4]. With a focus on the phenotype, it is defined as: "the proportion by which the population fitness (or whatever other trait is being considered) is decreased in comparison with an optimum genotype"[5]. This definition emphasizes the loss of fitness in the present population. However, it ignores any potential future loss in fitness caused by partially recessive deleterious mutations whose effects are not fully expressed. The "total load"[6] includes those masked mutations; for the purpose of this Review, the term 'genetic load' refers to 'total load'.

In the field of conservation genomics, understanding this total genetic load is important to assess its impact on the health and viability of endangered populations, both now and in the future. According to the International Union for Conservation of Nature (IUCN) Red List of Threatened Species, 30,449 out of 65,364 (46.6%) species of animals, fungi and plants are in decline[7]. During population decline, the composition of the genetic load changes, with many previously masked mutations becoming expressed. Hence, understanding the genetic load, as well as its composition, helps to guide future conservation efforts.

Empirical studies on the relationship between inbreeding and fitness estimates have improved our understanding of genetic loads in populations that have undergone demographic changes[8–12]. Such studies are relevant for evolutionary biology and conservation but assessing individuals' fitness is challenging in wild populations, especially in threatened species. Genomes contain valuable information about an individual's ancestry and health, including its genetic load. Recent studies have been able to analyse genetic load dynamics using whole-genome data even in the absence of fitness data (Supplementary Table 1).

Here, we review the classic population genetic definition of the genetic load[13]. We discuss the unit the load is measured in, that is, the lethal equivalents, and show how components of the load can be defined at the population and individual levels. We then describe current approaches that estimate the load using whole-genome sequence data without direct fitness estimates. Furthermore, we summarize what we have learnt from empirical studies that have estimated the load using whole genomes in wild animal populations, which have mainly focused on vertebrates (**Supplementary Table 1).** Finally, we discuss promising research areas and potential future developments in the theoretical and applied study of genetic load. We focus on sexually reproducing outcrossing diploid animals, although some of the concepts we cover are also relevant to plants and microorganisms, depending on their evolutionary dynamics. Also, we consider only (unconditionally) deleterious mutations, and not the genetic load that can be introduced by novel, better-adapted mutations that replace original variants.

[H1] The genetic load components

From a population genetics perspective, the genetic load can be defined as a statistic that summarizes the selection and dominance coefficients of deleterious mutations as a function of their frequencies in a population (**Box 2**, equations [1], [2] and [3][13,14]). The genetic load can also be expressed as a function of the genotype of these mutations in an individual genome (**Box 2**, equations [5], [6] and [7]). Furthermore, assuming that the fitness effects of mutations act independently and multiplicatively across loci, the impact of the genetic load on fitness is approximated by equations [8] and [9] (**Box 3**). Some confusion has arisen in the literature because the fitness effects of mutations are generally calculated multiplicatively, particularly for traits that affect

survival probability. However, the value of the load (that is, the lethal equivalents) is calculated by summing the selection coefficients (**Box 2**).

In diploid organisms, the genetic load can be partitioned into the realized load (also known as expressed load[6]) and the masked load (also known as the inbreeding load[15,16] or potential load[17,18]). The realized load reduces the fitness in the current generation, whereas the masked load quantifies the potential fitness loss due to (partially) recessive deleterious mutations that may become expressed in future generations depending on the population's demography (for example, inbreeding, population decline or subdivision). The genetic load is the sum of the realized load and the masked load. This terminology emphasizes the effect of the load rather than processes that give rise to the load (**Box 1**). Put simply, the genetic load is binary: the deleterious fitness effects of a mutation can either be expressed (in the realized load) or not expressed (in the masked load). Furthermore, the mean fitness of a population subject to the negative effects of deleterious mutations is approximately equal to the negative exponent of the realized load (**Box 3**).

The unit of the load components is the lethal equivalent, and it was originally defined in terms of additive fitness effects as "a group of mutant genes of such number that, if dispersed in different individuals, they would cause on the average one death, e.g., one lethal mutant, or two mutants each with 50 per cent probability of causing death, *etc.*"[13]. However, under a model of multiplicative fitness effects, a more accurate definition of a lethal equivalent is a group of mutant alleles with a summed selection coefficient equal to one. Assuming that the fitness effects of mutations are multiplicative across loci, the survival probability of an individual with a realized load of one lethal equivalent is approximately equal to e^{-1} (~36.8%), and with two lethal equivalents equal to e^{-2} (~13.5%). It can also be understood as the zero term in a Poisson distribution whose mean is in lethal equivalents, that is the proportion of individuals that carry zero lethal equivalents[19].

Our definition assumes that each mutation reduces fitness by a given probability, which might be more appropriate for studies on the survival rate, such as in the analysis of inbreeding depression or extinction risk in conservation genomics. According to this definition, individuals with a realized load of more than one lethal equivalent can still survive. It makes the explicit assumption that (semi)lethal mutations have low frequencies in the population and, hence, that homozygous lethals are rare. If (semi)lethal mutations are common, fitness is not equal to the negative exponent of

the realized load. For example, death is certain for an individual that expresses a lethal mutation (and not e^{-1}~36.8%) (see Supplementary Information SI1–SI2 and Supplementary Figures S1-S2 for further details).

Depending on the age of a population, the genetic load has a non-linear relationship with the effective population size (N_e) (**Fig. 1**). Small populations that have persisted for long periods of time are expected to possess the highest genetic load because many slightly deleterious mutations have become fixed, which elevates the realized load. By contrast, small populations tend to possess a low masked load. Hence, they are not expected to show much inbreeding depression[17], as mating between close relatives does not substantially increase the number of homozygous deleterious loci. The masked load increases with population size, making large populations more prone to inbreeding depression during population decline (**Fig. 1**). Although bottlenecks purge some highly deleterious mutations, thereby reducing the genetic load[20–22], they also convert the masked load into the realized load. Furthermore, prolonged bottlenecks eventually result in a persistent realized load due to the fixation of deleterious mutations[23] (**Fig. 2**). Consequently, even after demographic recovery, migration or genetic rescue might be required to replace fixed deleterious variants and restore fitness[18].

[H1] From genomes to load estimates

In absence of direct fitness information, genetic load estimates can be obtained from whole-genome sequencing data in two steps. First, the deleterious effects of mutations are predicted. Second, the deleterious scores of these mutations are summarized to produce a load index, or a set of indices, which can be considered as proxies and used to study the load components.

[H2] Predicting deleterious mutations

Various methods have been developed to predict the fitness consequences of mutations that can be broadly grouped in two categories (**Table 1**): (1) methods analysing the evolutionary conservation across a multi-species alignment; and (2) approaches estimating the expected effects based on the type of substitution and its known effects in model species. A third, less widely used source of information is the level of gene expression of the mutated gene.

Evolutionary conservation approaches can be applied to whole-genome sequences without annotations. A multiple sequence alignment is built to assess the level of conservation at individual nucleotide or genomic regions across lineages. Point mutations, insertion and deletions (indels) and structural variants can all contribute to the genetic load[24], but most predictors focus on point mutations (**Table 1**, **Supplementary Table 1**). A score quantifying the harmful effects of each variant is then assigned under the assumption that the level of conservation in the alignment reflects its functional importance[25–28]. The accuracy of the prediction increases with the number and the phylogenetic distance of the species in the alignment, as long as anchor species at intermediate evolutionary distances are also included[26,29]. Large alignments can be challenging given that there is considerable turnover of constrained sites across phylogenies[30], and because they are computationally intensive. Lineage-specific substitutions should be considered with great care, as they may represent adaptations, or they may reflect changes in selection pressures, rather than unconditionally deleterious mutations[30].

One of the most widely used methods based on evolutionary conservation, especially in wild animals, is genomic evolutionary rate profiling (GERP)[26,31]. A GERP score is a continuous variable estimating the average number of substitutions that would have accumulated under neutrality but have been removed ('rejected') by purifying selection. High scores reflect many rejected mutations and indicate that substitutions are only rarely tolerated during evolution[26,31]. Sites with high GERP scores are assumed to experience strong negative selection, although turnover of sites and changes in selection pressures across the phylogeny set limitations on the use of comparative genomic approaches[30].

An alternative approach to evolutionary conservation methods consists of the direct inference of possible mutation effects using information from biochemical studies or functional annotations of the site or region where the mutation occurs[32,33]. For example, the chemical properties of amino acids can be used to assign numerical scores to any of the possible changes[34]. Also, substitutions in a coding region can be categorized as either synonymous (more likely to be selectively neutral) or non-synonymous (possibly deleterious). This type of information can be translated into a categorical variable such as low, moderate or high according to their predicted nearly neutral, mildly deleterious or highly harmful type of change. Additional information based on protein structure, experimentally known effects of the mutation or more

detailed genome annotations can provide further data on the effects of variants such as missense, frameshift, stop-gain, loss of function, as done, for example, by SNPEff[32,35] A derived mutation can erroneously be predicted as deleterious when it has been identified using one or a few outgroups or when multiple mutations occur at the same site. Such errors can be avoided by integrating the probability of multiple mutations at the same site based on the overall mutational spectrum[36] in the target species or population and applying corrections based on the DNA sequence context[37].

Predictions of the negative impact of a mutation become more precise and reliable as the quality of the genome annotation increases, and if evolutionary conservation is included as one of the annotation variables (for example, combining conservation scores and mutation type with local DNA structure, GC content, distance to splice sites and gene expression)[33,38–40] (**Table 1**). Some of these ensemble approaches, such as SIFT[38,41], PolyPhen[39], ANNOVAR[33] and VEP[40], directly target functional effects of regions that are already known from existing databases[33,38–41], whereas others, such as CADD[42] or GWAVA[43], move beyond the available information and assign a deleteriousness score to all potential mutations genome-wide, instead of relying on only the established ones[42–44]. These methods provide reliable evidence of the fitness effect of variants, but are mostly available for model organisms and require specific genomic alignments and annotations. When focusing on ultra-conserved elements[45], scores obtained from the genome of a model animal can potentially be lifted over to a closely related non-model species.

Gene expression data can also be used to predict the potential fitness impact of specific variants. This approach takes advantage of the negative correlation observed between gene expression and protein polymorphism: highly expressed genes are usually associated with highly conserved coding sequences[46–49]. Mutations in highly expressed genes should therefore be prioritized in genetic load estimates. However, this approach relies on an additional correlation step between the data and the prediction and requires known gene expression levels[22]. Furthermore, gene expression data cannot be applied to single nucleotides, limiting their current application to filtering out highly down-regulated genes from genetic load analyses[22].

[H2] Translating deleteriousness scores into genetic load proxies

The predicted effects of variants have been used to produce genetic load proxies for single individuals or populations, and several approaches have been proposed (**Fig.**

3, **Supplementary Table 1**). If numerical scores are available (for example, GERP or CADD scores), these can be summed or averaged across the genome of each individual[50–53]. These values can be further averaged across individuals to obtain quantities related to the population's genetic load, which can be standardized using neutral variation[22,54,55]. When variants can only be classified into categories according to the type of substitution or the intensity of their harmful effect, load indices based on the number or ratios of the observed variants in each category can be computed[56–62]. These indices can also be computed separately for homozygous and heterozygous loci[56–58,63–66], or they can be calculated for deleterious variants that occur at high frequency or are fixed in different populations[56,67]. Furthermore, indices suitable for comparing different populations or samples with different ages (for example, ancient versus modern) have been developed (for example, R_{XY}[22,66,68]).

Unfortunately, given the various approaches used to translate deleteriousness scores into genetic load proxies, there is no agreed gold standard that enables comparison of the load components across studies (**Supplementary Table 1**). Even in humans, the load estimates differ markedly depending on how they are calculated[69–72].

[H1] Empirical genetic load estimates

Until recently, the approaches described above were applied mainly to model and domesticated species, where genomic data are abundant and the crucial validation of the predicted damage of each variant is frequently possible through functional studies. However, their use in wild non-model species is rapidly increasing (**Supplementary Table 1**).

[H2] Genetic load estimates in model and domesticated species

Model organisms, such as *Drosophila melanogaster*, have substantially contributed to our understanding of how genetic load manifests in genomes[73–76]. Experimental validations were used to assess the fitness effects of specific mutations[76], and often populations containing known harmful mutations were used to shed light on the synergistic epistatic effects and the role of mutations in inbreeding depression[77,78].

These genomic studies have started to reveal important insights into the impact of deleterious mutations both in coding and non-coding regions, which have been validated experimentally in some cases. Causative non-coding variants in human

diseases are often predicted through scores based on diverse genomic features derived from gene model annotations, evolutionary constraints and functional predictions, such as GERP, GWAWA and CADD[31,42,43,79] (**Table 1, Supplementary Table 1**). These multi-source informed predictions have recently been extended to non-human organisms, such as mouse[44], chicken[80] and pig[81]. The challenge now lies in connecting genome-wide bioinformatics predictions and experimental validation of fitness effects. Meta-analyses in livestock show consistent fitness reduction when genomics-derived inbreeding measurements increase[82]. For example, a 1% increment in inbreeding results in an average decrease of 1.3% for a given trait value[83,84]. Furthermore, genetic load seems to decrease with haplotype age, suggesting purging of deleterious variants within the breeding population[82,84,85]. Potentially recessive lethal variants can be identified by screening pedigrees for missing homozygous haplotypes[86,87], and causal mutations can be validated by combining functional annotations with carrier x carrier mating. Such variants can greatly compromise population fertility, and lethal recessive haplotypes likely constitute >10% of the genetic causes of stillbirths in purebred pigs[88,89].

Besides phenotypic validations of predicted harmful mutations in living organisms, novel developments in organoids, single-cell analyses and CRISPR–Cas genome editing greatly enhance our understanding of the functional impact of mutations (see [90,91] for reviews). Such studies help bridge the gap between sequence-derived predictions and fitness-related estimates of genetic load, and their insights may eventually be used to alleviate load in target populations[92].

[H2] Genetic load estimates in wild animals

Recent empirical insights into the genetic load have been obtained using whole-genome data without fitness data in wild animals (see, for example,[22,51,63,64,93]; **Supplementary Table 1**). Empirical studies in natural populations show that large ancestral populations accumulate substantial masked load[10,51]. During population decline, part of this load is lost by random genetic drift, and part is purged by selection as recessive deleterious mutations with large fitness effects become exposed by inbreeding[53,56,94]. However, a proportion of the masked load is converted into realized load, resulting in inbreeding depression[18]. Both the type of demographic contraction and the distribution of the fitness effects of mutations play important roles in determining these processes[10,21]. Harmful variants can become fixed during extreme

bottlenecks, but those with larger fitness effects are likely to be purged during long-term declines[56,67,95–99]. Persistently small, isolated populations are expected to exhibit increased load accumulation due to strong genetic drift[18,59,62,64]. In this case, highly harmful mutations might be effectively purged when exposed to selection under most conditions, but mildly deleterious variants can accumulate during demographic collapse[22,100] (**Fig. 2**).

The patterns observed in wild animals highlight that the genetic architecture and degree of dominance also determine the fate of deleterious mutations[21,101]. Accumulation of genetic load has been linked to an increased homozygosity of strongly deleterious recessive alleles[51,63,93,102] or a rise in frequency of harmful mutations[53,59] or both[58,100]. Within genomes, non-coding regions, chromosomal rearrangements, regions with runs of homozygosity and introgressed loci carrying disadvantageous alleles showed an enrichment of genetic load[53,59,103,104].

Remarkably, there is little correlation between the IUCN Red List status of species and their estimated genetic load[105,106]. While some endangered species have a higher genetic load than their more abundant counterparts[55,57,58,63,64,67,93,96,98,102], others seem to have a low load[51,107]. This poor concordance could be partly explained by the inconsistency amongst studies in estimating and reporting the genetic load, hindering effective conservation management. In addition, on its own, the genetic load is not a particularly informative statistic to assess the genetic health of a population. For example, the number of lethal equivalents calculated using equation [1] is not (immediately) affected by inbreeding or genetic drift, as it takes time to purge deleterious mutations[108]. Moreover, mildly deleterious mutations are expected to accumulate slowly in populations with a long-term small effective population size, counteracting the purifying effects of selection (**Fig. 1**). Once fixed in a population, these mutations no longer contribute to inbreeding depression[10]. Therefore, more informative than merely reporting the total genetic load is to delineate this statistic into its components, i.e., the masked load and the realized load. These components capture the current loss in fitness of the population, as well as the predicted future fitness loss caused by deleterious mutations that may lower the population's long-term viability[17]. Such information is crucial for correct management decisions, including genetic rescue, assisted gene flow, population supplementation and reintroduction programs (see [10,63,64,109,110]). Conservation geneticists tend to prioritize maximizing genetic diversity rather than minimizing the genetic load[111], but with better data and

understanding of the genetic load, a more balanced approach is likely to improve conservation outcomes[112].

[H1] Future directions

[H2] Towards a standardized use of genomic data to predict load components

Previous studies quantified the genetic load by counting the scores of derived homozygotes twice, and the scores of heterozygotes once[65,69,97,113] (**Box 2**). With estimates of the dominance coefficients, the relative realized load and relative fitness of individuals could be calculated (**Box 2**, **Box 3**). However, these would be rough approximations, and more accurate estimates of masked load and realized load require: 1) the identification (and possibly the validation with empirical fitness data) of the relationship between deleteriousness scores and selection coefficients; and 2) improved estimates of the dominance coefficients. Such data may become available — for example, as specific distributions for different bins of selection coefficients[114–118] — with the whole-genome sequencing of thousands of species[119–121]. The improved alignments and annotations of large datasets of whole genomes[42,79,120,122,123] and the integration of phylogenomic and population genomics approaches[124] will help to rapidly advance the field. Genomic sequences generated for individuals with fitness data across their entire life history, from deceased embryos to healthy adults, could be very useful for validating the genomic measures of realized load. Fitness data of threatened vertebrates and samples collected post-mortem in zoos could provide a largely overlooked, valuable resource for such studies[125].

[H2] The genotype–fitness relationship

The relationship between genome sequence-derived estimates of the load and their fitness consequences is mostly untested and usually relies on previously identified disease-causing mutations in model organisms and humans[126]. Whereas functional studies on specific mutations predicted to be deleterious are important, most potentially harmful mutations will not be experimentally validated (or even explored) in most species. In genetic load investigations, a gap has emerged between fitness-oriented studies and sequence-oriented studies, which infer potential fitness consequences based on large species alignments or studies on model species. We therefore need improved tools to bioinformatically identify deleterious mutations, such

as EVE (evolutionary model of variant effect)[127], which predicts the pathogenicity of protein variants by modelling the distribution of sequence variation across species, and VIVID, which integrates evolutionary conservation and functional analyses of variants with 3D protein models[128]. In addition, more field-based studies are needed that correlate phenotypic fitness values to the genomic predictions of the load components of individuals in their natural environment. Our understanding of the relationship between load scores, selection and dominance coefficients and fitness effects can be further improved by simulations[23,85,109,129].

[H2] How ancient DNA data can contribute to genetic load investigations

Empirically, the effects of demographic events, such as bottlenecks, population fragmentation and population size decline, can best be studied using temporal genomic samples (that is, ancient and historical museum-preserved DNA). Time series genomic data have been successfully used to show how the accumulation of genetic load increases in response to early domestication bottlenecks (the "cost of domestication"[130,131]) and to more recent artificial breeding practices[50,132]. Comparisons between contemporary and historical samples from museum collections have also shown increases in the number of deleterious mutations associated with higher inbreeding after demographic bottlenecks[96,98]. Theoretically, only the realized load is expected to increase after a rapid population decline, and hence, it would be interesting to re-examine these data and calculate the separate load components.

Ancient DNA data from extinct populations and species can yield information on the genetic load dynamics before extinction[133]. They can also help to test the predicted correlation between increased genetic load and extinction probability[134], and further our understanding of the genomic signature of population decline and extinction[135]. The assembly of reference genomes from museum samples or from ancient samples represents an important development to avoid the bias introduced when mapping ancient data to evolutionarily distant modern genomes[136,137]. However, the calling of genotypes from ancient DNA data remains challenging[138–140], and it may require additional laboratory and computational efforts to confirm the presence of deleterious mutations.

[H2] Practical applications in conservation biology

Both *ex situ* and *in situ* conservation of endangered species are increasingly implementing genome-level information to more accurately maintain high genetic diversity, maximize the representation of different subpopulations, minimize kinship and inbreeding, assess genetic introgression and preserve local adaptation[22,63,141]. Future conservation actions could benefit from including proxies of genetic load components, for example, to guide captive breeding[11], genetic rescue[111,129] and reintroduction programmes[109]. We foresee that individual-based modelling approaches will become increasingly important, for example, to assess the genetic load that has accumulated in the ancestral population. This could be inferred by simulating the historic population size and demographic trajectories. Such data can be derived from analyses that use whole-genome sequence data to estimate the time to the most recent common ancestor[142,143], and methods to infer N_e from the spectrum of linkage disequilibrium between pairs of loci[144]. Furthermore, individual-based models can be used to assess the future conservation needs of species[145].

The classical genetic conservation paradigm that globally high diversity equals a healthy population is currently being debated, and it is increasingly clear that we need to better understand the consequences of the different components (neutral, beneficial or deleterious) of diversity[10,23,146–148]. Genome-wide diversity correlates positively with higher (current) fitness[149–151], and higher diversity is thought to increase the (future) adaptive potential of wild populations[152]. Similarly, highly diverse populations typically carry an elevated masked load. For example, computer simulations suggest that high diversity in source populations used for genetic rescue could introduce deleterious variation into the recipient population, ultimately resulting in more inbreeding depression and a greater extinction risk[129]. Such advice goes against decades of conservation wisdom[147,148,153] and has made the application of genomic simulations in conservation a much-discussed issue[111]. Ultimately, the specific effects of different types of genetic variation on fitness are complex and will depend on the ancestral population size (large *versus* small[17,66]), recent demographic dynamics (for example, severity and duration of bottlenecks[99]), the expected amount of environmental change[154,155] and management interventions such as assisted gene flow and genetic rescue[109,156]. We hope that an improved framework to estimate the genetic load and its components will help to deepen these conversations within the conservation genetic community.

References

1. Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96–98 (1973).
2. Haldane, J. B. S. The cost of natural selection. *J. Genet.* **55**, 511 (1957).
3. Haldane, J. B. S. The effect of variation of fitness. *Am. Nat.* **71**, 337–349 (1937).
4. Muller, H. J. Our load of mutations. *Am. J. Hum. Genet.* **2**, 111–176 (1950).

First explicit use of the term "genetic load" with an explanation on the relationship between mutation rate and load, and controversial speculation on the possible effects of relaxed selection in humans.

5. Crow, J. F. Some possibilities for measuring selection intensities in man. *Hum. Biol.* **30**, 3–13 (1958).
6. Crow, J. F. Genetic loads and the cost of natural selection. in *Mathematical Topics in Population Genetics* (ed. Kojima, K.) 128–177 (Springer, 1970). doi:10.1007/978-3-642-46244-3_5.
7. IUCN Standards and Petitions Committee. *The IUCN Red List of Threatened Species. Version 2021-2*. https://www.iucnredlist.org (2021).
8. Barrett, S. C. H. & Charlesworth, D. Effects of a change in the level of inbreeding on the genetic load. *Nature* **352**, 522–524 (1991).
9. Saccheri, I. J., Lloyd, H. D., Helyar, S. J. & Brakefield, P. M. Inbreeding uncovers fundamental differences in the genetic load affecting male and female fertility in a butterfly. *Proc. Biol. Sci.* **272**, 39–46 (2005).
10. van Oosterhout, C. Mutation load is the spectre of species conservation. *Nat. Ecol. Evol.* **4**, 1004–1006 (2020).
11. van Oosterhout, C. *et al.* The guppy as a conservation model: implications of parasitism and inbreeding for reintroduction success. *Conserv. Biol. J. Soc. Conserv. Biol.* **21**, 1573–1583 (2007).
12. Fox, C. W., Scheibly, K. L. & Reed, D. H. Experimental evolution of the genetic load and its implications for the genetic basis of inbreeding depression. *Evol. Int. J. Org. Evol.* **62**, 2236–2249 (2008).
13. Morton, N. E., Crow, J. F. & Muller, H. J. An estimate of the mutational damage in man from data on consanguineous marriages. *Proc. Natl. Acad. Sci. U. S. A.* **42**, 855–863 (1956).

Definition of lethal equivalents and use of regression analysis to estimate the inbreeding load or masked load.

14. Cavalli-Sforza, L. L. & Bodmer, W. *Human population genetics*. (Dover Publications, 1971).
15. Caballero, A., Bravo, I. & Wang, J. Inbreeding load and purging: implications for the short-term survival and the conservation management of small populations. *Heredity* **118**, 177–185 (2017).
16. Nietlisbach, P., Muff, S., Reid, J. M., Whitlock, M. C. & Keller, L. F. Nonequivalent lethal equivalents: Models and inbreeding metrics for unbiased estimation of inbreeding load. *Evol. Appl.* **12**, 266–279 (2019).

17. Mathur, S. & DeWoody, J. A. Genetic load has potential in large populations but is realized in small inbred populations. *Evol. Appl.* **14**, 1540–1557 (2021).

Empirical study measuring load components in populations with different demography and the interplay between inbreeding effects and purging.

18. Mathur, S., Tomeček, J., Tarango-Arámbula, L., Perez, R. & DeWoody, A. An evolutionary perspective on contemporary genetic load in threatened species to inform future conservation efforts. *Authorea* (2021) doi:10.22541/au.162495929.94655412/v1.
19. Lewontin, R. *The genetic basis of evolutionary change*. (Columbia University Press, 1974).
20. Wang, J., Hill, W. G., Charlesworth, D. & Charlesworth, B. Dynamics of inbreeding depression due to deleterious mutations in small populations: mutation parameters and inbreeding rate. *Genet. Res.* **74**, 165–178 (1999).
21. Glémin, S. How are deleterious mutations purged? Drift versus nonrandom mating. *Evol. Int. J. Org. Evol.* **57**, 2678–2687 (2003).
22. Grossen, C., Guillaume, F., Keller, L. F. & Croll, D. Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. *Nat. Commun.* **11**, 1001 (2020).

Analysis of whole genome sequence data showing how highly deleterious mutations are purged, while mildly deleterious mutations accumulate in populations characterized by different series of bottleneck events.

23. Kardos, M. *et al.* The crucial role of genome-wide genetic variation in conservation. *Proc. Natl. Acad. Sci.* **118**, e2104642118 (2021).

Critical evaluation of the use of functional genetic diversity in conservation genetics, concluding that the conservation of genome-wide genetic variation should be prioritized.

24. Mills, R. E. *et al.* Natural genetic variation caused by small insertions and deletions in the human genome. *Genome Res.* **21**, 830–839 (2011).
25. Boffelli, D. *et al.* Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science* **299**, 1391 (2003).
26. Cooper, G. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* **15**, 901–913 (2005).

Introduction of genomic evolutionary rate profiling (GERP) to identify genomic regions that are evolutionarily constrained (i.e., exhibit nucleotide substitution deficits) using whole genomes alignments.

27. Asthana, S., Roytberg, M., Stamatoyannopoulos, J. & Sunyaev, S. Analysis of sequence conservation at nucleotide resolution. *PLOS Comput. Biol.* **3**, e254 (2007).
28. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).

29. Braasch, I. *et al.* The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat. Genet.* **48**, 427–437 (2016).

30. Huber, C. D., Kim, B. Y. & Lohmueller, K. E. Population genetic models of GERP scores suggest pervasive turnover of constrained sites across mammalian evolution. *PLOS Genet.* **16**, e1008827 (2020).

Analysis and discussion on the use and limitations of GERP scores to identify deleterious mutations.

31. Davydov, E. V. *et al.* Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. *PLOS Comput. Biol.* **6**, e1001025 (2010).
32. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).

This paper introduces a software to categorize the effects of genetic variants based on gene annotations.

33. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
34. Grantham, R. Amino acid difference formula to help explain protein evolution. *Science* **185**, 862–864 (1974).
35. Herrgard, S. *et al.* Prediction of deleterious functional effects of amino acid mutations using a library of structure-based function descriptors. *Proteins Struct. Funct. Bioinforma.* **53**, 806–816 (2003).
36. Harris, K. & Pritchard, J. K. Rapid evolution of the human mutation spectrum. *eLife* **6**, e24284 (2017).
37. Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Context dependence, ancestral misidentification, and spurious signatures of natural selection. *Mol. Biol. Evol.* **24**, 1792–1800 (2007).
38. Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
39. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **76**, 7.20. 1-7.20. 41 (2013).
40. McLaren, W. *et al.* The Ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
41. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
42. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).

Introduction of Combined Annotation–Dependent Depletion (CADD) to estimate the fitness effect of known variants and expand this prediction to new mutations.

43. Ritchie, G. R. S., Dunham, I., Zeggini, E. & Flicek, P. Functional annotation of noncoding sequence variants. *Nat. Methods* **11**, 294–296 (2014).
44. Groß, C., de Ridder, D. & Reinders, M. Predicting variant deleteriousness in non-human species: applying the CADD approach in mouse. *BMC Bioinformatics* **19**, 373 (2018).
45. Bejerano, G. *et al.* Ultraconserved elements in the human genome. *Science* **304**, 1321 (2004).

Identification, annotation, and analysis of ultra-conserved elements (UCEs) in the vertebrate genome.

46. Pál, C., Papp, B. & Hurst, L. D. Highly expressed genes in yeast evolve slowly. *Genetics* **158**, 927–931 (2001).
47. Drummond, D. A. & Wilke, C. O. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134**, 341–352 (2008).
48. Zhang, J. & Yang, J.-R. Determinants of the rate of protein sequence evolution. *Nat. Rev. Genet.* **16**, 409–420 (2015).
49. Marek, A. & Tomala, K. The contribution of purifying selection, linkage, and mutation bias to the negative correlation between gene expression and polymorphism density in yeast populations. *Genome Biol. Evol.* **10**, 2986–2996 (2018).
50. Schubert, M. *et al.* Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl. Acad. Sci.* **111**, E5661 (2014).
51. Van Der Valk, T., Manuel, M. de, Marques-Bonet, T. & Guschanski, K. Estimates of genetic load suggest frequent purging of deleterious alleles in small populations. *bioRxiv* 696831 (2021) doi:10.1101/696831.
52. Willemsen, D., Cui, R., Reichard, M. & Valenzano, D. R. Intra-species differences in population size shape life history and genome evolution. *eLife* **9**, e55794 (2020).
53. Orlando, L. & Librado, P. Origin and evolution of deleterious mutations in horses. *Genes* **10**, 649 (2019).
54. Khan, A. *et al.* Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. *Proc. Natl. Acad. Sci.* **118**, https://doi.org/10.1073/pnas.2023018118 (2021).
55. Liu, L. *et al.* Genetic consequences of long-term small effective population size in the critically endangered pygmy hog. *Evol. Appl.* **14**, 710–720 (2021).
56. Xue, Y. *et al.* Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science* **348**, 242–245 (2015).
57. Wang, P. *et al.* Genomic consequences of long-term population decline in brown eared pheasant. *Mol. Biol. Evol.* **38**, 263–273 (2021).
58. Hu, Y. *et al.* Genomic evidence for two phylogenetic species and long-term population bottlenecks in red pandas. *Sci. Adv.* **6**, eaax5751 (2020).

59. Leroy, T. *et al.* Island songbirds as windows into evolution in small populations. *Curr. Biol.* **31**, 1303-1310.e4 (2021).
60. Jay, P. *et al.* Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nat. Genet.* **53**, 288–293 (2021).
61. de Valles-Ibáñez, G. *et al.* Genetic load of loss-of-function polymorphic variants in great apes. *Genome Biol. Evol.* **8**, 871–877 (2016).
62. Kutschera, V. E. *et al.* Purifying selection in corvids is less efficient on islands. *Mol. Biol. Evol.* **37**, 469–474 (2020).
63. Robinson, J. A. *et al.* Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. *Sci. Adv.* **5**, eaau0757 (2019).
64. Robinson, J. A. *et al.* Genomic flatlining in the endangered island fox. *Curr. Biol.* **26**, 1183–1189 (2016).
65. Ochoa, A. & Gibbs, H. L. Genomic signatures of inbreeding and mutation load in a threatened rattlesnake. *Mol. Ecol.* **00**, 1–16 (2021).
66. Dussex, N. *et al.* Population genomics reveals the impact of long-term small population size in the critically endangered kākāpō. *Cell Genomics* **1**, 100002 (2021).
67. Benazzo, A. *et al.* 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).
68. Do, R. *et al.* No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nat. Genet.* **47**, 126–131 (2015).
69. Henn, B. M., Botigué, L. R., Bustamante, C. D., Clark, A. G. & Gravel, S. Estimating the mutation load in human genomes. *Nat. Rev. Genet.* **16**, 333–343 (2015).

Review paper showing that the predicted burden of deleterious alleles can differ substantially in human populations, depending on the methodology to calculate load.

70. Simons, Y. B. & Sella, G. The impact of recent population history on the deleterious mutation load in humans and close evolutionary relatives. *Curr. Opin. Genet. Dev.* **41**, 150–158 (2016).
71. Simons, Y. B., Turchin, M. C., Pritchard, J. K. & Sella, G. The deleterious mutation load is insensitive to recent population history. *Nat. Genet.* **46**, 220–224 (2014).
72. Lohmueller, K. E. The distribution of deleterious genetic variation in human populations. *Curr. Opin. Genet. Dev.* **29**, 139–146 (2014).
73. Wallace, B. Studies on irradiated populations of *Drosophila melanogaster*. *J. Genet.* **54**, 280–293 (1956).
74. Keightley, P. D. Nature of deleterious mutation load in *Drosophila*. *Genetics* **144**, 1993–1999 (1996).

75. Wallace, B. The estimation of adaptive values of experimental populations. *Evolution* **6**, 333–341 (1952).
76. Mukai, T., Chigusa, S. I., Mettler, L. E. & Crow, J. F. Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **72**, 335–355 (1972).
77. Domínguez-García, S., García, C., Quesada, H. & Caballero, A. Accelerated inbreeding depression suggests synergistic epistasis for deleterious mutations in *Drosophila melanogaster*. *Heredity* **123**, 709–722 (2019).
78. Whitlock, M. C. Fixation of new alleles and the extinction of small populations: drift load, beneficial alleles, and sexual selection. *Evolution* **54**, 1855–1861 (2000).
79. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
80. Groß, C. *et al.* Prioritizing sequence variants in conserved non-coding elements in the chicken genome using chCADD. *PLOS Genet.* **16**, e1009027 (2020).
81. Groß, C. *et al.* pCADD: SNV prioritisation in *Sus scrofa*. *Genet. Sel. Evol.* **52**, 4 (2020).
82. Doekes, H. P., Bijma, P. & Windig, J. J. How depressing is inbreeding? A meta-analysis of 30 years of research on the effects of inbreeding in livestock. *Genes* **12**, 926 (2021).
83. Leroy, G. Inbreeding depression in livestock species: review and meta-analysis. *Anim. Genet.* **45**, 618–628 (2014).
84. Doekes, H. P. *et al.* Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein–Friesian dairy cattle. *Genet. Sel. Evol.* **51**, 54 (2019).
85. Stoffel, M. A., Johnston, S. E., Pilkington, J. G. & Pemberton, J. M. Mutation load decreases with haplotype age in wild Soay sheep. *Evol. Lett.* **5**, 187–195 (2021).
86. Pausch, H. *et al.* Homozygous haplotype deficiency reveals deleterious mutations compromising reproductive and rearing success in cattle. *BMC Genomics* **16**, 312 (2015).
87. VanRaden, P. M., Olson, K. M., Null, D. J. & Hutchison, J. L. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *J. Dairy Sci.* **94**, 6153–6161 (2011).
88. Derks, M. F. L. *et al.* Loss of function mutations in essential genes cause embryonic lethality in pigs. *PLOS Genet.* **15**, e1008055 (2019).
89. Charlier, C. *et al.* NGS-based reverse genetic screen for common embryonic lethal mutations compromising fertility in livestock. *Genome Res.* **26**, 1333–1341 (2016).
90. Tait-Burkard, C. *et al.* Livestock 2.0 – genome editing for fitter, healthier, and more productive farmed animals. *Genome Biol.* **19**, 204 (2018).
91. Clark, E. L. *et al.* From FAANG to fork: application of highly annotated genomes to improve farmed animal production. *Genome Biol.* **21**, 285 (2020).

92. Johnsson, M. *et al.* Removal of alleles by genome editing (RAGE) against deleterious load. *Genet. Sel. Evol.* **51**, 14 (2019).
93. Robinson, J. A., Brown, C., Kim, B. Y., Lohmueller, K. E. & Wayne, R. K. Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in island foxes. *Curr. Biol.* **28**, 3487-3494.e4 (2018).
94. Crnokrak, P. & Barrett, S. C. H. Perspective: purging the genetic load: a review of the experimental evidence. *Evolution* **56**, 2347–2358 (2002).
95. Schumer, M. *et al.* Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* **360**, 656 (2018).
96. Van Der Valk, T., Díez-del-Molino, D., Marques-Bonet, T., Guschanski, K. & Dalén, L. Historical genomes reveal the genomic consequences of recent population decline in Eastern gorillas. *Curr. Biol.* **29**, 165-170.e6 (2019).

This paper shows the relevance of museum samples to monitor genetic load accumulation in declining populations.

97. Cui, R. *et al.* Relaxed selection limits lifespan by increasing mutation load. *Cell* **178**, 385-399.e20 (2019).
98. Feng, S. *et al.* The genomic footprints of the fall and recovery of the crested ibis. *Curr. Biol.* **29**, 340-349.e7 (2019).

This is an evaluation of temporal genomic erosion by directly comparing historical and contemporary re-sequenced genomes of the crested ibis, a species once on the brink of extinction.

99. Bortoluzzi, C. *et al.* The type of bottleneck matters: Insights into the deleterious variation landscape of small managed populations. *Evol. Appl.* **13**, 330–341 (2020).
100. Beichman, A. C. *et al.* Aquatic adaptation and depleted diversity: a deep dive into the genomes of the sea otter and giant otter. *Mol. Biol. Evol.* **36**, 2631–2655 (2019).
101. van Oosterhout, C. *et al.* Inbreeding depression and genetic load of sexually selected traits: how the guppy lost its spots. *J. Evol. Biol.* **16**, 273–281 (2003).
102. de Manuel, M. *et al.* The evolutionary history of extinct and living lions. *Proc. Natl. Acad. Sci.* **117**, 10927–10934 (2020).
103. Moran, B. M. *et al.* The genomic consequences of hybridization. *eLife* **10**, e69016 (2021).
104. Jones, M. R., Mills, L. S., Jensen, J. D. & Good, J. M. The origin and spread of locally adaptive seasonal camouflage in snowshoe hares. *Am. Nat.* **196**, 316–332 (2020).
105. Díez-del-Molino, D., Sánchez-Barreiro, F., Barnes, I., Gilbert, M. T. P. & Dalén, L. Quantifying temporal genomic erosion in endangered species. *Trends Ecol. Evol.* **33**, 176–185 (2018).
106. Hedrick, P. W. & Garcia-Dorado, A. Understanding inbreeding depression, purging, and genetic rescue. *Trends Ecol. Evol.* **31**, 940–952 (2016).

107. Van Der Valk, T. *et al.* The genome of the endangered Dryas monkey provides new insights into the evolutionary history of the vervets. *Mol. Biol. Evol.* **37**, 183–194 (2020).
108. Pérez-Pereira, N. *et al.* Long-term exhaustion of the inbreeding load in *Drosophila melanogaster*. *Heredity* **127**, 373–383 (2021).
109. Hansson, B., Morales, H. E. & van Oosterhout, C. Comment on “Individual heterozygosity predicts translocation success in threatened desert tortoises”. *Science* **372**, eabh1105 (2021).
110. Pérez-Pereira, N., Caballero, A. & García-Dorado, A. Reviewing the consequences of genetic purging on the success of rescue programs. *bioRxiv* 2021.07.15.452459 (2021) doi:10.1101/2021.07.15.452459.
111. Ralls, K., Sunnucks, P., Lacy, R. C. & Frankham, R. Genetic rescue: A critique of the evidence supports maximizing genetic diversity rather than minimizing the introduction of putatively harmful genetic variation. *Biol. Conserv.* **251**, 108784 (2020).

A direct critique of the view that genetic rescue should avoid the use of individuals from large populations.

112. von Seth, J. *et al.* Genomic insights into the conservation status of the world’s last remaining Sumatran rhinoceros populations. *Nat. Commun.* **12**, 2393 (2021).
113. Marsden, C. D. *et al.* Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc. Natl. Acad. Sci.* **113**, 152–157 (2016).
114. Kim, B. Y., Huber, C. D. & Lohmueller, K. E. Deleterious variation shapes the genomic landscape of introgression. *PLOS Genet.* **14**, e1007741 (2018).
115. Agrawal, A. F. & Whitlock, M. C. Mutation load: the fitness of individuals in populations where deleterious alleles are abundant. *Annu. Rev. Ecol. Evol. Syst.* **43**, 115–135 (2012).
116. Huber, C. D., Durvasula, A., Hancock, A. M. & Lohmueller, K. E. Gene expression drives the evolution of dominance. *Nat. Commun.* **9**, 2750 (2018).
117. Henn, B. M. *et al.* Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. *Proc. Natl. Acad. Sci.* **113**, E440 (2016).

Purifying selection at conserved genomic position is stronger in African compared to non-African populations, but many deleterious mutations evolved as nearly neutral during the Out-of-Africa expansion.

118. Boyko, A. R. *et al.* Assessing the evolutionary impact of amino acid mutations in the human genome. *PLOS Genet.* **4**, e1000083 (2008).
119. Rhie, A. *et al.* Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737–746 (2021).
120. Feng, S. *et al.* Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**, 252–257 (2020).
121. Lewin, H. A. *et al.* Earth BioGenome Project: sequencing life for the future of life. *Proc. Natl. Acad. Sci.* **115**, 4325 (2018).

122. Genereux, D. P. *et al.* A comparative genomics multitool for scientific discovery and conservation. *Nature* **587**, 240–245 (2020).
123. Armstrong, J. *et al.* Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).
124. Chen, J., Bataillon, T., Glémin, S. & Lascoux, M. Hunting for beneficial mutations: conditioning on SIFT scores when estimating the distribution of fitness effect of new mutations. *Genome Biol. Evol.* (2021) doi:10.1093/gbe/evab151.
125. Hemmings, N. L., Slate, J. & Birkhead, T. R. Inbreeding causes early death in a passerine bird. *Nat. Commun.* **3**, 863 (2012).
126. Kono, T. J. Y. *et al.* Comparative genomics approaches accurately predict deleterious variants in plants. *G3 GenesGenomesGenetics* **8**, 3321–3329 (2018).
127. Frazer, J. *et al.* Disease variant prediction with deep generative models of evolutionary data. *Nature* (2021) doi:10.1038/s41586-021-04043-8.
128. Tichkule, S. *et al.* VIVID: a web application for variant interpretation and visualisation in multidimensional analyses. *bioRxiv* 2021.11.16.468904 (2021) doi:10.1101/2021.11.16.468904.
129. Kyriazis, C. C., Wayne, R. K. & Lohmueller, K. E. Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. *Evol. Lett.* **5**, 33–47 (2021).
130. Bosse, M. No “doom” in chicken domestication? *PLOS Genet.* **15**, e1008089 (2019).
131. Makino, T. *et al.* Elevated proportions of deleterious genetic variation in domestic animals and plants. *Genome Biol. Evol.* **10**, 276–290 (2018).
132. Librado, P. *et al.* Ancient genomic changes associated with domestication of the horse. *Science* **356**, 442–445 (2017).
133. Rogers, R. L. & Slatkin, M. Excess of genomic defects in a woolly mammoth on Wrangel island. *PLOS Genet.* **13**, e1006601 (2017).

On the use of ancient DNA data to show the accumulation of detrimental mutations (deletions, stop codons) and retrogenes in a small populations.

134. Frankham, R. Genetics and extinction. *Biol. Conserv.* **126**, 131–140 (2005).
135. Chang, D. & Shapiro, B. Using ancient DNA and coalescent-based methods to infer extinction. *Biol. Lett.* **12**, 20150822 (2016).
136. Martiniano, R., Garrison, E., Jones, E. R., Manica, A. & Durbin, R. Removing reference bias and improving indel calling in ancient DNA data analysis by mapping to a sequence variation graph. *Genome Biol.* **21**, 250 (2020).
137. Günther, T. & Nettelblad, C. The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLOS Genet.* **15**, e1008302 (2019).
138. Marciniak, S. & Perry, G. H. Harnessing ancient genomes to study the history of human adaptation. *Nat. Rev. Genet.* **18**, 659–674 (2017).

139. Orlando, L., Gilbert, M. T. P. & Willerslev, E. Reconstructing ancient genomes and epigenomes. *Nat. Rev. Genet.* **16**, 395–408 (2015).
140. Pääbo, S. *et al.* Genetic analyses from ancient DNA. *Annu. Rev. Genet.* **38**, 645–679 (2004).
141. Flanagan, S. P., Forester, B. R., Latch, E. K., Aitken, S. N. & Hoban, S. Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. *Evol. Appl.* **11**, 1035–1052 (2018).

A practical roadmap for effectively implementing genomics into species management and conservation.

142. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919–925 (2014).
143. Heled, J. & Drummond, A. J. Bayesian inference of population size history from multiple loci. *BMC Evol. Biol.* **8**, 289 (2008).
144. Santiago, E. *et al.* Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium. *Mol. Biol. Evol.* **37**, 3642–3653 (2020).
145. Grace, M. K. *et al.* Testing a global standard for quantifying species recovery and assessing conservation impact. *Conserv. Biol.* (2021) doi:10.1111/cobi.13756.
146. Teixeira, J. C. & Huber, C. D. The inflated significance of neutral genetic diversity in conservation genetics. *Proc. Natl. Acad. Sci.* **118**, e2015096118 (2021).

A perspective piece arguing that loss of neutral genetic diversity does not predict extinction risk.

147. García-Dorado, A. & Caballero, A. Neutral genetic diversity as a useful tool for conservation biology. *Conserv. Genet.* **22**, 541–545 (2021).
148. DeWoody, J. A., Harder, A. M., Mathur, S. & Willoughby, J. R. The long-standing significance of genetic diversity in conservation. *Mol. Ecol.* **30**, 4147–4154 (2021).
149. Willoughby, J. *et al.* The reduction of genetic diversity in threatened vertebrates and new recommendations regarding IUCN conservation rankings. *Biol. Conserv.* **191**, 495–503 (2015).
150. Li, H. *et al.* Large numbers of vertebrates began rapid population decline in the late 19th century. *Proc. Natl. Acad. Sci.* **113**, 14079–14084 (2016).
151. Hansson, B. & Westerberg, L. On the correlation between heterozygosity and fitness in natural populations. *Mol. Ecol.* **11**, 2467–2474 (2002).
152. Harrisson, K. A., Pavlova, A., Telonis-Scott, M. & Sunnucks, P. Using genomics to characterize evolutionary potential for conservation of wild populations. *Evol. Appl.* **7**, 1008–1025 (2014).
153. Frankham, R. Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. *Mol. Ecol.* **24**, 2610–2618 (2015).
154. Rellstab, C., Dauphin, B. & Exposito-Alonso, M. Prospects and limitations of genomic offset in conservation management. *Evol. Appl.* **14**, 1202–1212 (2021).

155. Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M. & Keller, S. R. Genomic prediction of (mal)adaptation across current and future climatic landscapes. *Annu. Rev. Ecol. Evol. Syst.* **51**, 245–269 (2020).
156. Whiteley, A. R., Fitzpatrick, S. W., Funk, W. C. & Tallmon, D. A. Genetic rescue to the rescue. *Trends Ecol. Evol.* **30**, 42–49 (2015).
157. Haller, B. C. & Messer, P. W. SLiM 3: forward genetic simulations beyond the Wright–Fisher model. *Mol. Biol. Evol.* **36**, 632–637 (2019).
158. Frankham, R., Bradshaw, C. J. A. & Brook, B. W. Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biol. Conserv.* **170**, 56–63 (2014).
159. Lynch, M., Conery, J. & Burger, R. Mutation accumulation and the extinction of small populations. *Am. Nat.* **146**, 489–518 (1995).
160. Wallace, B. Fifty years of genetic load. *J. Hered.* **78**, 134–142 (1987).
161. Kimura, M., Maruyama, T. & Crow, J. F. The mutation load in small populations. *Genetics* **48**, 1303–1312 (1963).
162. Lynch, M. & Gabriel, W. Mutation load and the survival of small populations. *Evol. Int. J. Org. Evol.* **44**, 1725–1737 (1990).
163. Whitlock, M. C., Ingvarsson, P. K. & Hatfield, T. Local drift load and the heterosis of interconnected populations. *Heredity* **84**, 452–457 (2000).
164. Willi, Y., Griffin, P. & Van Buskirk, J. Drift load in populations of small size and low density. *Heredity* **110**, 296–302 (2013).
165. Bürger, R. & Lynch, M. Evolution and extinction in a changing environment: a quantitative-genetic analysis. *Evol. Int. J. Org. Evol.* **49**, 151–163 (1995).
166. Chevin, L.-M., Cotto, O. & Ashander, J. Stochastic evolutionary demography under a fluctuating optimum phenotype. *Am. Nat.* **190**, 786–802 (2017).
167. Kimura, M. & Crow, J. F. The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738 (1964).
168. García-Ramos, G. & Kirkpatrick, M. Genetic models of adaptation and gene flow in peripheral populations. *Evolution* **51**, 21–28 (1997).
169. Bolnick, D. I. & Nosil, P. Natural selection in populations subject to a migration load. *Evol. Int. J. Org. Evol.* **61**, 2229–2243 (2007).
170. Fierst, J. L. & Hansen, T. F. Genetic architecture and postzygotic reproductive isolation: evolution of Bateson-Dobzhansky-Muller incompatibilities in a polygenic model. *Evol. Int. J. Org. Evol.* **64**, 675–693 (2010).
171. Charlesworth, B. & Barton, N. H. Recombination load associated with selection for increased recombination. *Genet. Res.* **67**, 27–41 (1996).
172. Haag, C. R. & Roze, D. Genetic load in sexual and asexual diploids: segregation, dominance and genetic drift. *Genetics* **176**, 1663–1678 (2007).
173. Nunez, J. C. B. *et al.* Ecological load and balancing selection in circumboreal barnacles. *Mol. Biol. Evol.* **38**, 676–685 (2021).
174. Van Valen, L. M. A new evolutionary law. *Evol. Theory* **1**, l–30 (1973).
175. Crow, J. F. & Morton, N. E. The genetic load due to mother-child incompatibility. *Am. Nat.* **94**, 413–419 (1960).

176. Sandler, L. & Novitski, E. Meiotic drive as an evolutionary force. *Am. Nat.* **91**, 105–110 (1957).
177. Lindholm, A. K. *et al.* The ecology and evolutionary dynamics of meiotic drive. *Trends Ecol. Evol.* **31**, 315–326 (2016).
178. Gompel, N. & Prud'homme, B. The causes of repeated genetic evolution. *Dev. Biol.* **332**, 36–47 (2009).
179. Hansen, T. F. & Price, D. K. Age- and sex-distribution of the mutation load. *Genetica* **106**, 251–262 (1999).
180. Mee, J. A. & Yeaman, S. Unpacking conditional neutrality: genomic signatures of selection on conditionally beneficial and conditionally deleterious mutations. *Am. Nat.* **194**, 529–540 (2019).
181. Stone, J. L. Sheltered load associated with S-alleles in *Solanum carolinense*. *Heredity* **92**, 335–342 (2004).
182. van Oosterhout, C. A new theory of MHC evolution: beyond selection on the immune genes. *Proc. R. Soc. B Biol. Sci.* **276**, 657–665 (2009).
183. Brady, S. P. *et al.* Understanding maladaptation by uniting ecological and evolutionary perspectives. *Am. Nat.* **194**, 495–515 (2019).
184. Rice, W. & Chippindale, A. The evolution of hybrid infertility: Perpetual coevolution between gender-specific and sexually antagonistic genes. *Genetics* **116**, 179–188 (2002).
185. Hedrick, P. *Genetics of populations*. (Jones & Bartlett Learning, 2011).
186. García-Dorado, A. Understanding and Predicting the Fitness Decline of Shrunk Populations: Inbreeding, Purging, Mutation, and Standard Selection. *Genetics* **190**, 1461–1476 (2012).
187. Jinks, J. L. & Jones, R. M. Estimation of the components of heterosis. *Genetics* **43**, 223–234 (1958).

Acknowledgements

The authors thank Deborah Charlesworth and Armando Caballero for helpful comments on a previous version of the manuscript. CvO was supported by the Royal Society International Collaborations Award (ICA\R1\201194) and the Earth and Life Systems Alliance (ELSA). GB and FR were supported by the University of Ferrara (Italy). GB, FR, AI, and ET were funded by the MIUR PRIN 2017 grant 201794ZXTL to GB. MB was financially supported by the Dutch NWO-VENI grant n. 016.Veni.181.050. HEM was funded by an EMBO fellowship (grant 1111-2018) and the European Union's Horizon 2020 research and innovation program under a Marie Sklodowska-Curie grant (840519).

Author contributions

G.B., F.R., H.E.M. and C.v.O. researched the literature. G.B., F.R., E.T., H.E.M. and C.v.O. substantially contributed to discussions of the content. G.B., F.R., M.B., E.T., H.E.M. and C.v.O. wrote the article. All authors reviewed and/or edited the manuscript before submission.

Competing interests

The authors declare no competing interests.

Peer review information

Nature Reviews Genetics and the authors thanks Katerina Guschanski, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Table 1. Main approaches to predict deleteriousness from genomic data

	Principle	**Pros**	**Cons**	**Examples**
Evolutionary conservation	Mutations at sites with reduced number of substitutions compared to neutral expectations in multiple alignments are likely harmful	Annotation-free; allows direct comparisons between many species	Alignments among distant species could imply errors and missing regions; computationally intensive	GERP[31], PhyloP[28]
Basic annotation, physicochemical properties	Well-known functional effects of mutations in coding regions across species are used to classify a mutation	Alignment -free; simple to use and robust in classifying major classes of deleteriousness (e.g., non-synonymous, stop codons, loss of function)	Require basic annotation; focused on coding regions; numerical scores are not always predicted	SnpEff[32], Grantham scores[34]

Extended annotation	Diverse information (e.g., experimental evidence, physicochemical properties and evolutionary conservation) on the predicted harmful effects of variants is weighted and integrated into one metric	Exploit multiple information types	Still largely limited to humans, model and domesticated organisms where multiple data sources are available	PolyPhen[39], SIFT[38], ANNOVAR[33], VEP[40], CADD[42], GWAVA[43]

These approaches are designed principally to quantify the load due to SNPs. Other types of polymorphisms (for example, copy number variation, short tandem repeats, transposable elements) are not necessarily captured with these methods.

Figure 1. The effects of effective population size on the genetic load partition. Individual-based forward simulations in SLiM[157] of genetic load expressed in lethal equivalents (LEs), simulating a gamma distribution of selection (*s*) and dominance (*h*) coefficients that reflect empirical observations (see Supplementary Information SI3 for further details). Simulations with constant effective population size (N_e), with yellow dots and bars representing the mean and standard deviation across replicates, and blue dots indicating individual replicates. The three columns refer to populations at quasi-equilibrium (after a burn-in of $2N_e$ generations), and 4,000 and 8,000 generations later, respectively. Note that the x-axis shows the N_e rather than generation time, and that the LEs plotted on the y-axis are the quasi-equilibrium values of the load reached after $t=2N_e$ generations. The masked load increases with N_e and it reaches an equilibrium after circa $2N_e$ generations (top row). By contrast, the realized load continues to increase, particularly in small populations (middle row). Strong genetic drift across many loci overwhelms the purifying effects of selection in populations with small N_e, resulting in the gradual fixation of slightly deleterious mutations[158]. Such mutational meltdown generates a steady decline in fitness[20] and could even lead to extinction[159].

Figure 2. **The effects of demographic bottlenecks on the genetic load partition.** Individual-based forward simulations in SLiM[157] of genetic load expressed in lethal equivalents (LEs), simulating a gamma distribution of selection (*s*) and dominance (*h*)

coefficients that reflect empirical observations (see Supplementary Information SI3 for further details). The effects of demographic bottlenecks of different duration (t= 10, 100 or 500 years) on the genetic load dynamics (average generation is 2.6 years). Six stages are considered: first for $2N_e$ to reach quasi-equilibrium, followed by another 100 years at full N_e, 100 years of exponential collapse, a bottleneck stage (N_e=50) of different duration, 100 years of recovery and another 100 years after recovery. During a bottleneck, inbreeding and drift increase homozygosity, converting the masked load into the realized load. This load conversion has two main consequences. Firstly, purifying selection purge some (mostly highly) deleterious mutation, thereby reducing the genetic load[20–22]. Secondly, many (mostly mildly) deleterious mutations escape purifying selection, resulting in an overall increase of the realized load[20–22]. After demographic recovery, purifying selection reduces the realized load. However, a prolonged bottleneck results in a persistent realized load due to the fixation of deleterious mutations. These simulations illustrate the dynamics of load conversion and do not assess fitness or extinction risk[129]. Fitness and population viability of post-bottleneck populations are affected also by factors not simulated in this model, such as overall loss of adaptive variation, compensatory adaptive mutations, correlation between population density and fitness (Allee effect), environmental change, and loci under balancing selection[111].

Figure 3. Genetic load proxies used with whole genomes in wild animals ‘Categorical, basic’ refers to very simple annotations of a variant. ‘Categorical, deleteriousness’ refers to partitions where some direct and supported class of damage produced by a variant can be identified. ‘Numerical’ refers to scores with a continuous value for the impact of the substitution. For each category, we map in the lower section of the figure different statistics that are frequently used to summarize the scores in populations or individuals and obtain metrics related to the genetic load. When homozygous or heterozygous genotypes for a derived allele are identified, the ratio between the realized load and masked load can be calculated, thereby providing information about the population’s vulnerability to (future) inbreeding. Additional details and references are available in the main text and **Supplementary Table 1**. Abbreviations: ROH, runs of homozygosity; SFS, site frequency spectrum.

Box 1. Definitions of different types of load

Different types of load have been introduced in the literature (reviewed in[6,160]). Some definitions emphasize the processes giving rise to the load, whereas others were introduced to describe effects on genetic variation, fitness, or population persistence. Although these definitions are historically interesting and conceptually insightful, the differences can be confusing. In addition, these loads are often of limited practical relevance when analysing whole-genome sequence data of wild organisms because they lack the analytical framework to study them quantitatively.

[b1] Mutation (or mutational) load[4,115,161,162]

The reduction of fitness due to (recurrent) deleterious mutations, or the reduction of fitness at the mutation–selection equilibrium.

[b2] Drift load[159,163,164]

The reduction of fitness due to the increase of frequency (up to fixation: fixation load; during range expansion: expansion load) of deleterious mutations due to random genetic drift.

[b3] Evolution load (also referred to as evolutionary, transitory, lag or substitution load)[165,166]

The reduction of fitness due deleterious (sub-optimal, maladapted) mutations during the spread of superior adaptive variants.

[b4] Inbreeding load[15,17]

The reduction of fitness due to deleterious recessive mutations unmasked by inbreeding. The inbreeding load is the same as the masked load or potential load[17].

[b5] Segregation (or segregating) load, sometimes used as a synonym of balanced load[6,167]

The reduction of fitness due to segregating deleterious mutations; balanced load is the reduction of fitness due to mutations that are deleterious in both homozygous genotypes, but not in the heterozygote.

[b6] Migration load and hybrid load[168–170]

The reduction of fitness due to deleterious (maladapted) mutations introgressed from a different population or species after migration or hybridization.

[b7] Recombination load[171,172]

The reduction of fitness due to the breakup of favourable combinations of alleles at different loci due to recombination.

[b8] Other types of load

Ecological load[173], heterogeneous environment load[6], environmental load[174], incompatibility load[5,175], meiotic drive load[176,177], pleiotropic load[178], gametic load[179], non-local load[180], sheltered load[181–183] and gender load[184].

Box 2. Genetic load components

[b1] A population perspective

A population's genetic load at the gametic level is the sum of the selection coefficient s_i (the fitness reduction due to deleterious mutation *i*) of mutations at *L* loci, multiplied by their frequencies q_i[13];:

$$Genetic\ load(population) = \sum_{i=1}^{L} q_i\, s_i$$

[1]

The genetic load is independent of genotype frequencies and, hence, theoretically is not affected by inbreeding, random genetic drift or recombination. Only mutation and natural selection cause a directional change by increasing and decreasing the genetic load, respectively. By contrast, the portion of the genetic load that is realized (that is, whose fitness effects are expressed) does depend on genotype frequencies, as shown in equation 2. The first term expresses the homozygous effects, and the second term captures the effects of loci with heterozygous mutations, where h_i is the dominance coefficient of mutation *i*:

$$Realized\ load(population) = \sum_{i=1}^{L} q_i^2 s_i + 2\sum_{i=1}^{L} q_i[1-q_i]h_i s_i$$

[2]

The remaining part of the genetic load is not expressed and thus stays hidden from selection. Crow[6] referred to this as the inbreeding load, but we prefer to call this as the masked load considering that inbreeding tends to reduce this load by unmasking it. This terminology also avoids the confusion between a condition (i.e., the fitness effects of mutations that have remained masked) and a process (inbreeding):

$$Masked\ load(population) = \sum_{i=1}^{L} q_i\, s_i - \sum_{i=1}^{L} q_i^2 s_i - 2\sum_{i=1}^{L} q_i[1-q_i]h_i s_i$$

[3]

Inbreeding converts the masked load into the realized load, by increasing the frequency of homozygotes by $(1 - F)q_i^2 + Fq_i$, and by reducing the frequency of heterozygotes by $2(1 - F)\ q_i(1 - q_i)$, where *F* is the inbreeding coefficient[185].

Substituting these genotype frequency increases the realized load whilst reducing the masked load by the same amount:

$$Masked\ load\ (inbred\ population) = \sum_{i=1}^{L} q_i\, s_i - \left\{\sum_{i=1}^{L}(1-F)q_i^2 s_i + F q_i s_i\right\} - \left\{2\,(1-F)\sum_{i=1}^{L} q_i[1-q_i]h_i s_i\right\}$$

[4]

Equations [1-4] are visualized in Supplementary Figure S6.

[b2] An individual perspective

The different types of loads can also be calculated for individuals, which may be of practical use when individual genomes have been sequenced. The individual genetic load represents the potential fitness burden of mutations in the individual's genome, affecting its fitness and the fitness of its descendants. It is equal to the sum of the selection coefficients (s_i) across all loci *i* that are homozygous for the mutant allele, plus half the selection coefficients (s_j) of all heterozygous mutant loci *j* (see, for example,[69]):

$$Genetic\ load\ (individual) = \sum_{i=1}^{L(hom)} s_i + \sum_{j=1}^{L(het)} 0.5\, s_j$$

[5]

The factor 0.5 associated with the heterozygous loci in equation [5] expresses the probability that a given deleterious mutation at a heterozygous locus is passed on to an offspring (not the dominance coefficient).

The loss in fitness due to an individual's genetic load is captured by its realized load, and it is equal to the sum of all selection coefficient of all homozygous mutant loci, plus the sum of the product of the dominance coefficient (h_j) and selection coefficient of all heterozygous loci:

$$Realized\ load\ (individual) = \sum_{i=1}^{L(hom)} s_i + \sum_{j=1}^{L(het)} h_j s_j$$

[6]

Similarly, to equation [3], also at the individual level, part of the genetic load at the heterozygous sites is not expressed, and this constitutes an individual's masked load:

$$Masked\ load\ (individual) = \sum_{j=1}^{L(het)} (0.5 - h_j)s_j$$

[7]

This equation shows that additive mutations (*h*=0.5) do not contribute to the masked load, and indeed, these mutations do not contribute to inbreeding depression. This expression is also known as the purging coefficient, d = s(1/2 – h), defined by[186].

Under Hardy-Weinberg genotype frequencies, the values of the loads calculated using equations [5 – 7] and averaged over multiple individuals approximate to the population's values computed using equations [1], [2], and [3], respectively. Furthermore, equations [5 – 7] can be approximated by using average *s* and *h* coefficients instead of their sums (see **Supplementary Information SI4**).

Box 3. The load components under different conditions and the effect on fitness

Theoretically, a population could possess a large genetic load and still have a high fitness. For example, the genetic load of F1 hybrids of parents from two genetically diverged populations is mainly present as a masked load, and hence, the deleterious fitness effects of these mutations are not completely expressed (*cf.* heterosis[187]). The fitness of individuals is only affected by the realized load, and for fitness-related traits such as viability and survival, it can be calculated by multiplying the fitness effect across all homozygous and heterozygous loci:

$$Individual\ fitness = \prod_{i=1}^{L(hom)} (1 - s_i) \prod_{j=1}^{L(het)} (1 - h_j s_j)$$

[8]

Here, s_i and s_j are the selection coefficients of mutations in homo- and heterozygous loci, respectively, and h_j is the dominance coefficient. The mean population fitness can be approximated by taking the negative exponent of the realized load:

$$Mean\ population\ fitness \approx \left(1 - \left\{\bar{s}\bar{q}^2 + 2\bar{q}(1-\bar{q})\bar{h}\bar{s}\right\}\right)^L \approx e^{-\left\{\bar{s}\bar{q}^2 + 2\bar{q}(1-\bar{q})\bar{h}\bar{s}\right\}.L} \approx e^{-Realized\ Load}$$

[9]

Here, $\bar{s}$ and $\bar{h}$ are the mean selection and dominance coefficient averaged across loci, $\bar{q}$ the mean frequency of deleterious mutations, and *L* the number of loci carrying deleterious mutations. Both equation [8] and [9] ignore other (non-genetic) causes for fitness loss (*cf.* the intercept of the regression line in[5]), and they assume that the fitness effects are independent and act multiplicatively across loci. Furthermore, this equation assumes Hardy-Weinberg genotype frequencies and that (semi)lethal mutations are rare (see also **Supplementary information SI1 and SI2**).

GLOSSARY

Anchor species
A species that is added to the genome alignment between two evolutionarily distantly related taxa to help connecting them and facilitate the identification of conserved elements

Bottleneck (population bottleneck)
A sharp reduction in the effective population size over one or multiple generations.

Deleterious mutation
A poorly adapted genetic variant that reduces fitness relative to the wild type variant, thereby contributing to the genetic load. The mutations that remain deleterious across alternative environments are called “unconditionally deleterious”.

DNA sequence context
Extended patterns of DNA sequence composition, as di-nucleotide, triplet or longer patterns.

Effective population size
The number of individuals in an idealized population that shows the same amount of genetic drift (that is, random fluctuation of allele frequencies and loss of gene diversity) as the actual population.

Fitness
A measure of the capability of an individual or genotype to survive and reproduce.

Genetic rescue
Artificial (re)introducing of new (or rare) genetic variants into a population with the aim of reducing inbreeding depression, increasing genetic variation and population viability.

Inbreeding depression
The reduction of fitness due to breeding between relatives

Mutational spectrum
Rate of different types of DNA mutations in different sequence context

Ultra-conserved elements
Highly conserved genomic regions with (near) identical nucleotide sequences in evolutionarily distant taxa