The distribution of fitness effects varies phylogenetically across animals

Meixi Lin[1,2,a], Sneha Chakraborty[1], Carlos Eduardo G. Amorim[3], Sergio F. Nigenda-Morales[4,5], Annabel C. Beichman[6], Paulina G. Nuñez-Valencia[4,7,b], Jonathan Mah[8], Jacqueline A. Robinson[9], Christopher C. Kyriazis[1,10], Christian D. Huber[11], Andrew E. Webb[12,13], Sarah D. Kocher[12,13], Frederick I. Archer[14], Andrés Moreno-Estrada[4], Robert K. Wayne[1,c], and Kirk E. Lohmueller[1,8,a]

[1]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, California, 90095, USA
[2]Department of Integrative Biology, University of California, Berkeley, Berkeley, California, 94720, USA
[3]Biology Department, California State University, Northridge, California, 91330, USA
[4]Advanced Genomics Unit, National Laboratory of Genomics for Biodiversity (Langebio), Center for Research and Advanced Studies (Cinvestav), Irapuato, Guanajuato, 36824, Mexico
[5]Department of Biological Sciences, California State University San Marcos, San Marcos, California, 92096, USA
[6]Department of Genome Sciences, University of Washington, Seattle, Washington, 98195, USA
[7]Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México (UNAM), Cuernavaca, Morelos, 62210, Mexico
[8]Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, California, 90095, USA
[9]Institute for Human Genetics, University of California, San Francisco (UCSF), San Francisco, California, 94143, USA
[10]Conservation Genetics, San Diego Zoo Wildlife Alliance, Escondido, California, 92027, USA
[11]Department of Biology, Pennsylvania State University, University Park, Pennsylvania, 16802, USA
[12]Dept. of Ecology and Evolutionary Biology, Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey, 08544, USA
[13]Howard Hughes Medical Institute, Chevy Chase, Maryland, 20815, USA
[14]Marine Mammal and Turtle Division, Southwest Fisheries Science Center, La Jolla, California, 92037, USA
[a]Corresponding authors: meixilin@berkeley.edu; klohmueller@g.ucla.edu
[b]Current address: Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, 08003, Spain; Universitat Pompeu Fabra (UPF), Barcelona, Spain.
[c]Deceased

May 13, 2025

1 Abstract

The distribution of fitness effects (DFE) describes the selection coefficients (s) of newly arising mutations and fundamentally influences population genetic processes. However, the extent and mechanisms of DFE variation have not been systematically investigated across species with divergent phylogenetic histories and ecological functions. Here, we inferred the DFE in natural populations of eleven animal (sub)species, including humans, mice, fin whales, vaquitas, wolves, collared flycatchers, pied flycatchers, halictid bees, *Drosophila*, and mosquitoes. We find that the DFE co-varies with phylogeny, where the expected mutation effects are more similar in closely related species (Pagel's $\lambda = 0.84$, P = 0.01). Additionally, mammals have a higher proportion of strongly deleterious mutations (22% to 47% in mammals; 0.0% to 5.4% in insects and birds) and a lower proportion of weakly deleterious mutations than insects and birds. Population size is significantly negatively correlated with the expected impact of new deleterious mutations (PGLS$_\lambda$, P = 0.03), and the proportion of new beneficial mutations ($R^2_{\mathrm{adj}}= 0.73$, P < 0.001). These findings align with Fisher's Geometric Model (FGM), which defines organismal complexity as the number of phenotypes under selection. Consistent with the FGM's predictions, we observe that mutations are more deleterious in complex organisms, while beneficial mutations occur more frequently in smaller populations to compensate for the drift load. Our study demonstrates strong phylogenetic constraints in the evolution of a fundamental population genetics parameter, and proposes that, through mechanisms of global epistasis, long-term population size and organismal complexity drive variation in the DFE across animals.

2 Significance Statement

Understanding how mutations affect fitness is fundamental in evolution, but little is known about how and why the distribution of fitness effects (DFE) varies across species. In this study, we examine the DFE in diverse animal populations and show that closely related species exhibit similar patterns of mutation effects, with new mutations being more deleterious in mammals compared with birds and insects. Our findings corroborate Fisher's Geometric Model, which explains the variation in the DFE across species as a function of organismal complexity and long-term population size. By connecting organismal complexity and population size with the DFE, we offer a phylogenetic view into the selective forces shaping species adaptation and evolution.

Keywords: distribution of selection coefficients, Fisher's geometric model, deleterious variations, organismal complexity, effective population size

3 Introduction

Mutations are the raw material on which other evolutionary forces act. The distribution of fitness effects (DFE) represents the relative proportion of new mutations with different selection coefficients (s) affecting fitness of the individuals carrying them. In addition to its intrinsic value for understanding how mutations affect fitness, the DFE is a fundamental quantity in evolutionary genetics [1]. The DFE is essential for studying the maintenance of genetic variation, identifying causal variants underlying complex human disease, controlling for background selection in adaptive evolution, and understanding the long-term survival of small populations [1–6]. Consequently, considerable effort has been put into characterizing the biological factors underlying DFE evolution, and estimating the DFE for a given species of interest [1, 7]. However, the patterns of and mechanism behind the variation in the DFE have not been systematically surveyed across many species on a phylogenetic timescale [3].

Several theoretical models have proposed that organismal complexity and long-term population sizes (N_a) are pivotal biological factors influencing how the DFE might differ across species. However, while these two factors are common to many models, predictions for how they affect the DFE vary across studies [8–10]. While intuitively comprehensible, precise and comprehensive definitions of complexity remain elusive. For example, complexity can refer to the number of unique cell types and protein-protein interactions [11, 12]. Theoretically, complexity can refer to the number of genetically uncorrelated phenotypic traits under selection, denoting the dimensionality of phenotypic space in Fisher's Geometric Model (FGM; [13, 14]). We adopt this latter definition throughout this paper. Regardless, how complexity influences the DFE depends on the underlying model and associated assumptions. On the one hand, in complex organisms, highly connected networks confer robustness against the deleterious effects of mutations, therefore, more complex organisms are predicted to have fewer deleterious mutations [9, 15]. Alternatively, increased organismal complexity results in larger effect sizes for random mutations, predicting a higher proportion of deleterious mutations in complex organisms [7, 10, 16].

Similarly, the effect of long-term population size on the DFE is multifaceted. In species with larger effective population sizes, more efficient natural selection favors stable proteins that are then less susceptible to subsequent mutations that disrupt protein functions, thereby diminishing the average deleterious effects of new mutations [8, 17]. Conversely, the reduced selective strength in small populations leads to the accumulation of deleterious mutations, moving the population from the fitness optimum. In turn, an increased proportion of beneficial mutations is expected in species with smaller N_a [10], to compensate for the drift load. Taken together, the distribution of fitness effects at population scale ($2N_a s$) could be conserved across species as the interplay between selective strength and protein stability may offset variation in long-term population sizes [8].

Despite the growing number of published DFE in animal species, a lack of standardized DFE comparisons across lineages has hindered testing hypotheses regarding the DFE evolution [3]. In natural populations, the DFE is usually estimated by contrasting allele frequencies, summarized by the site frequency spectrum (SFS), of neutrally evolving and selected variants in genomic data sets [18, 19]. While DFE estimates for model organisms like humans [19, 20], *Drosophila* [18, 21], mice [7, 22], and nematodes [23] have been available, the DFEs for non-model organisms, such as flycatchers [24], great apes [25], Hawaiian monk seals [26], and vaquitas [27], are emerging on a per-lineage basis. These DFE estimates agree that most amino-acid changing mutations are nearly neutral ($-10^{-5} < s \leq 0$) or weakly deleterious ($-0.001 < s \leq -10^{-5}$), whereas beneficial ($s > 0$) and strongly deleterious ($s \leq -0.01$) mutations are less common [1]. Additionally, the DFE appears to vary across divergent species [24–27]. For example, humans

3

were found to carry a higher proportion of strongly deleterious mutations than *Drosophila* [7, 22]. Interestingly, closely related species demonstrated similar inferred DFE parameters [25, 28]. Unfortunately, the scarcity of high-quality genomic data and genetic parameter estimates (e.g. mutation rates) in non-model organisms [29], coupled with methodological variations and computational inefficiencies in DFE inference [20, 30, 31], have prevented a more comprehensive comparison.

The recent increase in available genomic resources for non-model organisms now offers unprecedented opportunities for comparative population genomics [29]. Leveraging such data, here we infer the DFE in eleven animal (sub)species with diverse phylogenetic relationships and life history traits using *varDFE*. We find strong signals of phylogenetic dependence in the evolution of DFE across species, with mammals bearing more deleterious mutations than insects and birds. To evaluate the biological processes behind the evolution of this fundamental population genetics parameter, we correlate proxies for organismal complexity and long-term population size with the expected mutation effects in different species. Mutations are more deleterious in more complex and smaller populations, suggesting patterns of macroscopic epistasis in DFE evolution.

4   Results

4.1   Population genetic inference for eleven animal (sub)species

We retrieved high-quality population-level polymorphism datasets for eleven animal (sub)species. These data include three insect species (mosquitoes, *Drosophila*, and halictid bees), two bird species (pied flycatchers and collared flycatchers), and six mammal (sub)species (arctic wolves, gray wolves, vaquitas, fin whales, mice, and humans; Table 1, Dataset S1). The eleven (sub)species we evaluated are distinctive in the potential biological features that affect the DFE, such as phylogenetic position (Fig. 1A), effective population size, and life history traits (Dataset S1). For example, the long-term population sizes (N_a) varied from 6264 (vaquitas) to 2.8×10^6 (*Drosophila*) individuals (Table 1). All datasets were filtered with the same stringency to retain only high-confidence genotype calls and at least eight diploid high-quality samples in coding regions (Text S1). We tallied the number of variants at different minor allele frequencies and generated folded site frequency spectra (SFS) in synonymous and nonsynonymous/missense regions (SYN-SFS and MIS-SFS, respectively) for each species (Fig. S1, Dataset S2). To implement a robust but flexible workflow to compare the DFE across species, we developed *varDFE* https://github.com/meixilin/varDFE, a python API based on the $\partial a \partial i$ and Fit$\partial a \partial i$ packages [20, 32] (Fig. S2).

To estimate long-term population size for each species, and control for population history and linked selection in the coding regions for downstream DFE inference [20, 31], we inferred demographic parameters from the putatively neutral SYN-SFS (Fig. S3, Dataset S3). A demographic model with a single change in population size (i.e. a two-epoch model) provided a good fit to SYN-SFS and reasonable parameter estimates for most species except for the mosquitoes, halictid bees, gray wolves, and collared flycatchers. For the gray wolves and collared flycatchers datasets, the three-epoch demographic model (i.e. two changes in population size) performed better with more realistic parameter estimates or drastically improved fit to the SYN-SFS. For the mosquitoes and halictid bees datasets, no demographic model based on the full SYN-SFS fit the data well, and using the singleton-masked SYN-SFS substantially improved model fit (Dataset S4, Text S1). All the inferred demographic models are qualitatively consistent with previous

4

139 estimates (Dataset S5). We utilized the best-fit demographic parameters from the chosen models (Table 1)

140 for downstream DFE inference.

4.2 Mutations are more deleterious in mammals compared with insects and birds

143 Conditional on the inferred demography, we estimated the DFE for new nonsynonymous mutations for

144 each species, assuming that mutations are neutral or deleterious ($s \leq 0$) and the DFE follows a gamma

145 distribution [20]. The gamma distribution yielded a robust fit to the observed MIS-SFS across all datasets

146 (Fig. S4, Dataset S6).

147   On average, mutations are 4.8 to 4923 times more deleterious in mammals (the expected selection

148 coefficient, $E[|s|]$ is 0.010 in mice to 0.67 in arctic wolves, n = 6), compared to insects ($E[|s|] = 0.00014$ in

149 *Drosophila* to 0.00056 in mosquitoes, n = 3) and birds ($E[|s|] = 0.00052$ in pied flycatchers and $E[|s|] =$

150 0.0021 in collared flycatchers; Fig. 1B). Noticeably, the scale (β') parameter for the arctic wolves reached

151 the upper boundary of the parameter space during inference. To rule out potential data artifacts, we

152 analyzed an independently generated whole-genome dataset from Russian Karelia gray wolves [33], which

153 also yield a high β' value, confirming $E[|s|]$ is very deleterious in wolves. The shape (α) parameter of the

154 gamma distribution differs between species groups as well. On average, mammals have lower estimates for

155 the shape parameter ($\alpha = 0.11$ in arctic wolves to $\alpha = 0.21$ in mice, n = 6) compared to insects and birds

156 ($\alpha = 0.22$ in collared flycatchers to $\alpha = 0.36$ in *Drosophila*, n = 5; Fig. 1C). The extent of variation in α

157 that we detected is in line with previous findings [28].

158   We also observed that mammals have much higher proportions of strongly deleterious mutations than

159 insects and birds (Fig. 1D). The maximum-likelihood gamma distribution for each species demonstrated

160 that 22% (vaquitas) to 47% (arctic wolves) of mutations in mammals are very strongly deleterious

161 ($|s| \geq 0.01$), compared to 0.0% (*Drosophila*) to 5.4% (collared flycatchers) in insects and birds. On the

162 other hand, insects and birds have a larger proportion of weakly deleterious mutations ($10^{-5} \leq |s| < 0.001$,

163 41% in collared flycatchers to 68% in *Drosophila*, n = 5), compared to mammals (16% in arctic wolves to

164 30% in mice, n = 6). The proportions of neutral to nearly neutral mutations ($0 \leq |s| < 10^{-5}$) are similar in

165 all eleven (sub)species. To directly estimate the proportion of (nearly) neutral mutations, we considered a

166 different functional form, where DFE followed a mixture of gamma distribution with an additional point

167 mass at neutrality (neugamma DFE [7], eq. 1, Dataset S7). This neugamma DFE improved model fit for

168 most datasets (lower AIC scores) compared to the gamma DFE (Figs. S5, S6), except for mosquitoes, gray

169 wolves, and vaquitas. Similar to the patterns obtained assuming a gamma DFE (Fig. 1D), the inferred

170 proportion of neutral mutations (p_{neu}) did not exhibit a phylogenetic pattern (Figs. 1E, S7). p_{neu} varied

171 from 0.081% (mosquitoes) to 16% (pied flycatchers) in insects and birds (n = 5), and from 0.089%

172 (vaquitas) to 28% (humans) in mammals (n = 6). We also attempted to directly estimate the proportion of

173 lethal mutations ($s = -0.5$) [20, 34]. However, our approach likely lacks sufficient power to confidently

174 infer the proportion of lethal mutations, due to poor model fit and unrealistic estimates of the proportion

175 of lethal mutations (Figs. S6-S8, Dataset S8, Text S2).

176   To evaluate the robustness of our analysis, we tested if different probability distributions for the

177 DFE, demographic models, or decisions to mask singletons in the SFS affected DFE inference (Fig. S9).

178 We examined the gamma, neugamma, and lognormal distributions (Fig. S10, Dataset S9) for the DFE.

179 When assuming the neugamma DFE, $E[|s|]$ remained more deleterious in mammals compared with insects

180 and birds. When assuming the lognormal DFE, the phylogenetic signal was less evident due to reduced

5

181 $E[|s|]$ estimates in mammals (Figs. S9A, S11). However, the AIC for the lognormal DFE is the lowest only

182 in four datasets (mosquitoes, gray wolves, arctic wolves, and vaquitas; Fig. S12), suggesting that the $E[|s|]$

183 derived from the lognormal DFE is less supported by the data. The gamma distribution was confirmed as

184 an adequate functional form for DFE comparisons, given its parsimony, consistent second-lowest AIC

185 across most datasets (Fig. S12), and good visual fit to the observed SFS (Fig. S4). Next, we tested the

186 potential impacts of demographic model misspecification and singleton-masking treatments. The average

187 mutation effects and proportion of each mutation category remain unchanged qualitatively (Figs. S9B-C,

188 S13, S14; Datasets S10, S11).

189   To account for the combinative effects of DFE and demographic model specifications, we generated

190 model-averaged $E[|s|]$ estimates that weigh outputs from different DFE and demographic models by their

191 AIC (Fig. 2A, Dataset S12). Using the full SFS, the model-averaged $E[|s|]$ is more deleterious in mammals

192 ($E[|s|] = 0.0034$ in humans to 0.28 in arctic wolves, n = 5) compared to insects ($E[|s|] = 2.2 \times 10^{-5}$ in

193 halictid bees to 3.7×10^{-5} in mosquitoes, n = 3) and birds ($E[|s|] = 0.00019$ in pied flycatchers and $E[|s|]$

194 $= 0.00099$ in collared flycatchers), with the exception that model-averaged $E[|s|]$ estimates in mice is now

195 0.00078, due to the improved fit of a more complex demographic models with extreme demographic events

196 (Text S3). The proportion of strongly deleterious mutations ($|s| \geq 0.01$) remains higher in mammals (21%

197 in fin whales to 56% in arctic wolves) except for the mice dataset (0.89%), compared to insects and birds

198 (0% in mosquitoes to 0.95% in *Drosophila*, n = 5, Figs. 2B, S15). We also evaluated the model-averaged

199 $E[|s|]$ in all datasets using the singleton-masked SFS, because no model fits the full SFS well in the

200 mosquito dataset ($min(\Delta LL) = 2018$). Using singleton-masked SFS, the model-averaged $E[|s|]$ is more

201 deleterious in mammals and birds ($E[|s|] = 0.0011$ in collared flycatchers to 0.06 in arctic wolves, n = 8),

202 compared to insects ($E[|s|] = 3.6 \times 10^{-5}$ in *Drosophila* to 0.00034 in mosquitoes; Figs. 2A, S15).

203   To confirm our DFE inference can confidently detect strongly deleterious mutations even in very

204 large populations, we performed a SLiM simulation using demographic parameters inferred from

205 *Drosophila* and DFE parameters inferred from humans. The DFE inferred from the simulation is similar to

206 humans (average $E[|s|] = 0.017$, n = 84), indicating our methods can confidently detect strongly

207 deleterious mutations in large populations (Fig. S16). Therefore, the lack of strongly deleterious mutations

208 inferred in species with large population sizes is likely not due to biases in our inference procedure, and

209 instead reflect true differences in the DFE. Overall, our DFE inference proved mostly robust to analytical

210 choices, with some exceptions noted in specific datasets (Text S3).

211   In summary, the distribution of fitness effects is more similar in closely related species, with

212 mammals harboring more strongly deleterious and less weakly deleterious variations compared with insects

213 and birds.

4.3   The DFE co-varies with phylogeny in animals

215 The apparent co-variation of the DFE with phylogeny motivated us to test for phylogenetic signal more

216 formally. Pagel's λ quantifies how trait evolution depends on the underlying phylogenetic tree across

217 species, where $\lambda = 0$ suggests no phylogenetic signal and $\lambda = 1$ indicates strong phylogenetic dependence

218 [35]. Assuming a gamma-distributed DFE, we calculated Pagel's λ for log10 transformed expected selection

219 coefficients [$\log_{10}(E[|s|])$], and the shape parameter (α, Fig. 1B-C). For $\log_{10}(E[|s|])$, Pagel's λ was 0.84 (P

220 $= 0.01$, likelihood ratio test [LRT]). For the shape parameter, Pagel's λ was 0.86 (P = 0.004, LRT). Since

221 the DFE parameters for arctic wolves reached upper boundaries, Pagel's λ was recalculated excluding the

222 arctic wolves dataset. Strong phylogenetic signal was consistently observed for both metrics ($\lambda = 0.86$, P =

0.008 for $\log_{10}(E[|s|])$; $\lambda = 0.81$, P = 0.01 for α). To acount for model uncertainties, we repeated Pagel's λ analysis for the model-averaged $E[|s|]$ estimates, and strong phylogenetic signal remains ($\lambda = 0.80$, P = 0.01 using full SFS; $\lambda = 0.74$, P = 0.02 using singleton-masked SFS). Both $\log_{10}(E[|s|])$ and α suggested a significant tendency that the evolution of DFE is influenced by the shared evolutionary history determined by the phylogenetic structure (Dataset S13).

To formally test whether the DFE varies across species, we examined a null model where the shape and scale parameters of the gamma distribution were constrained to be the same across the eleven (sub)species datasets, while still accounting for their individual demographic histories. The maximum likelihood estimate (MLE) for this null model is found by summing the individual likelihood surfaces of each species' DFE using a grid-search approach (Fig. 3A). The alternative model allows each species to have its own shape and scale parameters, as described in previous sections (Fig. 3B-C). This alternative model fits the data (MIS-SFS) significantly better than the null model (LRT, $\Lambda = 18203$, df = 20, P $< 10^{-16}$; Figs. 3A-C, S17). Therefore, the variation of DFE across taxa is statistically supported by our analyses.

To further investigate the relationship between phylogenetic proximity and DFE similarity, we next used the likelihood ratio test framework to compare pairs of species. Here, the null model is relaxed to constrain DFE parameters to be identical only for the two species compared, rather than across all species. The resulting LRT statistic (Λ) serves as a measure of DFE similarity between species pairs, with lower values indicating higher similarity (eq. 2). Although the hierarchical clustering of log10-scaled pairwise LRT statistics [$\log_{10}(\Lambda)$] did not fully recover the underlying phylogenetic relationships of the eleven (sub)species, insects and mammals clustered into two main lineages, the pied flycatchers fell into the insect lineage while the collared flycatchers clustered with vaquitas (Fig. 3D). In addition, $\log_{10}(\Lambda)$ is significantly lower (P = 0.02, Wilcoxon test) in pairwise comparisons within the same taxonomic class (average $\log_{10}(\Lambda) = 2.06 \pm 0.71$, n = 19), such as mammal-mammal pairs, compared to inter-class pairwise comparisons (average $\log_{10}(\Lambda) = 2.59 \pm 0.76$, n = 36), such as mammal-insect pairs (Fig. 3E). In summary, species that are more closely related phylogenetically, tend to have more similar DFEs compared to species that are more distantly related.

4.4 Population size and body mass are correlated with the evolution of the DFE

We next explored the relationship between the DFE and candidate life history traits related to long-term population size and organismal complexity, using the phylogenetic generalized least squares model (PGLS$_\lambda$ [36]). PGLS$_\lambda$ estimates the amount of phylogenetic signal (λ) in the regression residuals, and incorporates λ in regression parameter estimates, to control for interspecific autocorrelation due to phylogeny. As the long-term population size decreased (Fig. 4A), on average, mutations became more deleterious in a population ($\log_{10}(E[|s|]) \sim \log_{10}(N_a)$, PGLS$_\lambda$, slope (b) = -0.83, $\lambda = 0$, P = 0.03). Although significant correlation exists between N_a and $E[|s|]$ the log-likelihood surface of λ had another local optimum at $\lambda = 0.78$ ($LL_{\lambda=0.78} = -13.24$, $LL_{\lambda=0} = -13.10$), questioning the amount of phylogenetic signal in the correlation (Fig. S18). On the other hand, body mass was positively correlated with $E[|s|]$ (Fig. 4D, PGLS$_\lambda$, b = 0.26, $\lambda = 0$, P = 0.002). The two wolf (sub)species' DFE deviated the most from the regression, likely due to their very deleterious estimated DFEs. Other parameters we tested, including the generation time, age at maturity, maximum longevity, and mutation rates, were not significantly correlated with $E[|s|]$ after considering phylogenetic dependence (Fig. S19). To account for model uncertainty, we repeated PGLS$_\lambda$ analysis for the model-averaged $E[|s|]$ estimates. Significant correlation remains between

265   N_a and $E[|s|]$ (b = -0.98, $\lambda = 0$, P = 0.02 using full SFS; b = -0.88, $\lambda = 0$, P = 0.02 using singleton-masked

266   SFS; Fig. 4B-C), body mass and $E[|s|]$ (b = 0.31, $\lambda = 0$, P = 9×10^{-4} using full SFS; b = 0.20, $\lambda = 0$, P

267   = 0.02 using singleton-masked SFS; Figs. 4E-F, S19, S20). We repeated the PGLS$_\lambda$ analyses excluding the

268   arctic wolves outlier and for the shape (α) parameter (Text S4). Long-term population size and body mass

269   remains correlated with both $E[|s|]$ and α regardless of outlier treatments (Fig. S21, Dataset S14).

270   The negative correlation of population size and expected mutation effects at the individual level

271   suggests potential constancy for population-scaled mutation effects ($\gamma = 2N_a s$), an important prediction on

272   the DFE when considering protein thermostability [8]. Therefore, we next sought to test if population size

273   by itself can explain the variation in the DFE observed across species by examining the population-scaled

274   DFE. Assuming the population-scaled DFE follows a gamma distribution, the expected mutation effects

275   ($E[|2N_a s|]$) still vary across species (Fig. 4G-H). Compared to the co-variation of $E[|s|]$ with phylogeny, we

276   no longer observe similar $E[|2N_a s|]$ within lineages. The expected population-scaled mutation effects are

277   the highest in two wolves populations ($E[|2N_a s|]= 1.0 \times 10^5$ in arctic wolves and 8120 in gray wolves), and

278   the lowest in the vaquitas (161) and humans (195; Fig. 4G). The proportion of each type of mutation

279   fluctuated for each species as well (Fig. 4H). Adopting the grid-search analysis for population-scaled DFE

280   comparisons, we found that the alternative model where each species has its own population-scaled shape

281   and scale parameters fit the data significantly better than the null model assuming a constant γ in different

282   species (LRT, $\Lambda = 99508$, df = 20, P $< 10^{-16}$; Fig. S22). When comparing the population-scaled DFE

283   between species pairs, $\log_{10}(\Lambda)$ for γ remained significantly lower (P = 0.01, Wilcoxon test) in pairwise

284   comparisons within the same taxonomic class ($\log_{10}(\Lambda) = 2.98 \pm 0.62$, n = 19) compared to inter-class

285   comparisons (3.47 ± 0.75, n = 36; Fig. S23), as observed in the $\log_{10}(\Lambda)$ statistics for s (Fig. 3E).

286   Therefore, although population size is strongly correlated with the expected selection coefficient (Fig.

287   4A), it is not the sole predictor for the variation in the DFE across species (Figs. 4G-H, S22, S23).

4.5   Fisher's geometric model enables concurrent examination of population size and organismal complexity's influence on the DFE

290   Previous work has suggested that Fisher's geometric model (FGM) can explain patterns of DFE variation

291   across divergent species. FGM models the fitness landscape with a minimal set of parameters to study the

292   emergent properties of mutational effects [14]. In FGM, the fitness of an organism is characterized by a

293   multi-dimensional Gaussian decay function with a local optimum in n-dimension phenotype space.

294   Mutations create new phenotypes, modeled as random vectors with effect scale σ that act within a subset

295   of m phenotype dimensions ($m \leq n$) [10]. The phenotype dimensionality n is termed "organismal

296   complexity", and both m and σ could be a direct measurement of organismal complexity from different

297   perspectives [7, 16]. When the population is at the fitness optimum and mutations are universally

298   pleiotropic ($m = n$), the derived DFE is a gamma distribution with only neutral or deleterious mutations

299   ($s \leq 0$) [16]. When the population is under mutation-selection-drift equilibrium, the increase in fitness from

300   beneficial mutations should counteract the drift load, i.e. the decrease in fitness caused by fixed deleterious

301   mutations, and the population has an equilibrium phenotypic distance (z_{eq}; $z_{eq} > 0$) to the fitness

302   optimum. The formula for this FGM-derived DFE, which models the full DFE spectrum ($-0.5 \leq s \leq 0.5$),

303   is a function of three key parameters in FGM (eq. 3, derived in [10]): long-term population size (N_a),

304   mutation pleiotropy (m), and scale of mutation effects (σ).

305   We inferred the parameters for this FGM-derived DFE for eleven animal (sub)species (eq. 3). The

306   FGM-derived DFE provided an equally good or better fit to the MIS-SFS compared to the gamma DFE

<div align="center">8</div>

(Fig. S24, Dataset S15). In addition, the inferred long-term population sizes from the FGM-derived DFE largely agree with the estimates of N_a derived from demographic inference using genetic variation data (R^2_{adj}= 0.36, P = 0.03 linear regression; Fig. 5A). Both suggest that most populations are likely under mutation-selection-drift equilibrium and the FGM-based DFE describes their mutation effects well. The vaquita population's N_a estimate deviated the most between the two methods, likely reflective of its long-term small population size and recent bottleneck [27], thus departing from mutation-selection-drift equilibrium.

We found some support that the mutation effect scale (σ) is reflective of a species' phylogenetic position. Mammals have larger σ (0.077 to 9.5, n = 6) compared with insects and birds (0.0037 to 0.052, n = 5; Fig. 5B). The mutation pleiotropy (m) is negatively correlated with the inferred σ (R^2_{adj}= 0.76, P < 0.001, linear regression; Fig. S25), suggesting potential auto-correlations during inference. Although less conserved within lineages, mammals tend to have smaller m (0.17 to 0.60, n = 6) compared with insects and birds (0.49 to 4.5, n = 5). However, Pagel's λ did not suggest phylogenetic signal for either σ ($\lambda = 0$, P = 1) or m ($\lambda = 0.51$, P = 0.67; Dataset S13).

Next we examined the proportion of mutations with various ranges of s predicted from the FGM-derived DFE (Fig. 5D). Similar to the findings assuming the gamma DFE, strongly deleterious mutations are scarce in insects and birds populations (0.0% to 2.0% in collared flycatchers), but much more abundant in mammals (9.1% in mice to 55% in arctic wolves). Additionally, FGM predicts that as the population size decreases, the proportion of beneficial mutations increases to counteract the increased drift load [10]. Indeed, assuming FGM, smaller populations were found to harbor more beneficial mutations (R^2_{adj}= 0.73, P < 0.001, linear regression; Fig. 5C).

5 Discussion

Here, we evaluated the long-standing question of the evolutionary stability of the DFE across species, taking advantage of the advances in whole genome sequencing of non-model organisms. To do this, we curated high-quality whole-genome based polymorphism data with a large sample size per species. Our results suggest that the DFE evolution is contingent upon the underlying phylogenetic relationships across species, with closely-related species sharing a more similar DFE (Fig. 3). Overall, mutations are more deleterious in mammals, compared to insects and birds (Figs. 1, 2). We also observed strong correlations between the DFE with long-term population size and organismal complexity (Figs. 4, 5).

The DFE estimates we obtained align with those of previous studies (Dataset S16). In humans, the inferred gamma distribution shape parameter in our study (0.19 $\pm$ 0.0048), falls within the range of previous studies: from 0.12 $\sim$ 0.16 [28], 0.16 [25], 0.18 $\sim$ 0.21 [19], 0.17 $\sim$ 0.21 [20], to 0.2 [18]. In *Drosophila*, our inferred shape parameter of 0.36 $\pm$ 0.0017 aligns with previous estimates: from 0.35 [18], to 0.32 $\sim$ 0.41 [28]. Even in non-model organisms, our DFE inferences are concordant with published studies [22, 24, 27, 28]. The consistency of our DFE estimates with the literature further confirmed that the DFE variation observed across species is likely of biological relevance. Notably, we found an unusually high proportion of deleterious mutations in the DFE in two wolves species (Fig. 1B). As we currently lack a satisfying explanation for the distinct DFE in canids, it is worthy of further investigation [37]. Similar to prior research, the DFEs inferred in this study are restricted to nonsynonymous mutations in coding regions. We also assume all mutations are additive and future studies are needed to jointly infer the DFE with dominance [38–42], and to explore the DFE variations for non-coding and structural variants [43–45].

9

348   We found significant phylogenetic correlations during the evolution of the DFE in animals, (Pagel's λ

349 = 0.84, Fig. 1B), and the pairwise likelihood ratio test showed that the DFE is more similar in

350 species-pairs within the same taxonomic class (Fig. 3E). Several studies have suggested that the DFE is

351 similar in closely related species, but this phylogenetic signal has not been systematically evaluated until

352 now. For example, high levels of DFE correlation were reported between populations within species, in

353 humans, *Drosophila* and wild tomato [46], and between species within lineages, such as in great apes [25]

354 and *Populus* trees [47]. In addition, stark DFE differences have been consistently observed between species

355 with early divergences, such as humans and *Drosophila* [7, 18]. Previous work comparing the DFE across

356 diverse groups of animals had reported among-taxa variation [28, 48], similar to our findings. Chen et al.

357 [28] acknowledged the differences in DFE between invertebrates and vertebrates, whereas the uncertainties

358 for $E[|s|]$ in their study prevented further investigations. Although Galtier and Rousselle [48] considered

359 the strong group effect on the shape parameter to be a methodological artifact and deemed the DFE to be

360 the same across species, it has later been shown that their model's inclusion of lethal mutations flattened

361 the variations in entries of different SFSs, and was unsuitable for fitting a single gamma DFE to

362 multi-species data [49]. Recently, between-species, but not within-species, variation in the DFE was

363 reported for six forest tree species, corroborating the phylogenetic constraints we found in the animal DFEs

364 [50]. Emergent studies in comparative population genomics suggest that phylogenetic correlations exist not

365 only in mutation's fitness effects, as demonstrated here, but also in other fundamental population genetics

366 parameters, such as mutation rates and mutational spectra in vertebrates [51, 52]. Although our study did

367 not find correlations of mutation rate with $E[|s|]$ after correcting for phylogenetic signal (Fig. S19),

368 simulations have suggested a complex joint distribution of selection coefficients and mutation rates [53].

369 Future studies could leverage the broad phylogenetic context provided by macrogenetics to distinguish the

370 co-evolution of mutation rates and their effects, and to address core evolutionary hypotheses, such as

371 adaptive evolution and Lewontin's paradox [22, 29, 54].

372   Fisher's geometric model offers a simplistic yet versatile genotype-phenotype-fitness landscape to

373 explain the mechanism underlying phylogenetic signals in the DFE variation. In FGM, an organism's

374 phenotype is abstracted into a point in an n-dimensional space. Mutations move the phenotype with its

375 associated fitness, which decays with the distance to the phenotypic optimum [14]. Solutions for an

376 organism's DFE at mutation-selection-drift equilibrium can be derived under this model (eq. 3) and fit the

377 empirical data well (Fig. S24), suggesting that FGM's fitness landscape is appropriate for describing the

378 DFE across species. Additionally, our data aligned with two key predictions from the FGM-derived

379 equilibrium DFE regarding long-term population size (N_a) and organismal complexity (n). First, as N_a

380 decreases, a species' phenotype at mutation-selection-drift equilibrium is farther away from its fitness

381 optimum. The variance of the DFE increases, leading to the increased proportions of both beneficial and

382 strongly deleterious mutations in smaller populations (Fig. S26). Indeed, we observed that the proportion

383 of beneficial mutations increased in smaller populations (Fig. 5C). Moreover, mammals with lower N_a

384 exhibited a markedly higher proportion of highly deleterious mutations compared to insects and birds.

385 This trend is consistent across both the full DFE derived from the FGM (Fig. 5D), and the deleterious

386 component of the DFE modeled by the gamma distribution (Fig. 1D). Note that N_a by itself does not

387 explain all the variation in the DFE across species, because we did not observe a constant $E[|2N_a s|]$ across

388 populations (Fig. 4G-H). Second, the expected effect of mutations ($E[s]$) becomes more deleterious as

389 organismal complexity increases, because mutations are more likely to disrupt something important when

390 more dimensions are available in the phenotype space [7]. While biological complexity lacks a

straightforward definition, and the FGM's complexity parameter n (the dimensionality in the phenotype space) is not directly measurable, evidence suggests that mammals are more complex than insects. This is indicated by their greater number of unique cell types [55], larger and more complex genomes [56], and more extensive protein-protein interactions and transcriptional regulatory networks [57]. Although we cannot directly estimate n, the mutation effect scale (σ), which is correlated with n, can be inferred from the FGM-derived DFE [10, 16]. σ is also larger in mammals compared to insects and birds (Fig. 5B). Indeed, we discovered that mutations are on average more deleterious in the more complex mammal species, compared with insects (Fig. 1).

Beyond Fisher's geometric model, our findings that the DFE co-varies with ancestral population size and complexity can be interpreted under the broader concept of macroscopic or global epistasis. In other words, a mutation's fitness effect depends on the population's background fitness [58, 59]. Theoretically, a stable mutation-selection-drift equilibrium requires that the selective coefficients depend on the background fitness. To maintain stable fitness at equilibrium, the distribution of s needs to broaden as fitness declines [60]. With a log-concave fitness function, $w(z) = exp(-z^2)$, where z is the distance to phenotype optimum, the FGM has a built-in mechanism of epistasis [7, 14]. This leads to the prediction that the DFE variance increases as the equilibrium distance z increases in species with smaller ancestral population size and higher complexity. However, macroscopic epistasis is not exclusive to FGM, and can be observed from a neutral genotype network model [61] as well as first principles of population genetics. Natural selection is more efficient in larger populations than in smaller populations. As such, in larger populations, positive selection will be more effective at fixing mutations that increase protein stability, function, and robustness (Fig. 6A). When new deleterious mutations occur in large populations, these deleterious mutations occur in the optimal background and may have only weak effects. These are the mutations whose fitness effects have been measured in our present analysis. In smaller populations, however, selection will be unable to fine-tune the proteome (Fig. 6B), resulting in unstable proteins. When subsequent deleterious mutations occur in smaller populations, they occur in a less optimal background, and as such, they are more likely to be strongly deleterious. Evidence for global epistasis comes not only from DFEs derived from genetic variation data, as this study shows, but also from experimental findings. For example, the observation that larger populations have a lower proportion of beneficial mutations (Fig. 5C), is consistent with the widely observed diminishing-return epistasis in experimental microbial evolution, where the strength of beneficial mutations quickly declines as populations gain fitness in laboratory settings [62, 63]. Overall, the correlation in $E[|s|]$ with long-term population size is analogous to the drift-barrier hypothesis on mutation rate variation across large phylogenetic scales [64]. Our current results suggest that a similar drift-barrier exists for $E[|s|]$ (Fig. 6).

Our findings on the phylogenetic dependence of DFE variation have several practical implications. In conservation biology, genomics-informed simulations are increasingly popular to estimate the genetic health and quantify the deleterious load of small and fragmented populations [27, 65–67]. The relative conservation of the DFE parameters within closely-related lineages suggests that it is reasonable to assume DFE parameters from closely-related species when the study species' DFE parameter is not available. However, we caution against using constant DFE parameters across divergent lineages (e.g. mammals vs insects), as significant variation in the DFE exists. Our study also provides additional insights for adaptive evolution, by offering a phylogenetic constraint in the constant evolutionary processes of genetic drift and purifying selection, allowing for a more appropriate evolutionary null model for identifying adaptive variants [4].

11

434   Overall, we speculate that while N_a and its changes are similar in closely-related species due to
435   shared evolutionary histories, they are not inherently phylogenetically constrained. Long-term population
436   size and demographic events significantly influence a population's DFE by modifying the balance between
437   selection and drift, therefore altering the fitness background. Conversely, organismal complexity reflects the
438   physiological constraints imposed by phylogenetic structure, and probably governs the DFE at the
439   mutation-phenotype mapping level through molecular mechanisms, such as protein stability and functional
440   redundancy. Notably, we observed that mutations tend to be more deleterious in more complex mammal
441   species (Fig. 1), contradicting Kimura's classical proposition that more mutations are neutral in more
442   complex species [15]. In the future, we propose extending Fisher's geometric model from a single-species to
443   a multi-species model where organismal complexity across species is regulated by a hyperparameter that
444   adjusts the degree of phylogenetic dependence. In conclusion, our study demonstrated strong phylogenetic
445   correlations in the DFE variation across animals. Through mechanisms of global epistasis, long-term
446   population size and organismal complexity are likely underlying the phylogenetic dependence of the DFE.

6 Methods

6.1 Data sets and site frequency spectra

Polymorphism datasets in eleven animal (sub)species were compiled from various whole genome resequencing projects using a standardized curation workflow designed to retain only high-quality variants and samples for DFE inference (Text S1). Most datasets had pre-annotated variant effects. If not, we annotated the variant call files (VCF) using default options in snpEff (v.5.1 [68]). To retain only high-quality variants, we selected genomic sites that: 1. passed the filters set in the original dataset, 2. were in canonical coding region (CDS) but not in soft-masked, unknown or CpG islands of the genome, 3. were monomorphic or biallelic single nucleotide polymorphisms (SNP) that are annotated as "synonymous" or "nonsynonymous/missense". To retain only high-quality samples, we selected at least eight diploid samples for each species that: 1. had at least 8x average sequencing depth, 2. belonged to one natural population with no population substructure or admixture identified by principal component analysis (PCA), 3. were not related to other samples identified by kinship analysis.

To summarize polymorphism data, the obtained VCFs were further filtered to exclude genomic sites with more than 20% missing or 75% heterozygous genotypes. We projected down the sample size and computed folded synonymous and nonsynonymous/missense site frequency spectra (SYN-SFS and MIS-SFS, respectively) for each species. Computing a folded SFS mitigates uncertainties in ancestral state classifications. The projected sample size was calculated using the hypergeometric probability distribution to maximize the number of SNPs available and account for the sporadic missing genotypes (Dataset S1). The total synonymous and nonsynonymous sequence lengths (L_{SYN} and L_{MIS}) were obtained using previous estimates of nonsynonymous to synonymous mutation rate ratio $r_L = \frac{\theta_{\mathrm{MIS}}}{\theta_{\mathrm{SYN}}}$ [7].

To further understand the mechanisms of DFE evolution, we also compiled several life history traits for each species (sources and citations in Dataset S1).

6.2 Demographic inference

To control for how population size changes influence the SFS, we first inferred demographic parameters from the SYN-SFS using the `Demog1D_sizechangeFIM` module in *varDFE* (v.0.1.0, Fig. S2, based on $\partial a \partial i$ v.2.1.1 [20, 32] and python v.3.10.2). For each species, we fitted four models, including the standard neutral, two-epoch, three-epoch, and four-epoch models. To account for the uncertainties in genotype calls, we repeated the demographic inference by masking the singleton entries in the SYN-SFS. In total, eight demographic inference runs (four demographic models in SYN-SFS with or without masking singletons) for each species were conducted. All settings were the same across datasets except for the starting parameter positions, which were set based on knowledge from prior demographic inference (Datasets S1, S3). For each run, the best-fit parameters with the maximum multinomial log-likelihood in 100 replicates were chosen, and their uncertainties were estimated through Fisher's Information Matrix (FIM).

To choose the best-fit demographic model and singleton-masking treatment out of the eight runs per (sub)species for downstream DFE inference, we performed a step-wise likelihood-based model selection procedure (Fig. S27, Dataset S4, Text S1). At the start of model selection, the two-epoch demographic model based on full SYN-SFS run was chosen for all datasets. However, if it resulted in poor model fit, we chose a more complex demographic model or masked the singletons in the SYN-SFS. We calculated the population-scaled synonymous mutation rate θ_{SYN}, and ancestral population size using $N_a = \frac{\theta_{\mathrm{SYN}}}{4\mu L_{\mathrm{SYN}}}$ in the

13

487 chosen demographic run for downstream analyses (Table 1).

6.3 DFE inference

489 Conditional on the inferred demographic scenarios, we estimated the DFE for new nonsynonymous

490 mutations using methods from Fit$\partial a \partial i$ [20]. Briefly, Fit$\partial a \partial i$ takes advantage of the pattern that

491 deleterious mutations are less likely to segregate in the sample of individuals and those that are segregating

492 are more likely to be at lower frequency compared to neutral mutations. Assuming the MIS-SFS is under

493 selection while the SYN-SFS is putatively neutral, the best DFE parameters should fit the differences

494 between MIS-SFS and SYN-SFS in each minor allele frequency bin. We first computed and stored the

495 expected MIS-SFS for a range of population-scaled selection coefficients $\gamma = 2N_a s$ (-10000 $\leq \gamma \leq$ -10^{-5} and

496 $10^{-5} \leq \gamma \leq 100$) in each species, given their best fit demographic scenarios, using the `DFE1D_refspectra`

497 module in *varDFE*.

498   To infer the deleterious-only DFE, we assumed that the DFE follows a gamma distribution, given its

499 strong theoretical and empirical support [7, 16, 19]. We parameterized the DFE for each species as

500 $\gamma \sim Gamma(\alpha, \beta)$, where α ($\alpha > 0$; shape) and β ($\beta > 0$; scale) are parameters to be inferred. For each

501 species, 100 replicates were run from a permuted starting parameter of $\alpha = 0.2$ and $\beta = 4000$ using the

502 `DFE1D_inferenceFIM` module in *varDFE* [7]. The best-fit parameters with the maximum Poisson

503 log-likelihood were chosen, and the corresponding Akaike information criterion (AIC) was calculated. To

504 quantify uncertainty the parameter estimates, we Poisson resampled the MIS-SFS 200 times, and repeated

505 the DFE inference to obtain the 95% confidence interval [19, 20]. Recall that we optimized the gamma

506 distribution parameters for the population-scaled selection coefficient γ. To obtain the distribution of s, we

507 unscaled the gamma distribution by $1/(2N_a)$, therefore, $s \sim Gamma(\alpha, \beta')$, where $\beta' = \beta/2N_a$ and N_a had

508 been inferred from demography estimation. The expected selection coefficient $E[|s|]$ for each species was

509 calculated given the gamma distribution's property that $E[|s|] = \alpha\beta'$. Similarly, $E[|2N_a s|] = \alpha\beta$. To

510 compute the proportion of mutations with different values of s or $2N_a s$, we found the cumulative

511 probability in the given range using the `pgamma` function in R (v.3.6.2 [69]).

512   To estimate the proportion of effectively neutral mutations, we added a point mass at neutrality

513 ($p_{\text{neu}} \geq 0$) to the original gamma distribution. The DFE function can be written as

514 $\gamma \sim Neugamma(\alpha, \beta, p_{\text{neu}})$.

$$f(\gamma) = \begin{cases} \frac{p_{\text{neu}}}{10^{-5}} + (1 - p_{\text{neu}}) * Gamma(\gamma|\alpha, \beta), & 0 \leq |\gamma| < 10^{-5} \\ (1 - p_{\text{neu}}) * Gamma(\gamma|\alpha, \beta), & |\gamma| \geq 10^{-5} \end{cases} \tag{1}$$

515 The distribution of s is $s \sim Neugamma(\alpha, \beta', p_{\text{neu}})$, where $\beta' = \beta/(2N_a)$. The expected selection

516 coefficient $E[|s|]$ for each species is $E[|s|] = \frac{p_{\text{neu}}}{2*2N_a} * 10^{-5} + (1 - p_{\text{neu}})\alpha\beta'$.

517   To examine the robustness of DFE inference, we conducted additional analyses examining alternative

518 functional forms of the DFE, evaluating demographic model impacts, and assessing the effects of singleton

519 masking. For the lognormal distribution, we parameterized the DFE for each species as

520 $\gamma \sim Lognormal(\mu_s, \sigma^2)$. The distribution of s was unscaled as $s \sim Lognormal(\mu'_s, \sigma^2)$, where

521 $\mu'_s = \mu_s - ln(2N_a)$. $E[|s|]$ for each species was calculated using $E[|s|] = exp(\mu'_s)$. To account for uncertainty

522 in the functional form of the DFE, we performed DFE inference assuming three different functional forms

523 (gamma, neugamma, and lognormal) under both two-epoch and three-epoch demographic models, with

524 and without singleton masking. As masking singleton changes the data log-likelihood, for each species, we

14

525 generated separate model-averaged DFE estimates based on either the singleton-masked SFS or the full
526 SFS. The model-averaged DFE estimates were obtained by computing weighted-average in six combinations
527 of DFE and demographic models with weights determined by respective AIC values, as described in [25].

6.4 Phylogenetic signal detection

529 To evaluate whether the DFE co-varies with phylogeny, we calculated Pagel's λ for the log10 transformed
530 expected selection coefficient ($\log_{10}(E[|s|])$) and the shape parameter (α) in the gamma-distributed DFE
531 (Dataset S13) using the `phytools` package (v.1.2.0 [70]). The underlying phylogenetic structure and
532 divergence times for eleven (sub)species were compiled from the TimeTree database (http://timetree.org/).
533 To test if the candidate life history traits (X) are correlated with the observed DFE parameters (Y is $E[|s|]$
534 or α), and correct for phylogenetic dependency when appropriate, we utilized the PGLS_λ method in
535 `phytools`. The regression was specified as $\log_{10}(E[|s|]) \sim \log_{10}(X)$ or $\alpha \sim \log_{10}(X)$ to normalize the data.
536 The λ estimated in PGLS_λ does not equate to the Pagel's λ estimated individually for X or Y, but
537 represents whether the correlation of X and Y can be accounted for by phylogenetic dependence in residual
538 error. Candidate explanatory variables (X) include long-term population size (N_a), body mass in grams,
539 generation time in years per generation, age at maturity in days, maximum longevity in years and
540 mutation rates in mutations per bp per generation. To control for outliers and model uncertainties, we
541 repeated the PGLS_λ analysis excluding the arctic wolves data point, and on the model-averaged $E[|s|]$
542 estimates (Dataset S14).

543 To formally test for variation in the DFE across species, we used a grid-search approach
544 (`DFE1D_gridsearch` in *varDFE*) to compare each species' DFE estimates to a null model where the DFE
545 was constrained to be the same across species by likelihood-ratio tests (LRT). 2000 grid points were evenly
546 spaced in biologically meaningful ranges for shape ($\alpha = 0.1$ to 0.5) and scale ($\beta' = 0.0001$ to 0.5)
547 parameters. Using `integrate` in $\partial a \partial i$ we obtained the expected SFS for each $\alpha - \beta'$ pair and calculated the
548 Poisson log-likelihood relative to empirical MIS-SFS in each species. For the null model, we summed
549 log-likelihoods across species to find the MLE. To test whether the shape (α) and scale (β') are different in
550 any x number of species, the LRT was constructed as $\Lambda = -2 * ln(L_0/L_1) = -2 * (LL_0 - LL_1)$, where LL_0
551 is the log-likelihood for the null model, with two DFE parameters inferred
552 ($\hat{\alpha_1} = \hat{\alpha_2} = ... = \hat{\alpha_x}, \hat{\beta}'_1 = \hat{\beta}'_2 = ... = \hat{\beta}'_x$; params = 2) given the inferred demographic parameters for each
553 species ($\Theta_{D,1}, \Theta_{D,2}, ..., \Theta_{D,x}$), and LL_1 is the log-likelihood for the alternate model, where each species is
554 allowed to have its own DFE parameters ($\hat{\alpha_1}, \hat{\alpha_2}, \hat{\alpha_x}, ..., \hat{\beta}'_1, \hat{\beta}'_2, \hat{\beta}'_x$; params = 2x). Across the eleven
555 (sub)species tested, asymptotically, Λ should follow a χ^2 distribution with $df = 2x - 2 = 20$.

We also calculated the pairwise LRT statistics in all possible pairs of species ($x = 2$). For example, the LRT statistics comparing humans (population 1) and *Drosophila* (population 2) DFE can be written as follows:

$$\Lambda = -2 * ln[\frac{L(\widehat{\alpha_1 = \alpha_2}, \widehat{\beta'_1 = \beta'_2}|\Theta_{D,1}, \Theta_{D,2})}{L(\hat{\alpha_1}, \hat{\alpha_2}, \hat{\beta}'_1, \hat{\beta}'_2|\Theta_{D,1}, \Theta_{D,2})}] \tag{2}$$

556 Hierarchical clustering for $\log_{10}(\Lambda)$ was performed by the `hclust` function in R. To test the variations of
557 $\gamma = 2N_a s$ across species, we repeated the LRT by evenly spacing 2000 grid points in shape ($\alpha = 0.1$ to 0.5)
558 and population-scaled scale ($\beta = 100$ to 2.5×10^4) parameters.

6.5 Fitting the FGM-derived DFE

To investigate the genetic mechanisms underlying DFE variation and to incorporate beneficial mutations, we implemented a functional form of the DFE directly derived from the Fisher's geometric model [10]. The FGM proposes that fitness (w) can be described as a Gaussian function of the distance to the optimum (z), $w(z) = exp(-z^2)$, in an n-dimension phenotypic space. Random mutations likely affect a subset of m phenotypes $(m \leq n)$, with fitness effect size r following a zero mean Gaussian distribution with scale σ. The DFE for a well-adapted population can be described using the mutation pleiotropy (m), scale of mutation effects (σ) and long-term population size (N_a), $\gamma \sim Lourenco(m, \sigma, N_a)$. Here $\Gamma(.)$ is the gamma function and $K(.)$ is the modified Bessel function of the second kind [10].

$$f(s) = \frac{2^{\frac{1-m}{2}}\sqrt{N_a}(|s|)^{\frac{m-1}{2}}(1 + \frac{1}{N_a\sigma^2})^{\frac{1-m}{4}}exp(-N_as)}{\sqrt{\pi}\sigma^m\Gamma(\frac{m}{2})} \times K_{\frac{m-1}{2}}(N_a|s|\sqrt{1 + \frac{1}{N_a\sigma^2}}) \tag{3}$$

While complex in its expression, this FGM-derived distribution is functionally equivalent to the previously described gamma and lognormal distributions for generating mutational effects. Therefore, we estimated its parameters $(m, \sigma, \text{and } N_a)$ following methods described in the DFE inference section, through maximum-likelihood optimization by fitting the observed differences between MIS-SFS and SYN-SFS across minor allele frequency bins. To compute the proportion of mutations with different values of s, we found the cumulative probability for s in given ranges using the `scipy.integrate.quad` function (v.1.8.0) in python. The proportion of beneficial mutations was obtained by calculating the cumulative probability for $s > 0$. We performed two linear regressions using the `lm` function in R: first, $\log_{10}(N_a)$ inferred in this FGM-derived DFE and the $\log_{10}(N_a)$ obtained in demographic inference; second, the proportion of beneficial mutations and $\log_{10}(N_a)$. To test for phylogenetic signal in the inferred m and σ parameters, we calculated Pagel's λ (Dataset S13).

6.6 Simulating human DFE with drosophila demography

To investigate whether Fit$\partial a \partial i$ can accurately detect strongly deleterious mutations in species with large population sizes, we performed a simulation study using Drosophila demography and human DFE. We used the stdpopsim library to obtain coding sequence (CDS) annotations for *Drosophila melanogaster* chromosomes 2L, 2R, 3L, and 3R from the FlyBase_BDGP6.32.51_CDS [71]. CDS regions were labeled as "exon", and non-annotated regions as "bkgd". To simulate realistic genomic regions, we randomly sampled 84 (each 1 Mb) segments from these chromosomes, with probabilities proportional to chromosome lengths. Annotations for each segment were subset and converted to SLiM's coordinate format. Next, we implemented a two-epoch Drosophila demographic model in which the ancestral population size was set to 2,766,461, and the population expanded to 7,482,090 over 507,689 generations. We used a uniform recombination rate of 2.03806×10^{-8} which is an average of recombination rates for Drosophila chromosomes from the stdpopsim catalog. Mutation rate was set to 1.5×10^{-9} per bp per generation. Deleterious mutations followed a human-like gamma DFE (mean = –0.0276, shape = 0.189) with a nonsynonymous-to-synonymous ratio of 2.85. Background (noncoding) regions were assigned neutral mutations only. From each simulation, we output site frequency counts using a sample size of 100 and performed inference using methods described above. To reduce computational cost, we applied a scaling factor Q = 100, adjusting population size, mutation rate, recombination rate, and selection coefficients accordingly. We then inferred the DFE on each simulated dataset as done for the empirical data.

16

7   Acknowledgments

8   Funding

9   Author contributions

ML and KEL conceived the study. ML, CEGA, SN-M, ACB, PN-V, JAR, AEW, SDK, AME, and RKW generated the data. ML, CEGA, ACB, SC and JM performed the analysis of the data. SC and ML performed simulations. JM, CCK and CDH provided scripts for some analyses. FIA collected and contributed the samples and sample information. KEL performed funding acquisition. ML wrote the first draft of the manuscript. ML and KEL revised the manuscript with input from all the authors.

10   Data availability

The *varDFE* python package is available in this github repository: `https://github.com/meixilin/varDFE`. Additional scripts used to perform the data processing and analysis are available in this github repository: `https://github.com/meixilin/varDFE_analyses`. The raw sequence data for Arctic wolves and halictid bees will be available from NCBI's Sequence Read Archive database website upon publication. All other raw sequence data used in this analysis is already publicly available. Additional supportive information will be available from Zenodo upon publication.

17

11 Tables

Common name	Scientific name	Class	Demography	SFS mask	N_a	$E[|s|]$	α	β'
mosquito	*Anopheles coluzzii*	Insecta	three-epoch	TRUE	2286875	0.00056	0.30	0.0019
fruit fly	*Drosophila melanogaster*	Insecta	two-epoch	FALSE	2766461	0.00014	0.36	0.00038
halictid bee	*Lasioglossum calceatum*	Insecta	two-epoch	TRUE	280467	0.00051	0.30	0.0017
pied flycatcher	*Ficedula hypoleuca*	Aves	two-epoch	FALSE	246918	0.00052	0.29	0.0018
collared flycatcher	*Ficedula albicollis*	Aves	three-epoch	FALSE	138027	0.0021	0.22	0.0097
gray wolf	*Canis lupus lupus*	Mammalia	three-epoch	FALSE	45517	0.089	0.12	0.72
arctic wolf	*Canis lupus arctos*	Mammalia	two-epoch	FALSE	74139	0.67	0.11	6.30
vaquita	*Phocoena sinus*	Mammalia	two-epoch	FALSE	6264	0.013	0.13	0.096
fin whale	*Balaenoptera physalus*	Mammalia	two-epoch	FALSE	12411	0.027	0.14	0.19
house mouse	*Mus musculus castaneus*	Mammalia	two-epoch	FALSE	207682	0.010	0.21	0.048
human	*Homo sapiens*	Mammalia	two-epoch	FALSE	7041	0.014	0.19	0.073

Table 1: Summary of DFE inference in eleven (sub)species' datasets, assuming a gamma-distributed DFE for new neutral or deleterious mutations ($s \leq 0$). "Demography" is the demographic model of choice used for long-term population size inference. "SFS mask" reports if singletons are masked in the site-frequency spectrum during the inference. "N_a" is the inferred long-term population size in each dataset. "$E[|s|]$" is the expected mutation effects ($s \leq 0$). "α" and "β'" are the inferred shape and scale parameters respectively in the gamma DFE.
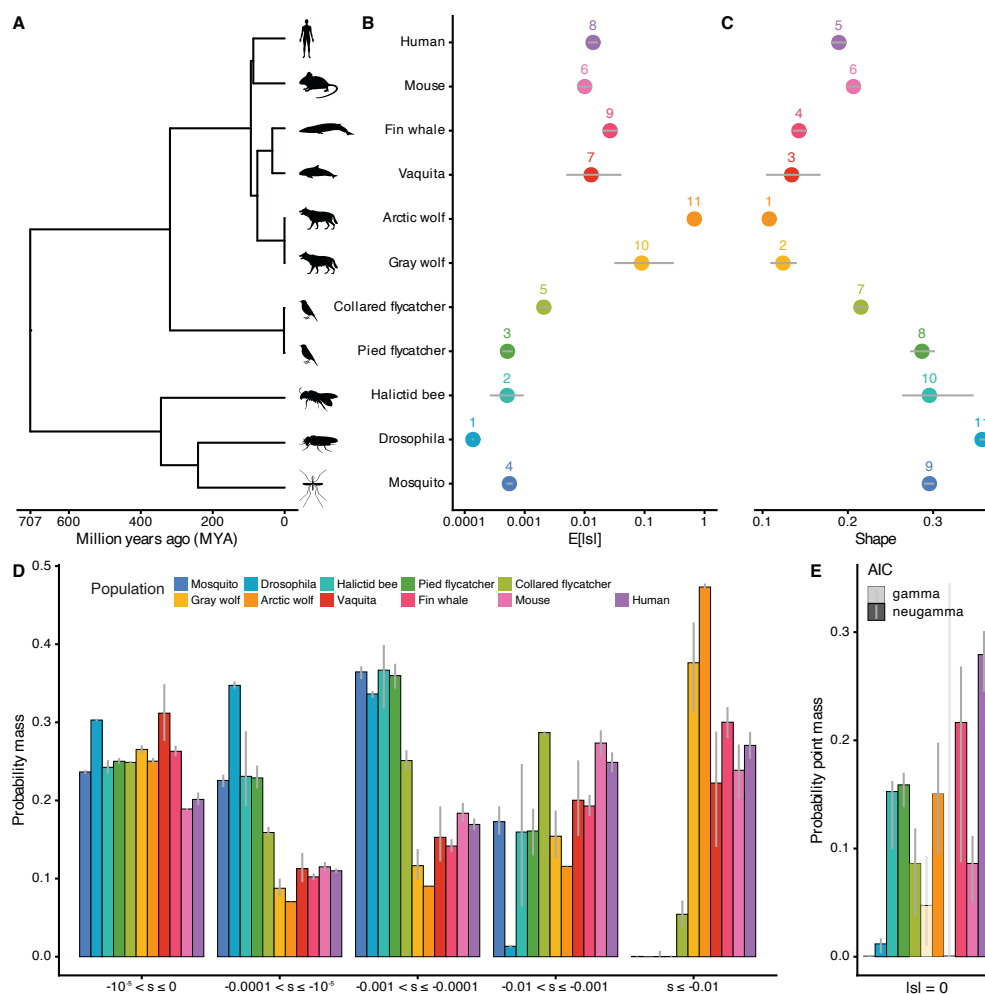
12 Figures

Figure 1: Mutations are more deleterious in mammals compared to insects and birds. (A) The phylogenetic relationship among the (sub)species included in this study (n = 11). (B) Assuming a gamma-distributed DFE of neutral and deleterious mutations, the expected deleterious mutation effects ($E[|s|]$) for eleven (sub)species. From top to bottom, species are increasingly divergent from humans: mice, fin whales, vaquitas, arctic wolves, gray wolves, collared flycatchers, pied flycatchers, halictid bees, *Drosophila* and mosquitoes. Numbers show the ranks of $E[|s|]$. (C) The shape parameter (α) in the gamma-distributed DFE for eleven (sub)species. Numbers show the ranks of α. (D) Proportions of mutations in various categories of $|s|$. From left to right, mutations range from (nearly) neutral ($-10^{-5} < s \leq 0$) to very strongly deleterious ($s < -0.01$). (E) The inferred probability point mass at $s = 0$ (p_{neu}) assuming a gamma-distributed DFE with a neutral point mass (neugamma DFE). In species where the Akaike Information Criterion (AIC) for the neugamma DFE is higher than that for the gamma DFE, which are mosquitoes, gray wolves and vaquitas, the bars are marked as translucent. (B-E) Gray lines represent 95% confidence intervals derived from 200 Poisson-resampled SFS.
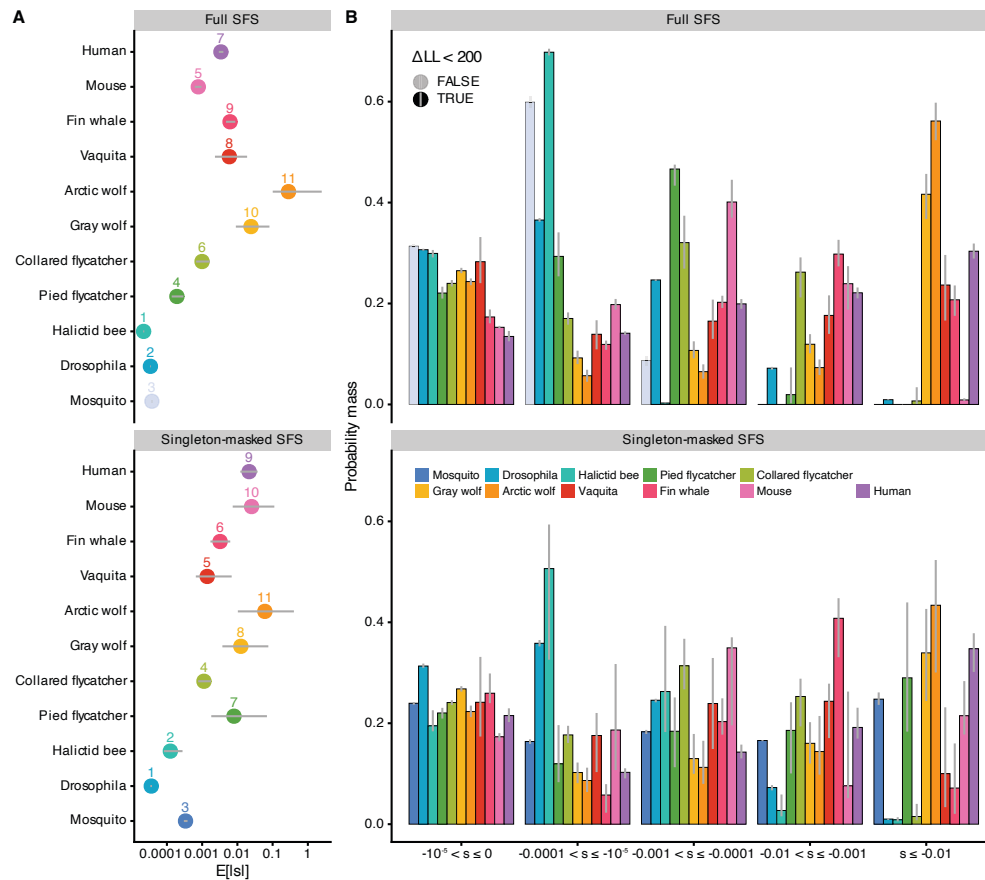
Figure 2: Model-averaged DFE estimates using full SFS or singleton-masked SFS. The estimates are weighted-averages of six DFE and demographic inference runs per species, combining outputs from gamma-, neugamma-, lognormal-distributed DFE given two-epoch or three-epoch demographic models by their AIC. Panels show whether full SFS or singleton-masked SFS is used in the inference. For results corresponding to singleton-masking treatments in Fig. 1, refer to Fig. S15. (A) The model-averaged expected deleterious mutation effects ($E[|s|]$) for eleven (sub)species. Numbers show the ranks of $E[|s|]$. (B) The model-averaged proportions of mutations in various categories of $|s|$. From left to right, mutations range from (nearly) neutral ($-10^{-5} < s \leq 0$) to very strongly deleterious ($s < -0.01$). (A-B) Gray lines represent 95% confidence intervals.
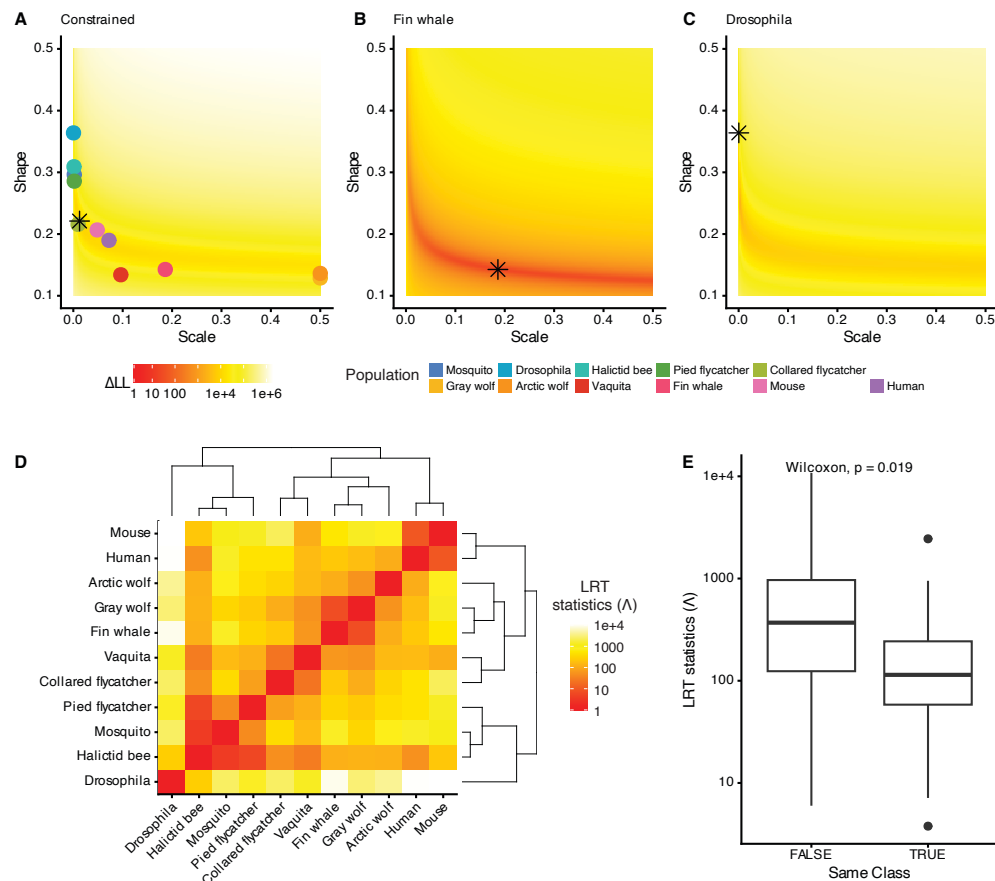
Figure 3: Grid-search-based likelihood-ratio tests confirm the phylogenetic signal in DFE evolution. (A-C) The log-likelihood surfaces for the shape (α) and scale (β') parameters (A) under the null model where all species are constrained to the same parameters or under the alternative model allowing species' parameters to vary in representative (B) fin whales and (C) *Drosophila* datasets (other species' log-likelihood surfaces shown in Fig. S17). Background colors from yellow to red indicate the differences in log-likelihood for given parameters to data. On each log-likelihood surface, the maximum likelihood estimate (MLE) derived for (A) the null model or (B-C) the alternative model in respective species is overlaid as the asterisk. In (A), the colored points showing MLEs for each species derived from the alternative model are overlaid for comparison. The two wolves datasets' MLEs derived from parameter optimization procedures (Fig. 1) exceed the grid-search derived MLE plotted here. For all other species, the MLEs from the grid search and parameter optimization are the same. (D) Pairwise LRT statistics (Λ) for each species pair is colored on a log10 scale and hierarchically clustered. Darker cells represent more similar DFE estimates in the species compared, such as the human-mouse pair, whereas lighter cells represent more distinct DFEs, such as the human-*Drosophila* pair. The dendrogram derived from hierarchical clustering is annotated. (E) The boxplot of pairwise LRT statistics in (D), categorized by whether species pair belong to the same Class. The y-axis is in log10 scale.
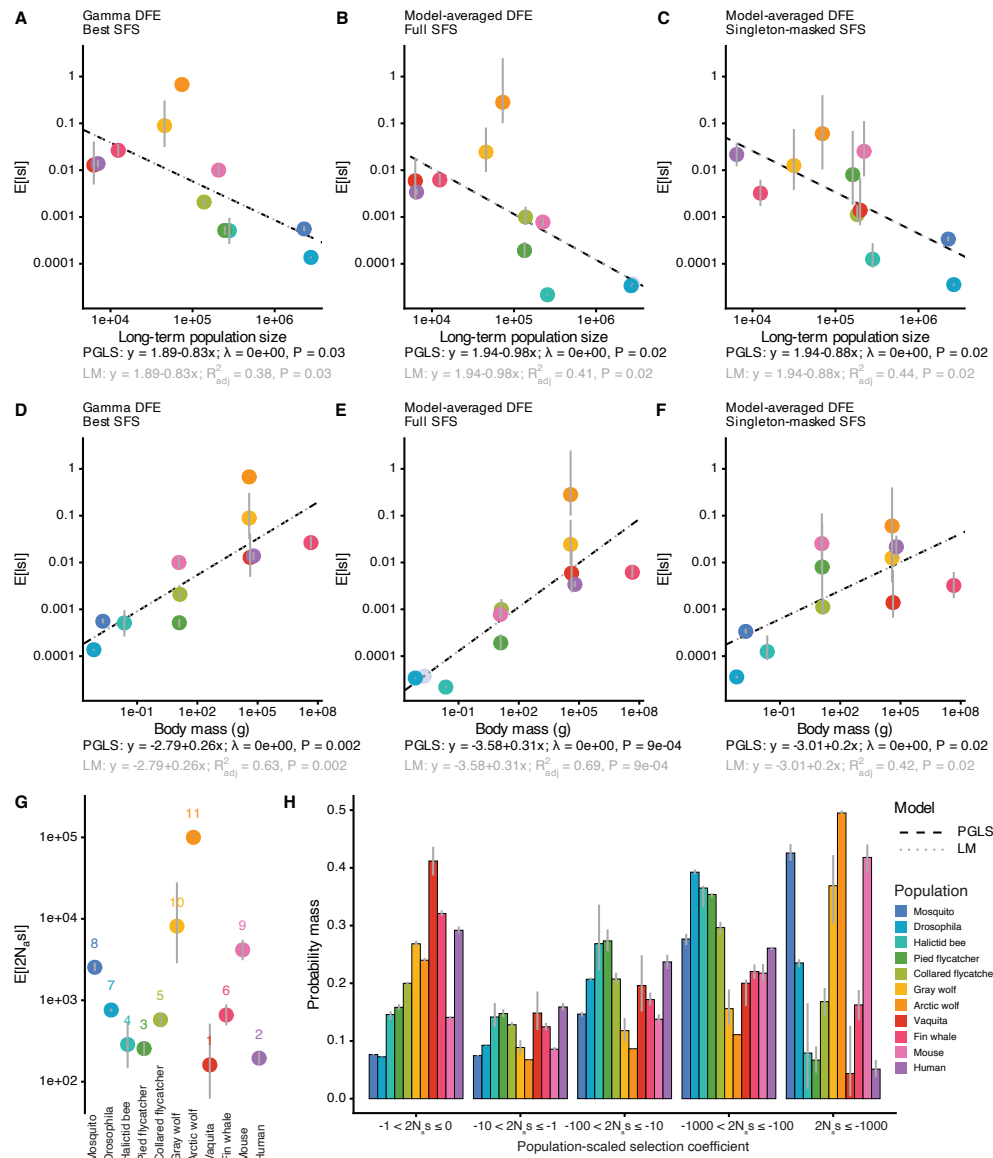
21

Figure 4: Long-term population size (N_a) and body mass are correlated with the expected mutation effects ($E[|s|]$), but N_a does not explain all the variations in the DFE across species. The phylogenetic generalized least square model with simultaneously inferred phylogenetic signal (PGLS$_\lambda$) was fitted for $E[|s|]$ with (A-C) long-term population size, (D-F) body mass, and other life history traits (Figs. S19, S20). $E[|s|]$ were computed from (A,D) gamma DFE, (B,E) model-averaged DFE using full SFS or (C,F) model-average DFE using singleton-masked SFS. The PGLS$_\lambda$ results are overlaid as a dashed black line, with equations, inferred phylogenetic signal (λ), and likelihood-ratio derived p-value annotated at the bottom. The linear regression results are overlaid as dotted gray line, with equations and p-values annotated at the bottom. (G) Assuming a gamma-distributed DFE, the expected population-scaled deleterious mutation effects ($E[|2N_a s|]$) are not conserved across eleven species. Numbers show the ranks of $E[|2N_a s|]$. (A-G) Each point represents one species, with gray vertical lines representing confidence intervals. All axes are in log10 scale. (H) Proportions of mutations with various ranges of $|2N_a s|$. From left to right, mutations ranged from (nearly) neutral ($-1 < 2N_a s \leq 0$) to very strongly deleterious ($2N_a s \leq -1000$). Gray lines represent 95% confidence intervals.
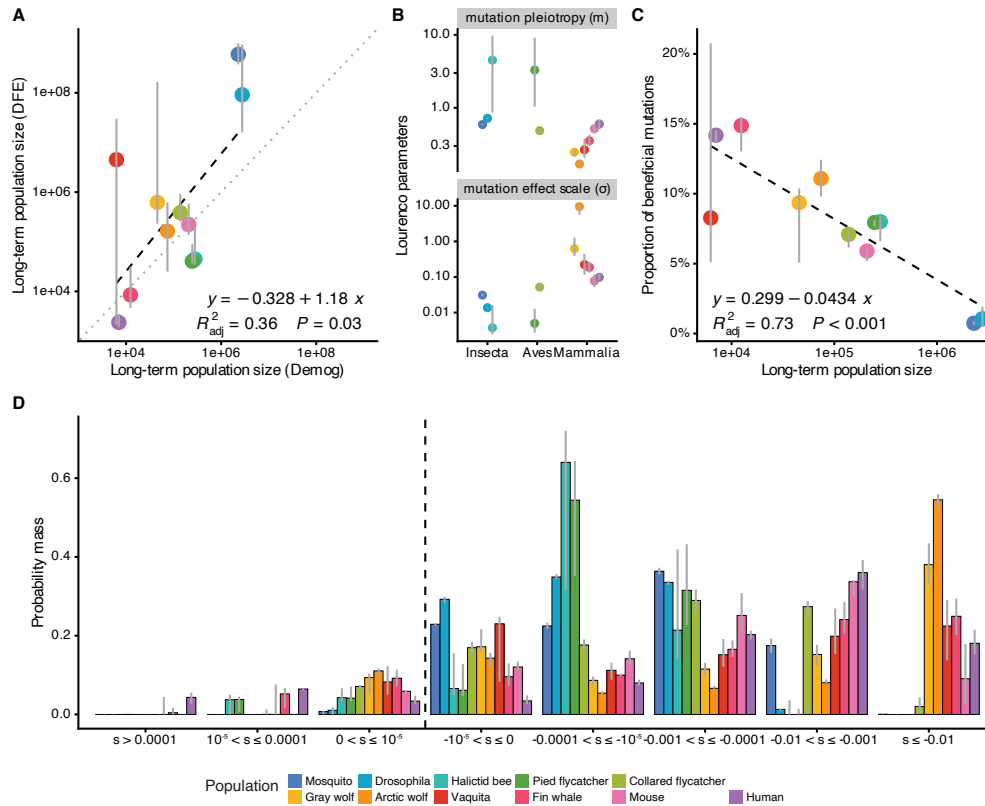
22

Figure 5: Fitting a Fisher's Geometric Model (FGM) derived DFE. (A) The inferred long-term effective population size (N_a) from the FGM-derived DFE (y-axis) is correlated with the effective size inferred from genetic variation data using $\partial a \partial i$ (x-axis). Axes are on a log10 scale. (B) The estimated scale of mutation effects (σ) and mutation pleiotropy (m) parameters for eleven (sub)species, grouped by the Class of each species. Note that σ is higher in mammals compared to insects. The y-axis is on a log10 scale. (C) The estimated proportion of beneficial mutations ($s > 0$) is negatively correlated with long-term population size inferred from $\partial a \partial i$. The x-axis is on a log10 scale. (D) Proportions of mutations in various categories of s estimated from the FGM-derived DFE for each species. From left to right, mutations range from beneficial ($s > 0$), (nearly) neutral ($-10^{-5} < s \le 0$) to very strongly deleterious ($s \le -0.01$). Dashed line marks $s = 0$. Gray lines represent 95% confidence intervals.
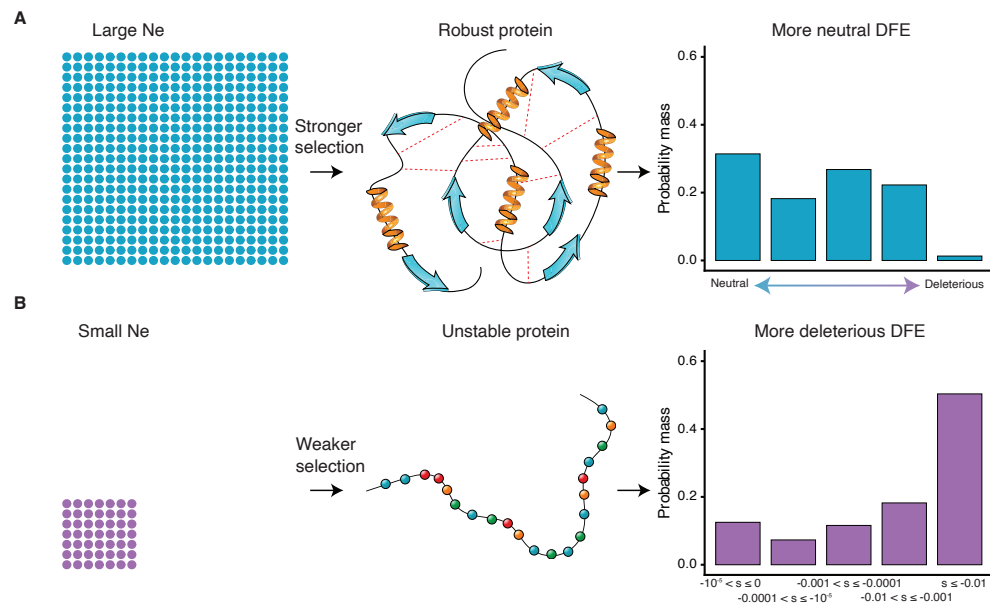
Figure 6: The drift-barrier of DFE evolution. (A) In large populations, efficient natural selection results in robust proteins. Because proteins are robust, subsequent deleterious mutations, whose effects we measured in our study, are only weakly deleterious. The inferred DFE becomes more neutral. (B) In smaller populations, natural selection is less effective at optimizing proteins. Thus, subsequent deleterious mutations, whose effects we measured in our study, are more strongly deleterious. The inferred DFE becomes more deleterious. Proportions of mutations in various categories of $|s|$ is generated assuming a gamma-distributed DFE, with the same population-scaled shape and scale parameters of $\alpha = 0.2, \beta = 5000$, but different N_a of 10^6 in (A) and 10000 in (B).

References

[1] Adam Eyre-Walker and Peter D. Keightley. "The Distribution of Fitness Effects of New Mutations". In: *Nature Reviews Genetics* 8.8 (2007), pp. 610–618. DOI: 10.1038/nrg2146.

[2] Adam Eyre-Walker. "Genetic Architecture of a Complex Trait and Its Implications for Fitness and Genome-Wide Association Studies". In: *Proceedings of the National Academy of Sciences* 107 (suppl_1 2010), pp. 1752–1756. DOI: 10.1073/pnas.0906182107.

[3] Jun Chen et al. "What Does the Distribution of Fitness Effects of New Mutations Reflect? Insights from Plants". In: *New Phytologist* 233.4 (2022), pp. 1613–1619. DOI: 10.1111/nph.17826.

[4] Parul Johri, Brian Charlesworth, and Jeffrey D Jensen. "Toward an Evolutionarily Appropriate Null Model: Jointly Inferring Demography and Purifying Selection". In: *Genetics* 215.1 (2020), pp. 173–192. DOI: 10.1534/genetics.119.303002.

[5] Jacqueline Robinson et al. "Deleterious Variation in Natural Populations and Implications for Conservation Genetics". In: *Annual Review of Animal Biosciences* 11 (Volume 11, 2023 2023), pp. 93–114. DOI: 10.1146/annurev-animal-080522-093311.

[6] Christopher C. Kyriazis, Jacqueline A. Robinson, and Kirk E. Lohmueller. "Using Computational Simulations to Model Deleterious Variation and Genetic Load in Natural Populations". In: *The American Naturalist* 202.6 (2023), pp. 737–752. DOI: 10.1086/726736.

[7] Christian D. Huber et al. "Determining the Factors Driving Selective Effects of New Nonsynonymous Mutations". In: *Proceedings of the National Academy of Sciences* 114.17 (2017), pp. 4465–4470. DOI: 10.1073/pnas.1619508114.

[8] Richard A. Goldstein. "Population Size Dependence of Fitness Effect Distribution and Substitution Rate Probed by Biophysical Model of Protein Thermostability". In: *Genome Biology and Evolution* 5.9 (2013), pp. 1584–1593. DOI: 10.1093/gbe/evt110.

[9] Mark L. Siegal and Jun-Yi Leu. "On the Nature and Evolutionary Impact of Phenotypic Robustness Mechanisms". In: *Annual Review of Ecology, Evolution, and Systematics* 45.1 (2014), pp. 495–517. DOI: 10.1146/annurev-ecolsys-120213-091705.

[10] João Lourenço, Nicolas Galtier, and Sylvain Glémin. "Complexity, Pleiotropy, and the Fitness Effect of Mutations". In: *Evolution* 65.6 (2011), pp. 1559–1571. DOI: 10.1111/j.1558-5646.2011.01237.x.

[11] Christoph Adami. "What Is Complexity?" In: *BioEssays* 24.12 (2002), pp. 1085–1094. DOI: 10.1002/bies.10192.

[12] Michael Lynch. "The Frailty of Adaptive Hypotheses for the Origins of Organismal Complexity". In: *Proceedings of the National Academy of Sciences* 104 (suppl_1 2007), pp. 8597–8604. DOI: 10.1073/pnas.0702207104.

[13] Olivier Tenaillon et al. "Quantifying Organismal Complexity Using a Population Genetic Approach". In: *PLOS ONE* 2.2 (2007), e217. DOI: 10.1371/journal.pone.0000217.

[14] O. Tenaillon. "The Utility of Fisher's Geometric Model in Evolutionary Genetics". In: *Annual Review of Ecology, Evolution, and Systematics* 45.1 (2014), pp. 179–201. DOI: 10.1146/annurev-ecolsys-120213-091846.

[15] M. Kimura. "Model of Effectively Neutral Mutations in Which Selective Constraint Is Incorporated". In: *Proceedings of the National Academy of Sciences of the United States of America* 76.7 (1979), pp. 3440–3444. DOI: 10.1073/pnas.76.7.3440.

[16] Guillaume Martin and Thomas Lenormand. "A General Multivariate Extension of Fisher's Geometrical Model and the Distribution of Mutation Fitness Effects Across Species". In: *Evolution* 60.5 (2006), pp. 893–907. DOI: 10.1111/j.0014-3820.2006.tb01169.x.

[17] Olin K. Silander, Olivier Tenaillon, and Lin Chao. "Understanding the Evolutionary Fate of Finite Populations: The Dynamics of Mutational Effects". In: *PLOS Biology* 5.4 (2007), e94. DOI: 10.1371/journal.pbio.0050094.

[18] Peter D Keightley and Adam Eyre-Walker. "Joint Inference of the Distribution of Fitness Effects of Deleterious Mutations and Population Demography Based on Nucleotide Polymorphism Frequencies". In: *Genetics* 177.4 (2007), pp. 2251–2261. DOI: 10.1534/genetics.107.080663.

[19] Adam R. Boyko et al. "Assessing the Evolutionary Impact of Amino Acid Mutations in the Human Genome". In: *PLOS Genetics* 4.5 (2008), e1000083. DOI: 10.1371/journal.pgen.1000083.

[20] Bernard Y. Kim, Christian D. Huber, and Kirk E. Lohmueller. "Inference of the Distribution of Selection Coefficients for New Nonsynonymous Mutations Using Large Samples". In: *Genetics* 206.1 (2017), pp. 345–361. DOI: 10.1534/genetics.116.197145.

[21] Adam Eyre-Walker and Peter D. Keightley. "Estimating the Rate of Adaptive Molecular Evolution in the Presence of Slightly Deleterious Mutations and Population Size Change". In: *Molecular Biology and Evolution* 26.9 (2009), pp. 2097–2108. DOI: 10.1093/molbev/msp119.

[22] Ying Zhen et al. "Greater Strength of Selection and Higher Proportion of Beneficial Amino Acid Changing Mutations in Humans Compared with Mice and Drosophila Melanogaster". In: *Genome Research* 31.1 (2021), pp. 110–120. DOI: 10.1101/gr.256636.119.

[23] Kimberly J Gilbert et al. "The Distribution of Mutational Effects on Fitness in Caenorhabditis Elegans Inferred from Standing Genetic Variation". In: *Genetics* 220.1 (2022), iyab166. DOI: 10.1093/genetics/iyab166.

[24] Paulina Bolívar et al. "Biased Inference of Selection Due to GC-Biased Gene Conversion and the Rate of Protein Evolution in Flycatchers When Accounting for It". In: *Molecular Biology and Evolution* 35.10 (2018), pp. 2475–2486. DOI: 10.1093/molbev/msy149.

[25] David Castellano et al. "Comparison of the Full Distribution of Fitness Effects of New Amino Acid Mutations Across Great Apes". In: *Genetics* 213.3 (2019), pp. 953–966. DOI: 10.1534/genetics.119.302494.

[26] Stephen John Gaughran. "Patterns of Adaptive and Purifying Selection in the Genomes of Phocid Seals". PhD thesis. United States – Connecticut: Yale University, 2021. 197 pp. ISBN: 9798534651188.

[27] Jacqueline A. Robinson et al. "The Critically Endangered Vaquita Is Not Doomed to Extinction by Inbreeding Depression". In: *Science* 376.6593 (2022), pp. 635–639. DOI: 10.1126/science.abm1742.

[28] Jun Chen, Sylvain Glémin, and Martin Lascoux. "Genetic Diversity and the Efficacy of Purifying Selection across Plant and Animal Species". In: *Molecular Biology and Evolution* 34.6 (2017), pp. 1417–1428. DOI: 10.1093/molbev/msx088.

26

[29] Deborah M. Leigh et al. "Opportunities and Challenges of Macrogenetic Studies". In: *Nature Reviews Genetics* 22.12 (12 2021), pp. 791–807. DOI: 10.1038/s41576-021-00394-0.

[30] Paula Tataru et al. "Inference of Distribution of Fitness Effects and Proportion of Adaptive Substitutions from Polymorphism Data". In: *Genetics* 207.3 (2017), pp. 1103–1119. DOI: 10.1534/genetics.117.300323.

[31] Paula Tataru and Thomas Bataillon. "polyDFEv2.0: Testing for Invariance of the Distribution of Fitness Effects within and across Species". In: *Bioinformatics* 35.16 (2019), pp. 2868–2869. DOI: 10.1093/bioinformatics/bty1060.

[32] Ryan N. Gutenkunst et al. "Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data". In: *PLOS Genetics* 5.10 (2009), e1000695. DOI: 10.1371/journal.pgen.1000695.

[33] Linnéa Smeds, Ilpo Kojola, and Hans Ellegren. "The Evolutionary History of Grey Wolf Y Chromosomes". In: *Molecular Ecology* 28.9 (2019), pp. 2173–2191. DOI: 10.1111/mec.15054.

[34] Emma E Wade et al. "Quantifying the Fraction of New Mutations That Are Recessive Lethal". In: *Evolution* 77.7 (2023), pp. 1539–1549. DOI: 10.1093/evolut/qpad061.

[35] Mark Pagel. "Inferring the Historical Patterns of Biological Evolution". In: *Nature* 401.6756 (1999), pp. 877–884. DOI: 10.1038/44766.

[36] Liam J. Revell. "Phylogenetic Signal and Linear Regression on Species Data". In: *Methods in Ecology and Evolution* 1.4 (2010), pp. 319–329. DOI: 10.1111/j.2041-210X.2010.00044.x.

[37] Carlos Eduardo G. Amorim et al. "Evolutionary consequences of domestication on the selective effects of new amino acid changing mutations in canids". In: *bioRxiv* (2024), pp. 2024–11. DOI: 10.1101/2024.11.13.623529.

[38] Aneil F Agrawal and Michael C Whitlock. "Inferences About the Distribution of Dominance Drawn From Yeast Gene Knockout Data". In: *Genetics* 187.2 (2011), pp. 553–566. DOI: 10.1534/genetics.110.124560.

[39] Zachary L. Fuller et al. "Measuring Intolerance to Mutation in Human Genetics". In: *Nature Genetics* 51.5 (5 2019), pp. 772–776. DOI: 10.1038/s41588-019-0383-1.

[40] Aaron P Ragsdale. "Local Fitness and Epistatic Effects Lead to Distinct Patterns of Linkage Disequilibrium in Protein-Coding Genes". In: *Genetics* 221.4 (2022), iyac097. DOI: 10.1093/genetics/iyac097.

[41] Gustavo V. Barroso and Kirk E. Lohmueller. "Inferring the Mode and Strength of Ongoing Selection". In: *Genome Research* (2023), gr.276386.121. DOI: 10.1101/gr.276386.121.

[42] Christopher C. Kyriazis and Kirk E. Lohmueller. *Constraining Models of Dominance for Nonsynonymous Mutations in the Human Genome*. 2024. DOI: 10.1101/2024.02.25.582010. Pre-published.

[43] Henry J Barton and Kai Zeng. "New Methods for Inferring the Distribution of Fitness Effects for INDELs and SNPs". In: *Molecular Biology and Evolution* 35.6 (2018), pp. 1536–1546. DOI: 10.1093/molbev/msy054.

27

[44]   Casey M. Bergman and Martin Kreitman. "Analysis of Conserved Noncoding DNA in Drosophila Reveals Similar Constraints in Intergenic and Intronic Sequences". In: *Genome Research* 11.8 (2001), pp. 1335–1345. DOI: 10.1101/gr.178701.

[45]   Athanasios Kousathanas et al. "Positive and Negative Selection on Noncoding DNA Close to Protein-Coding Genes in Wild House Mice". In: *Molecular Biology and Evolution* 28.3 (2011), pp. 1183–1191. DOI: 10.1093/molbev/msq299.

[46]   Xin Huang et al. "Inferring Genome-Wide Correlations of Mutation Fitness Effects between Populations". In: *Molecular Biology and Evolution* 38.10 (2021), pp. 4588–4602. DOI: 10.1093/molbev/msab162.

[47]   Shuyu Liu et al. "Demographic History and Natural Selection Shape Patterns of Deleterious Mutation Load and Barriers to Introgression across Populus Genome". In: *Molecular Biology and Evolution* 39.2 (2022), msac008. DOI: 10.1093/molbev/msac008.

[48]   Nicolas Galtier and Marjolaine Rousselle. "How Much Does Ne Vary Among Species?" In: *Genetics* 216.2 (2020), pp. 559–572. DOI: 10.1534/genetics.120.303622.

[49]   Anders P Charmouh, Greta Bocedi, and Matthew Hartfield. "Inferring the Distributions of Fitness Effects and Proportions of Strongly Deleterious Mutations". In: *G3 Genes—Genomes—Genetics* (2023), jkad140. DOI: 10.1093/g3journal/jkad140.

[50]   Jennifer James et al. "Between but Not Within-Species Variation in the Distribution of Fitness Effects". In: *Molecular Biology and Evolution* 40.11 (2023), msad228. DOI: 10.1093/molbev/msad228.

[51]   Lucie A. Bergeron et al. "Evolution of the Germline Mutation Rate across Vertebrates". In: *Nature* (2023), pp. 1–7. DOI: 10.1038/s41586-023-05752-y.

[52]   Annabel C Beichman et al. "Evolution of the Mutation Spectrum Across a Mammalian Phylogeny". In: *Molecular Biology and Evolution* 40.10 (2023). Ed. by Weiwei Zhai, msad213. DOI: 10.1093/molbev/msad213.

[53]   Bryan L. Gitschlag et al. "Mutation and Selection Induce Correlations between Selection Coefficients and Mutation Rates". In: *The American Naturalist* 202.4 (2023), pp. 534–557. DOI: 10.1086/726014.

[54]   Miles Roberts and Emily B. Josephs. *Previously Unmeasured Genetic Diversity Explains Part of Lewontin's Paradox in a k-Mer-Based Meta-Analysis of 112 Plant Species*. 2024. DOI: 10.1101/2024.05.17.594778. Pre-published.

[55]   Eva Schad, Peter Tompa, and Hedi Hegyi. "The Relationship between Proteome Size, Structural Disorder and Organism Complexity". In: *Genome Biology* 12.12 (2011), R120. DOI: 10.1186/gb-2011-12-12-r120.

[56]   Michael Lynch and John S. Conery. "The Origins of Genome Complexity". In: *Science* 302.5649 (2003), pp. 1401–1404. DOI: 10.1126/science.1089370.

[57]   Alan P. Boyle et al. "Comparative Analysis of Regulatory Information and Circuits across Distant Species". In: *Nature* 512.7515 (2014), pp. 453–456. DOI: 10.1038/nature13668.

[58]   Milo S. Johnson, Gautam Reddy, and Michael M. Desai. "Epistasis and Evolution: Recent Advances and an Outlook for Prediction". In: *BMC Biology* 21.1 (2023), p. 120. DOI: 10.1186/s12915-023-01585-3.

[59] Juan Diaz-Colunga et al. "Global Epistasis on Fitness Landscapes". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 378.1877 (2023), p. 20220053. DOI: `10.1098/rstb.2022.0053`.

[60] Joshua L Cherry. "Should We Expect Substitution Rate to Depend on Population Size?" In: *Genetics* 150.2 (1998), pp. 911–919. DOI: `10.1093/genetics/150.2.911`.

[61] Andreas Wagner. "Neutralism and Selectionism: A Network-Based Reconciliation". In: *Nature Reviews Genetics* 9.12 (2008), pp. 965–974. DOI: `10.1038/nrg2473`.

[62] Aisha I. Khan et al. "Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population". In: *Science* 332.6034 (2011), pp. 1193–1196. DOI: `10.1126/science.1203801`.

[63] Alejandro Couce et al. "Changing Fitness Effects of Mutations through Long-Term Bacterial Evolution". In: *Science* 383.6681 (2024), eadd1417. DOI: `10.1126/science.add1417`.

[64] Michael Lynch. "Evolution of the Mutation Rate". In: *Trends in Genetics* 26.8 (2010), pp. 345–352. DOI: `10.1016/j.tig.2010.05.003`.

[65] Christopher C. Kyriazis, Robert K. Wayne, and Kirk E. Lohmueller. "Strongly Deleterious Mutations Are a Primary Determinant of Extinction Risk Due to Inbreeding Depression". In: *Evolution Letters* 5.1 (2021), pp. 33–47. DOI: `10.1002/evl3.209`.

[66] Annabel C. Beichman et al. "Genomic Analyses Reveal Range-wide Devastation of Sea Otter Populations". In: *Molecular Ecology* (2022), mec.16334. DOI: `10.1111/mec.16334`.

[67] Sergio F. Nigenda-Morales et al. "The Genomic Footprint of Whaling and Isolation in Fin Whale Populations". In: *Nature Communications* 14.1 (2023), p. 5465. DOI: `10.1038/s41467-023-40052-z`.

[68] Pablo Cingolani et al. "A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff". In: *Fly* 6.2 (2012), pp. 80–92. DOI: `10.4161/fly.19695`.

[69] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. Vienna, Austria, 2019.

[70] Liam J. Revell. "Phytools: An R Package for Phylogenetic Comparative Biology (and Other Things)". In: *Methods in Ecology and Evolution* 3.2 (2012), pp. 217–223. DOI: `10.1111/j.2041-210X.2011.00169.x`.

[71] Jeffrey R Adrion et al. "A community-maintained standard library of population genetic models". In: *elife* 9 (2020), e54967.