

# Genetic diversity loss in the Anthropocene

Moises Exposito-Alonso<sup>1,2,3\*</sup>, Tom R. Booker<sup>4,5,&</sup>, Lucas Czech<sup>1,&</sup>, Tadashi Fukami<sup>2,&</sup>, Lauren Gillespie<sup>1,6,&</sup>, Shannon Hateley<sup>1,&</sup>, Christopher C. Kyriazis<sup>7,&</sup>, Patricia L. M. Lang<sup>2,&</sup>, Laura Leventhal<sup>1,2,&</sup>, David Nogues-Bravo<sup>8,&</sup>, Veronica Pagowski<sup>2,&</sup>, Megan Ruffley<sup>1,&</sup>, Jeffrey P. Spence<sup>9,&</sup>, Sebastian E. Toro Arana<sup>1,2,&</sup>, Clemens L. Weiß<sup>9,&</sup>, Erin Zess<sup>1,&</sup>.

<sup>1</sup>Department of Plant Biology, Carnegie Institution for Science, Stanford, CA 94305, USA.

<sup>2</sup>Department of Biology, Stanford University, Stanford, CA 94305, USA.

<sup>3</sup>Department of Global Ecology, Carnegie Institution for Science, Stanford, CA 94305, USA.

<sup>4</sup>Department of Zoology, University of British Columbia, Vancouver, Canada.

<sup>5</sup>Biodiversity Research Centre, University of British Columbia, Vancouver, Canada.

<sup>6</sup>Department of Computer Science, Stanford University, Stanford, CA 94305, USA.

<sup>7</sup>Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA.

<sup>8</sup>Center for Macroecology, Evolution and Climate, GLOBE Inst., Univ. of Copenhagen, Copenhagen, Denmark.

<sup>9</sup>Department of Genetics, Stanford University, Stanford, CA 94305, USA.

&Authors are listed alphabetically

\*To whom correspondence should be addressed: [moisesepositoalonso@gmail.com](mailto:moisesepositoalonso@gmail.com)

**Keywords:** extinction, genetic diversity, climate change, habitat loss, Anthropocene

**More species than ever before are at risk of extinction due to anthropogenic habitat loss and climate change. But even species that are not threatened have seen reductions in their populations and geographic ranges, likely impacting their genetic diversity. Although preserving genetic diversity is key to maintaining adaptability of species, we lack predictive tools and global estimates of genetic diversity loss across ecosystems. By bridging theories of biodiversity and population genetics, we introduce a mathematical framework to understand the loss of naturally occurring DNA mutations within decreasing habitat within a species. Analysing genome-wide variation data of 10,095 geo-referenced individuals from 20 plant and animal species, we show that genome-wide diversity follows a power law with geographic area (the mutations-area relationship), which can predict genetic diversity loss in spatial computer simulations of local population extinctions. Given pre-21<sup>st</sup> century values of ecosystem transformations, we estimate that over 10% of genetic diversity may already be lost, surpassing the United Nations targets for genetic preservation. These estimated losses could rapidly accelerate with advancing climate change and habitat destruction, highlighting the need for forecasting tools that facilitate implementation of policies to protect genetic resources globally.**

45 Anthropogenic habitat loss and climate change (1, 2) have led to the extinction of hundreds of species  
46 over the last centuries (1, 2) and approximately one million more species (25% of all known species)  
47 are at risk of extinction (3). It has been estimated that an even larger fraction—at least 47%—of plant  
48 and animal species have lost part of their geographic range in response to the last centuries of  
49 anthropogenic activities (4, 5). Though this loss might seem inconsequential compared to losing an  
50 entire species, this range contraction reduces genetic diversity, which dictates species' ability to adapt  
51 to new environmental conditions (6–8). The loss of geographic range can spiral into a feedback loop  
52 where diversity loss further increases the risk of species extinction (9, 10).

53  
54 Although genetic diversity is a key dimension of biodiversity (11), it has been overlooked in  
55 international conservation initiatives. Only in 2021 did the United Nations' Convention of Biological  
56 Diversity propose to preserve at least 90% of all species' genetic diversity (12, 13). Although analyses  
57 of genetic markers in animal populations sampled over time with the aim of quantifying recent genetic  
58 change are emerging (14, 15) and simulation studies with species distribution models or sensitivity  
59 analyses suggest within-species range variation may be strongly impacted (5, 16, 17), theory and  
60 scalable approaches to estimate genome-wide diversity loss across species do not yet exist, impairing  
61 prioritization and evaluation of conservation targets. Here, we introduce a framework to estimate global  
62 genetic diversity loss by bridging biodiversity theory with population genetics, and by combining data  
63 on global ecosystem transformations with newly available genomic datasets.

64  
65 The first studies that predicted biodiversity reductions in response to habitat loss and climate  
66 change in the 1990s and the 2000s projected species extinctions using the relationship of biodiversity  
67 with geographic area—termed the species-area relationship (SAR) (18) (see **Supplementary Materials**  
68 **[SM] I** for a comparison of mathematical models for predicting biodiversity). In this framework,  
69 ecosystems with a larger area ( $A$ ) harbour a larger number of species ( $S$ ) resulting from a balance of  
70 limited dispersal, habitat heterogeneity, and colonisation-extinction-speciation dynamics. The more a  
71 study area is extended, the more species are found. The SAR has been empirically shown to follow a  
72 power law,  $S = A^z$ . It scales consistently across continents and ecosystems (19), with a higher  $z$   
73 characterising more speciose and spatially structured ecosystems. Given estimates of decreasing  
74 ecosystem areas over time ( $A_{t-1} > A_t$ ), Thomas et al. (20) proposed rough estimates of the percentage of  
75 species extinctions in the 21<sup>st</sup> century ranging from 15 to 37% (**SM I.3**). Though this may be an  
76 oversimplification, SAR has become a common tool for policy groups including the Intergovernmental  
77 Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) (3).

78  
79 As species richness is for to ecosystems' biodiversity, within-species variation can be  
80 quantitatively described by the richness of genetic mutations within a species, defined here as DNA  
81 nucleotide variants appearing in individuals of a species. Although population genetics theory has long  
82 established that larger populations have higher genetic diversity (21), and it is known that geographic  
83 isolation between populations within the same species results in geographically separated accumulation  
84 of different mutations, there have been no attempts to describe the extent of genetic diversity loss driven  
85 by species' geographic range reduction using an analogous “mutations-area relationship” (MAR).

86  
87 We suspected that such a mutations-area relationship must exist given that another general  
88 assumption is shared with species studies, namely that when mutations appear they are first in only one  
89 individual, and they typically remain at low frequency in a population, though a few prevail to high  
90 frequency through stochastic genetic drift and natural selection (22). This principle of “commonness of  
91 rarity” is well-known for species (i.e. most species in an ecosystem are rare while only a few are

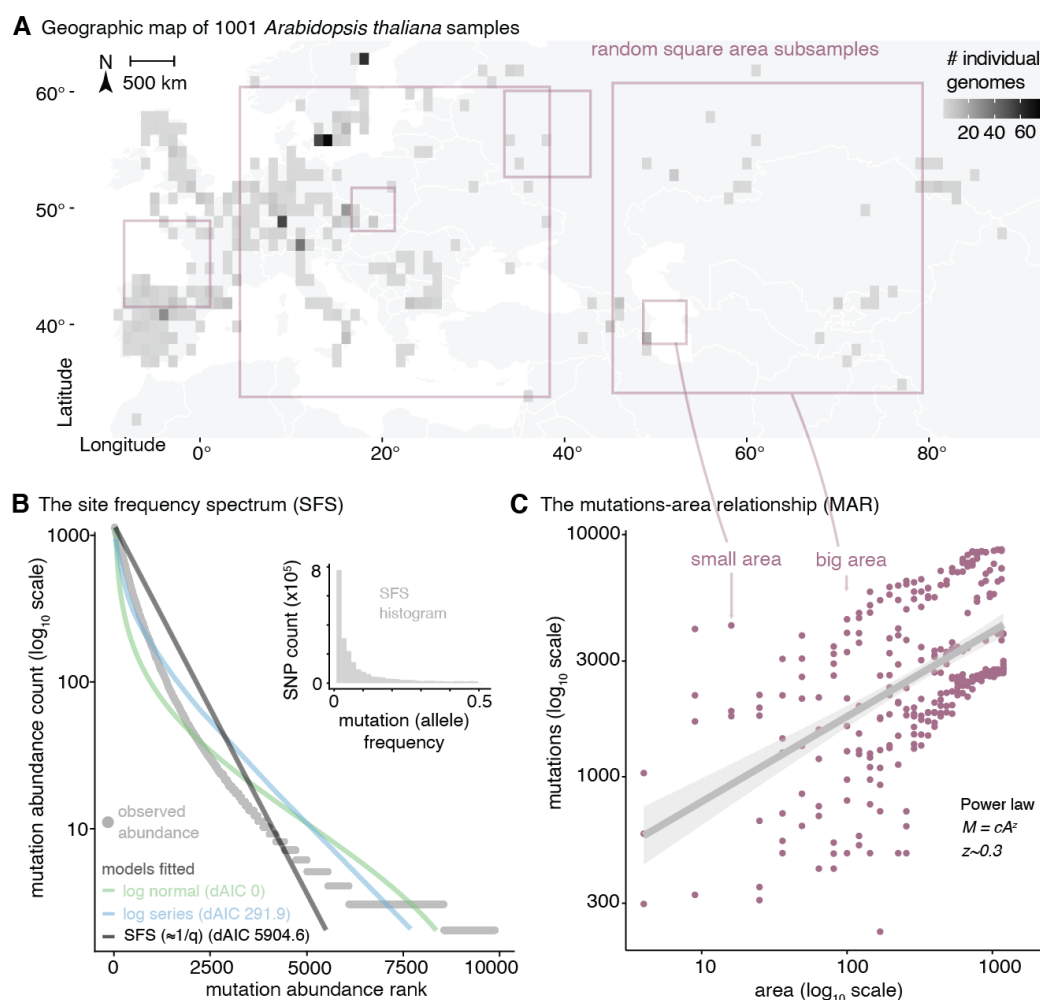
92 common) and, together with limited spatial dispersal of species and communities, is a key statistical  
 93 condition that led to the power-law SAR.

94

95 To examine the expectation of a power-law MAR, we begin quantifying the rarity of mutations  
 96 using millions of biallelic genetic variants of the *Arabidopsis thaliana* 1001 genomes dataset (Fig. 1A)  
 97 (23) by fitting several common models of species abundances (24) to the distribution of mutation  
 98 frequencies ( $q$ ), termed the Site Frequency Spectrum in population genetics (Fig. 1B, SM II.1). The  
 99 canonical L-shaped probability distribution ( $1/q$ ) of this spectrum—which is expected under  
 100 population-equilibrium and the absence of natural selection processes—fit this data well (Fig. 1B),  
 101 although the more parameter rich Preston’s species abundance log-normal model achieved the best AIC  
 102 value (Fig. 1B, SM III.1, Table S3, Table S10). Despite the small differences in fit, these models all  
 103 showcase the similarities of abundance distributions of mutations within species and species within  
 104 ecosystems, suggesting that they may behave similarly in their relationship to geographic area (22, 24).

105

106



107

108 **Fig. 1 | Mutations across populations follow a log-normal abundance distribution and a power**  
 109 **law with species range area.** (A) Density of individuals projected in a 1 x 1 degree latitude/longitude  
 110 map of Europe and exemplary subsample areas of different sizes. (B) Distribution of mutation (SNPs)  
 111 frequencies in 1,001 *Arabidopsis thaliana* plants using a site frequency spectrum histogram (grey inset)  
 112 and a Whittaker’s rank abundance curve plot, and the fitted models of common species abundance  
 113 functions in *A. thaliana* using a dataset random sample of 10,000 mutations also used in (C). The AIC  
 114 fit of the three models is indicated with respect to the top model, log-normal. (C) The mutations-area

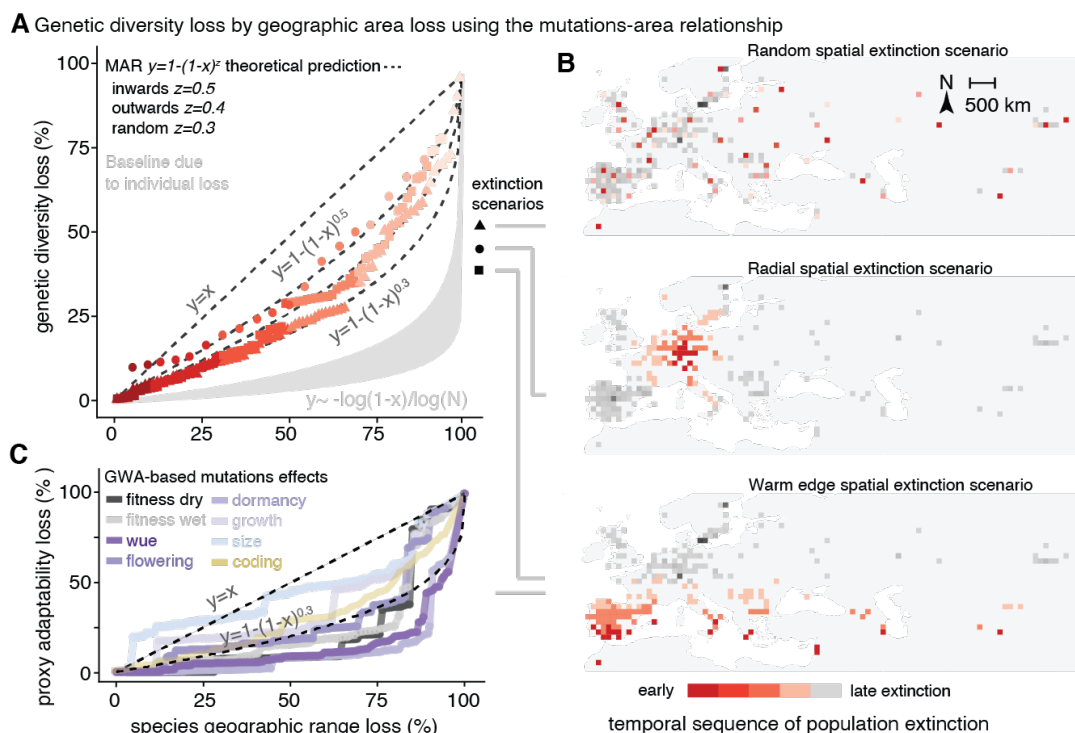
115 relationship (MAR) in log-log space built from 10 random subsamples of different areas of increasing  
116 size within *A. thaliana*'s geographic range along with the number of mutations discovered for each area  
117 subset.

118  
119 To quantify how genetic diversity within a species increases with geographic area, we  
120 constructed the MAR by subsampling different regions of different sizes of *Arabidopsis thaliana*'s  
121 native range using over one thousand geo-referenced genomes (**Fig. 1A, C**). As a metric of genetic  
122 diversity, we modelled the number of mutations ( $M$ ) in space (number of segregating sites) consistent  
123 with the species-centric approach of SAR, which uses species richness as the metric of biodiversity  
124 (**SM II.2**). The MAR also followed the power law relationship  $M = cA^z$  with a scaling value  $z_{MAR} =$   
125  $0.324$  (CI95% =  $0.238$ – $0.41$ ) (**Fig. 1C**). Naturally, subsamples of larger areas may also contain more  
126 individuals, and therefore should also have more mutations. But the observed power law relationship  
127 goes beyond what is expected from the increase of number of samples in an area (which only accounts  
128 for increases of  $M \approx \log(A)$ , see theoretical derivation **SM II.3**). The remainder may be attributed to  
129 population genetic drift and spatial natural selection causing structuring of genetic diversity across  
130 populations. The discovered power law scaling appears robust to different methods of area  
131 quantification, the effects of non-random spatial patterns, random area sampling, fully nested outward  
132 or inward sampling (*I9*), raster area calculations, raster grid resolution ( $\sim 10$ – $1,000$  km side cell size),  
133 and is adjusted for limited sample sizes (**SM II.3.2, III.3, Fig. S14-18, Tables S7-9**).

134  
135 We then wondered whether MAR can predict the loss of genetic diversity due to species' range  
136 contractions. We explored several scenarios of range contraction in *A. thaliana* by removing *in silico*  
137 grid cells in a map representing populations that are lost (**Fig. 2B**). Our simulations included random  
138 local population extinction as if deforestation was scattered across large continents, radial expansion of  
139 an extinction front due to intense localised mortality, or local extinction in the warmest regions within  
140 a species range (*4, 25*), among others (**SM III.4**). The MAR-based predictions of genetic loss, using  $1 -$   
141  $(I - A_t / A_{t-1})^z$  and assuming  $z = 0.3$ , conservatively followed the simulated local loss in *A. thaliana*  
142 (pseudo- $R^2 = 0.87$ , taking all simulations together) (**SM II.4, III.4**).

143  
144 Since genetic diversity is ultimately created by spontaneous DNA errors passed onto offspring  
145 every generation, the loss of genetic diversity seems reversible, as these mutations could happen again.  
146 However, the recovery of genetic diversity through natural mutagenesis is extremely slow (*57*),  
147 especially for mutations affecting adaptation. Simulating a species undergoing only a 5–10% in area  
148 reduction, it would take at least  $\approx 140$ – $520$  generations to recover its original genetic diversity ( $2,100$ –  
149  $7,800$  years for a fast-growing tree or medium-lifespan mammal of 15 year generation length), although  
150 for most simulations, recovery virtually never happened over millennia (see **SM II.4-5, Fig. S11, SM**  
151 **III.6**).

152  
153  
154























155  
 156 **Fig. 2 | The power law of genetic diversity loss with range area loss.** (A) Percentage of loss of total  
 157 genetic diversity in *Arabidopsis thaliana* from several stochastic simulations (red) of local extinction  
 158 in (B), and theoretical model projections of genetic diversity loss using the MAR (dotted lines). The  
 159 expectation for genetic diversity loss based only on individuals is in grey (using starting populations of  
 160  $N=10^4-10^9$ ) (SM II.4). (B) Cartoon of several possible range contractions simulated by progressively  
 161 removing grid cells across the map of Eurasia (red/grey boxes) following different hypothesised spatial  
 162 extinction patterns. (C) A metric of adaptive capacity loss during warm edge extinction in (B). Using  
 163 Genome Wide Associations (GWA) to estimate effects of mutation on fitness in different rainfall  
 164 conditions, water use efficiency [wue], flowering time, seed dormancy, plant growth rate, and plant  
 165 size. Plotted are the fraction loss of the summed squared effects ( $\sum a^2$ ) of 10,000 mutations from the top  
 166 1% tails of effects. We also plot (yellow) the fraction of protein-coding alleles lost (nonsynonymous,  
 167 stop codon loss/gain, and frameshift mutations).

168  
 169  
 170 To test the generality of the MAR, we searched in public nucleotide repositories for datasets of  
 171 hundreds to thousands of whole-genome sequenced individuals for the same species sampled across  
 172 geographic areas within their native ranges (Table 1, SM IV). In total, we identified 20 wild plant and  
 173 animal species with such published resources and assembled a dataset amassing a total of 10,095  
 174 individuals of these species, with 1,522 to 88,332,015 naturally occurring mutations per species,  
 175 covering a geographic area ranging from 0.03 to 115 million km<sup>2</sup>. Fitting MAR for these diverse species,  
 176 we recovered  $z_{MAR}$  values similar to *A. thaliana*, with many species overlapping in confidence intervals,  
 177 with the exception of some outliers (mean (SE)  $z_{MAR} = 0.31 (\pm 0.038)$ , median = 0.26, IQR =  $\pm 0.15$ ,  
 178 range=0.10–0.82, mean (SE)  $z^*_{MAR}$  scaled = 0.26 ( $\pm 0.048$ ). See Table 1, SM IV, Fig. S22, Table S10).  
 179 Theoretical derivations show that  $z_{MAR}$  is a consequence of fundamental evolutionary and ecological  
 180 forces (mutation, dispersal, selection) and should range from 0 to 1, depending on the strength of  
 181 population structure (SM II.3, see Fig. S10 for its relationship with isolation-by-distance). These  
 182 predictions were further confirmed by spatial population genetics coalescent and individual-based

183 simulations in 2D and continuous space (**SM II.3**), as well as with mainland-island community  
 184 assembly simulations according to the Unified Neutral Theory of Biodiversity (UNTB) (**SM V.3**).

185  
 186

187 **Table 1 |The mutations-area relationship across diverse species.** Summary statistics of individuals  
 188 sampled broadly across species distributions, sequencing method and mutations studied, and convex  
 189 hull area extent of all samples within a species. The mutations-area relationship (MAR) parameter  $z$ ,  
 190 which captures how spatially restricted mutations are, including a scaled correction  $z^*$  for low sampling  
 191 genomic effort. Percent area that needs to be kept for a species to maintain 90% of its genetic diversity,  
 192 using the per-species MAR value estimates. Area predictions are not provided for threatened species,  
 193 as these have likely already lost substantial genetic diversity and require protection of their full  
 194 geographic range (Fig. 3).

Species	N	M <sub>tot</sub> Method	A <sub>tot</sub> Km <sup>2</sup> x10 <sup>6</sup>	MAR z [CI95%]	MAR z* [CI95%]	scaledMin area 90% %
 <i>Arabidopsis thaliana</i>	1,135 (1,001) <sup>#</sup>	11,769,920 W	27.34	0.324 (0.238–0.41)	0.312 (0.305 - 0.32)	71–78
 <i>Arabidopsis lyrata</i>	108	17,813,817 W	2.79	0.236 (0.218–0.254)	0.151 (0.137–0.165)	50–66
 <i>Amaranthus tuberculatus</i>	162 (155)	1,033,443 W	0.80	0.109 (0.081–0.136)	0.142 (0.136–0.149)	48–65
 <i>Eucalyptus melliodora</i> <sup>VU</sup>	275 (36) <sup>*</sup>	9,378 GBS	0.95	0.466 (0.394–0.538)	0.403 (0.398–0.407)	77–82
 <i>Yucca brevifolia</i> <sup>CA</sup>	290	10,695 GBS	NA	<sup>?</sup> 0.128 (0.109–0.147)	0.049 (0.037–0.062)	-
 <i>Mimulus guttatus</i>	521 (286) <sup>#*</sup>	1,522 GBS	25.14	0.274 (0.259–0.29)	0.231 (0.221–0.241)	63–73
 <i>Panicum virgatum</i>	732 (576) <sup>†</sup>	33,905,044 W	6.29	0.232 (0.211–0.252)	0.126 (0.116–0.136)	43–63
 <i>Panicum hallii</i>	591	45,589 W	2.19	0.824 (0.719 - 0.928)	0.814 (0.745 - 0.883)	88–90
 <i>Pinus contorta</i>	929	32,449 GC	0.89	<sup>?</sup> 0.015 (0.014–0.016)	-0.061(-0.062-0.060)	-
 <i>Pinus torreyana</i> <sup>CR</sup>	242	478,238 GBS	NA	<sup>?</sup> 0.236 (0.19–0.282)	0.105 (0.099–0.11)	-
 <i>Populus trichocarpa</i>	882	28,342,826 W	1.12	0.275 (0.218–0.332)	0.165 (0.155–0.176)	53–67
 <i>Anopheles gambiae</i>	1142 (29) <sup>*</sup>	52,525,957 W	19.96	0.214 (0.164–0.264)	0.122 (0.111–0.132)	42–62
 <i>Acropora millepora</i> <sup>NT</sup>	253 (12) <sup>*</sup>	17,931,448 W	0.03	0.246 (0.209–0.283)	0.287 (0.28–0.294)	69–77
 <i>Drosophila melanogaster</i>	271 <sup>%</sup>	5,019 W	115.21	0.437 (0.397–0.477)	0.325 (0.314–0.336)	72–79
 <i>Empidonax traillii</i> <sup>Decline</sup>	219 (199) <sup>&amp;</sup>	349,014 GBS/GC	7.03	0.214 (0.174–0.254)	0.074 (0.047–0.102)	24–54
 <i>Setophaga petechia</i> <sup>Decline</sup>	199	104,711 GBS	15.17	0.251 (0.236 - 0.267)	0.149 (0.135 - 0.163)	49--66
 <i>Peromyscus maniculatus</i>	80 (78) <sup>&amp;</sup>	14,076 GBS	22.61	0.488 (0.264–0.713)	0.683 (0.615–0.751)	86–88
 <i>Dicerorhinus sumatrensis</i> <sup>CR</sup>	16	8,870,513 W	NA	<sup>?</sup> 0.412 (0.369–0.456)	0.127 (0.11–0.144)	-
 <i>Canis lupus</i>	349 (230) <sup>†</sup>	1,517,226 W	19.10	0.256 (0.232–0.28)	0.184 (0.175–0.193)	56–70
 <i>Homo sapiens</i>	2504 (24) <sup>*</sup>	88,332,015 W	80.76	<sup>?</sup> 0.431 (0.347–0.514)	0.281 (0.23–0.332)	NA

195 <sup>#</sup>Only individuals in the native range were used for the analyses.

196 <sup>&</sup>Only individuals with available coordinates or matching IDs were used for analyses.

197 <sup>%</sup>Numbers indicate pools of flies used for Pool-Sequencing.

198 <sup>\*</sup>Number of geographically separated populations, as multiple individuals were collected per population.

199 <sup>†</sup>Only natural populations were used, excluding breeds, landraces, and cultivars.

200 Area was not reported for species with unknown locations or where less than 2 populations were sampled.

201 <sup>?</sup>Values excluded from global averages used for conservation applications due to uncertain estimates, suboptimal genomic data type, or  
 202 because estimates should not be applied for conservation (i.e. humans or nearly extinct Sumatran rhinoceros).

203 Acronyms: W = whole-genome re-sequencing or discovery SNP calling. GBS = genotyping by sequencing of biallelic SNP markers. GC =  
 204 genotyping chip; CR = Red List Critically Endangered. VU= Red List Vulnerable. CA = included in the California Endangered Species Act.  
 205 Decline = population decline reported in the Red List.

206  
 207

208 Although we expect species-specific traits related to dispersibility or gene flow to affect  $z_{MAR}$   
 209 (e.g. migration rate and environmental selection in population genetic simulations significantly  
 210 influences  $z_{MAR}$ , **Table S2**), no significant association was found between  $z_{MAR}$  and different  
 211 ecologically-relevant traits, mating systems, home continents, etc., for the 20 species analysed. Perhaps  
 212 this is simply that there are still too few species that have large population genomic data to find such a  
 213 signal (**Table 1, Table S12-13**). Nevertheless, the relative consistency of  $z_{MAR}$  across largely different  
 214 species may be promising for conservation purposes, as an average  $z_{MAR} \sim 0.3$  (IQR  $\pm 0.15$ , **Table 1,**  
 215 **Table S11**) could be predictive of large-scale trends of genetic diversity loss in many range-reduced  
 216 species that lack genomic information. Further, although species will naturally have different starting

217 levels of total genetic diversity prior to range reductions, for instance, due to genome size, structure, or  
218 mating system differences (26), the application of  $z_{MAR}$  provides relative estimates of genetic diversity  
219 loss. For instance, assuming  $z_{MAR} \sim 0.3$ , we would predict that an area reduction of  $\sim 50\%$  creates an  
220 approximate loss of  $\sim 20\%$  of genetic diversity relative to the total genetic diversity of a given species.

221

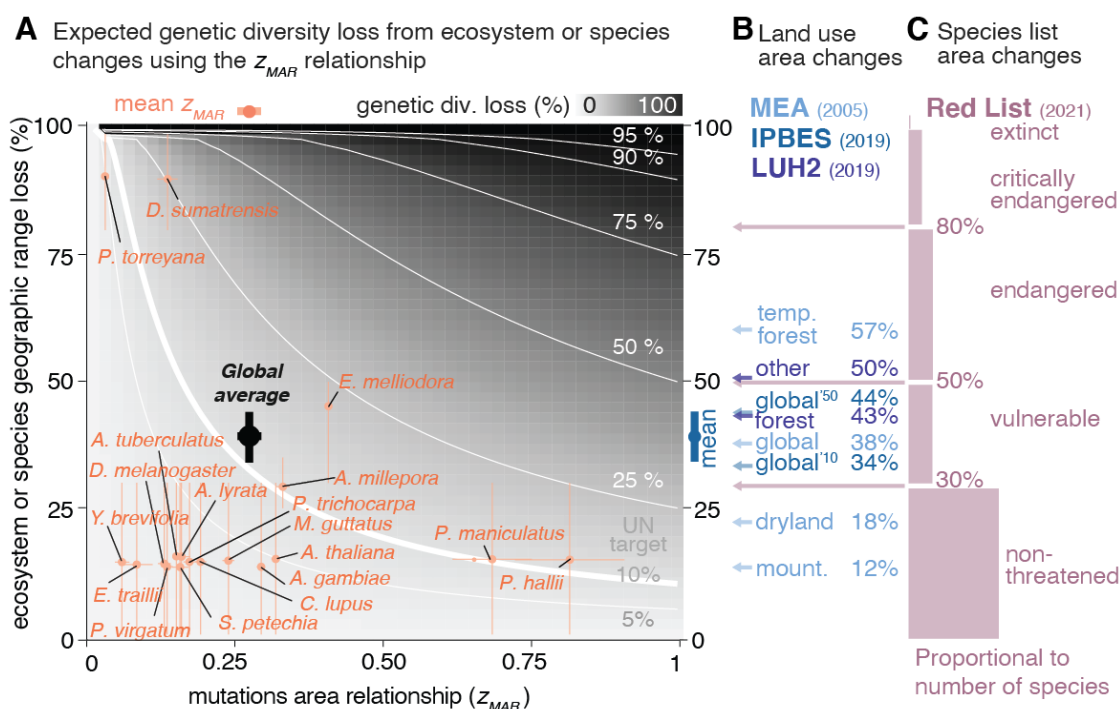
222 Finally, we used MAR to estimate the average global genetic diversity loss caused by pre-21<sup>st</sup>  
223 century land transformations. Although accurate species-specific geographic area reduction data in the  
224 last centuries are scarce, we leveraged global land cover transformations from primary ecosystems to  
225 urban or cropland systems (3, 27) (Table S14-15). Using the average scaled  $z_{MAR}^*$  (Table S18) and  
226 several global averages of Earth's land and coastal transformations for present day (38% global area  
227 transformation from (27), 34% from (28), and 43-50% from (29)), we estimate a 10-16% global genetic  
228 diversity loss on average across species (Fig. 3A). While these estimates may correctly approximate  
229 central values across species in an ecosystem, we expect a substantial variation in the extent of loss  
230 across species, ranging theoretically from 0 to 100% (Fig. 3, Fig. S26). One cause of this variation is  
231 the heterogeneity in land cover transformations across ecosystems; for example, more pristine high-  
232 altitude systems have only lost 0.3% of their area, while highly managed temperate forests and  
233 woodlands have lost 67% (Fig. 3B, Table S14-15).

234

235 Another cause for the variability in genetic loss among species (even within the same  
236 ecosystem) may be their differential geographic ranges and abundances, life histories, or species-  
237 specific threats. We gathered data from species red-listed by the International Union for Conservation  
238 of Nature (IUCN) (1), which evaluates recent population or geographic range area reduction over  $\pm 10$   
239 years /  $\pm 3$  generations to place assessed species in different threat categories using several thresholds  
240 (guidelines for assessments and thresholds available at [www.iucn.org](http://www.iucn.org)). Again, assuming that with the  
241 average  $z_{MAR} \sim 0.3$  we can capture general patterns, we translate these category thresholds into genetic  
242 diversity loss (Fig. 3C, see SM V, Table S17). *Vulnerable* species, having lost at least 30% of their  
243 geographic distribution, may have experienced  $>9\%$  of genetic diversity loss, *endangered* species,  
244 which have lost over 50% of their geographic distribution, should have incurred  $>16\%$  of genetic  
245 diversity loss, and *critically endangered* species, with over 80% area reduction, likely suffered  $>33\%$   
246 of genetic diversity loss (Fig. 3B). This clearly showcases that even species in no imminent risk of  
247 extinction (e.g. least concern, near threatened, vulnerable), such as the majority of species for which  
248 population genomic data exists, may already be losing substantial genetic diversity (Fig. 3A).

249

250



251  
 252 **Fig. 3 | The parameter space of genetic diversity loss mapping pre-21<sup>st</sup> century ecosystem**  
 253 **transformations and species threat categories against possible values of the mutations-area**  
 254 **relationship.** (A) Possible values of two key parameters, the mutations-area relationship scaling  
 255 parameter (MAR) and % of area reduction of a species geographic range (as a proxy of entire ecosystem  
 256 transformation). The theoretical % of genetic diversity loss is represented as filled grey colour, with  
 257 isolines in white. Estimates of scaled  $z_{MAR}^*$  from Table 1 per species are in orange with their 95%  
 258 confidence intervals (for unscaled  $z_{MAR}$  see Fig. S23). Although exact area losses per species are  
 259 unknown, species are plotted based on their IUCN Red list (C) status, using the broad ranges of  
 260 minimum and maximum recent population or area decline per category. The global average is calculated  
 261 with the average  $z_{MAR}$  across species and % of the Earth transformed from IPBES. (B) Percentage of  
 262 transformed ecosystem area from the Millennium Ecosystem Assessment (MEA) (27) are represented  
 263 by light blue arrows, from the Intergovernmental Science-Policy Panel for Biodiversity and Ecosystem  
 264 Services (IPBES) (28) for 2010 and 2050 are dark blue arrows, and from the Land Use Harmonization  
 265 2 (LUH2) dataset (29) are in dark purple. (C) The minimum criterion value of population or geographic  
 266 area loss to be classified in each category of the IUCN Red List are indicated with pink arrows (the near  
 267 threatened category does not have a range of values, instead we used 30%  $\pm$  10%). The number of plant  
 268 species (for which population abundance loss approximates area loss) included in each category is  
 269 shown as box sizes (I). The IUCN ranges were used to place ranges of estimates in (A) per species.

270  
 271 The ultimate challenge is to understand how genetic diversity loss relates to loss of adaptive  
 272 capacity of a species. To this end, we leveraged the extensive knowledge of the effect of mutations in  
 273 ecologically relevant traits in *A. thaliana* from Genome-Wide Associations (GWA) (Fig. 2C, SM III).  
 274 We again conducted spatial warm edge extinction simulations, this time tracking metrics of adaptive  
 275 capacity, including the total sum of effects estimated from GWA of remaining mutations ( $\sum_i a_i$  for  
 276  $i=1 \dots 10,000$  variants of putative  $a_i$  effect), the additive genetic variance ( $Va = \sum_i p_i(1-p_i)a_i^2$ , which  
 277 accounts for each variant's population frequency  $p_i$ ), and the loss of nonsynonymous mutations (SM  
 278 III.5). Although determining the effect of mutations through GWA is technically challenging even in  
 279 model species (30, 31), and variants may even be either deleterious or advantageous depending on  
 280 genomic backgrounds (32) or environments (33), our simulations suggest putatively functional



281 mutations may be lost more slowly ( $z < 0.3$ , **Fig. 2C**) than neutral genetic diversity (**Fig. 2A**). In fact,  
282 the additive variance  $V_a$  parameter, often equated to the rate of adaptation, appears rather stable (34)  
283 until just before the extinction event when it sharply collapses (**Fig. S21**; see also **Fig. 2C**, and **SM**  
284 **II.3.4** for simulations that replicate this pattern). This is analogous to the famous “rivet popper”  
285 metaphor where ecosystem structure and function may suddenly collapse as species are inadvertently  
286 lost (35). Projections of the MAR using genome-wide variation may crucially serve as early  
287 conservation tool in non-threatened species (36, 37), before species reach accelerating collapsing  
288 extinction dynamics—an acceleration that we expect to be even more dramatic due to elevated drift and  
289 accumulation of deleterious mutations of small critically-endangered populations (38, 39).

290

291 To achieve the recently published United Nations target to protect “at least 90% of genetic  
292 diversity within all species”(13), it will be necessary to aggressively protect as many populations as  
293 possible for each species. Here, we have discovered the existence of a mutations-area relationship  
294 (MAR) and provided a mathematical framework to forecast genetic diversity loss with shrinking  
295 geographic species ranges. The MAR contrasts with existing studies on the risk of losing entire species  
296 by focusing on quantifying the magnitude and dynamics of genetic diversity loss likely ongoing in most  
297 species. This framework demonstrates that even with conservative estimates, substantial area protection  
298 will be needed to meet the UN Sustainable Development Goals. For vulnerable or endangered species,  
299 we may have likely already failed.

300

## 301 **ADDITIONAL INFORMATION**

302

303 **Author contribution** M.E.-A. conceived and led the project. M.E.-A., J.P.S., M.R., S.H., L.G., L.C.,  
304 L.L., S.T.A., V.P., E.Z., P.L.M.L., C.C.K., T.B., C.W. conducted research, all authors interpreted the  
305 results and wrote the manuscript.

306

307 **Data availability.** The analysed datasets are publicly available or were shared by authors upon request  
308 (see Supplementary Materials for details). Code is available at Github  
309 (<https://github.com/moiepositoalonso/lab/mar>) and Zenodo (<https://doi.org/10.5281/zenodo.6408624>).

310

311 **Acknowledgements.** We are grateful for many researchers who made their genomic data publicly  
312 available and thus made this research possible: the 1001 Arabidopsis Genomes Consortium, the  
313 Anopheles gambiae 1000 Genomes Consortium, the 1000 (Human) Genome Consortium, the European  
314 "Drosophila Evolution over Space and Time" (DEST) Consortium, MacLachlan et al., Fuller et al.,  
315 Ruegg et al., Kingsley et al., Schweizer et al., and Royer et al.. In addition, we are grateful to all the  
316 authors who directly shared intermediate genome variant files: Sujan Mamidi, John Lovell, Dan  
317 Jacobson, Manesh Shah, Julia Kreiner, Kay Lucek, Yvonne Willi, Juan D. Palacio Mejia, Justin  
318 Borevitz, Megan Supple, Mario Vallejo-Marin, Nicolas Dussex, Lionel Di Santo, and Jill Hamilton.  
319 We thank John Wiens, Sue Rhee, Detlef Weigel, Ellie Armstrong, Marty Kardos, Dmitri Petrov, Rob  
320 Colwell, Rasmus Nielsen, Anna Michalak, Sean Hoban, Jonathan Pritchard, and members of the Moi  
321 and Mordecai Labs for comments, discussion, or references. M.E.-A. is supported by the Office of the  
322 Director of the National Institutes of Health’s Early Investigator Award with award number:  
323 1DP5OD029506-01; by the U.S. Department of Energy, Office of Biological and Environmental  
324 Research, grant number: DE-SC0021286; and by the Carnegie Institution for Science. J.P.S. is  
325 supported by an NIH training grant (5T32HG000044-23), and S.T.A by the NIGMS Center of the NIH  
326 under award number T32GM007276. S.H. and C.L.W. are supported by Stanford’s Center for  
327 Computational, Evolutionary, and Human Genomics. P.L.M.L. is supported by a Human Frontier  
328 Science Program Long-Term Fellowship (LT000330/2019-L). M. R. is supported by the NSF’s Plant

329 Genome Postdoctoral Research Fellowship in Biology. L. L. and L.G. are supported by NSF's GRFP.  
330 Computational analyses were done on the High-Performance Computing clusters *Memex* and *Calc*  
331 supported by the Carnegie Institution for Science.

332

## 333 REFERENCES

- 334 1. IUCN, The IUCN Red List of Threatened Species. <https://www.iucnredlist.org>.
- 335 2. S. Díaz, J. Settele, E. S. Brondízio, H. T. Ngo, J. Agard, A. Arneth, P. Balvanera, K. A.  
336 Brauman, S. H. M. Butchart, K. M. A. Chan, L. A. Garibaldi, K. Ichii, J. Liu, S. M.  
337 Subramanian, G. F. Midgley, P. Miloslavich, Z. Molnár, D. Obura, A. Pfaff, S. Polasky, A.  
338 Purvis, J. Razzaque, B. Reyers, R. R. Chowdhury, Y.-J. Shin, I. Visseren-Hamakers, K. J. Willis,  
339 C. N. Zayas, Pervasive human-driven decline of life on Earth points to the need for  
340 transformative change. *Science*. **366**, eaax3100 (2019).
- 341 3. IPBES, *Global Assessment Report on Biodiversity and Ecosystem Services* (IPBES Secretariat,  
342 Bonn, Germany, 2019; [https://www.ipbes.net/global-assessment-report-biodiversity-ecosystem-](https://www.ipbes.net/global-assessment-report-biodiversity-ecosystem-services)  
343 [services](https://www.ipbes.net/global-assessment-report-biodiversity-ecosystem-services)).
- 344 4. J. J. Wiens, Climate-Related Local Extinctions Are Already Widespread among Plant and  
345 Animal Species. *PLoS Biol.* **14**, e2001104 (2016).
- 346 5. W. Thuiller, S. Lavorel, M. B. Araújo, M. T. Sykes, I. C. Prentice, Climate change threats to  
347 plant diversity in Europe. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 8245–8250 (2005).
- 348 6. M. Exposito-Alonso, F. Vasseur, W. Ding, G. Wang, H. A. Burbano, D. Weigel, Genomic basis  
349 and evolutionary potential for extreme drought adaptation in *Arabidopsis thaliana*. *Nat Ecol*  
350 *Evol.* **2**, 352–358 (2018).
- 351 7. C. Parmesan, Ecological and Evolutionary Responses to Recent Climate Change. *Annu. Rev.*  
352 *Ecol. Evol. Syst.* **37**, 637–669 (2006).
- 353 8. T. Capblancq, M. C. Fitzpatrick, R. A. Bay, M. Exposito-Alonso, S. R. Keller, Genomic  
354 Prediction of (Mal)Adaptation Across Current and Future Climatic Landscapes. *Annu. Rev. Ecol.*  
355 *Evol. Syst.* **51**, 245–269 (2020).
- 356 9. M. Lynch, J. Conery, R. Burger, Mutation Accumulation and the Extinction of Small  
357 Populations. *Am. Nat.* **146**, 489–518 (1995).
- 358 10. D. Spielman, B. W. Brook, R. Frankham, Most species are not driven to extinction before  
359 genetic factors impact them. *Proceedings of the National Academy of Sciences.* **101** (2004), pp.  
360 15261–15264.
- 361 11. W. Steffen, K. Richardson, J. Rockström, S. E. Cornell, I. Fetzer, E. M. Bennett, R. Biggs, S. R.  
362 Carpenter, W. de Vries, C. A. de Wit, C. Folke, D. Gerten, J. Heinke, G. M. Mace, L. M.  
363 Persson, V. Ramanathan, B. Reyers, S. Sörlin, Planetary boundaries: Guiding human  
364 development on a changing planet. *Science*. **347** (2015), doi:10.1126/science.1259855.
- 365 12. S. Díaz, N. Zafra-Calvo, A. Purvis, P. H. Verburg, D. Obura, P. Leadley, R. Chaplin-Kramer, L.  
366 De Meester, E. Dulloo, B. Martín-López, M. R. Shaw, P. Visconti, W. Broadgate, M. W.  
367 Bruford, N. D. Burgess, J. Cavender-Bares, F. DeClerck, J. M. Fernández-Palacios, L. A.  
368 Garibaldi, S. L. L. Hill, F. Isbell, C. K. Khoury, C. B. Krug, J. Liu, M. Maron, P. J. K.  
369 McGowan, H. M. Pereira, V. Reyes-García, J. Rocha, C. Rondinini, L. Shannon, Y.-J. Shin, P.  
370 V. R. Snelgrove, E. M. Spehn, B. Strassburg, S. M. Subramanian, J. J. Tewksbury, J. E. M.  
371 Watson, A. E. Zanne, Set ambitious goals for biodiversity and sustainability. *Science*. **370**, 411–

- 372 413 (2020).
- 373 13. CBD, 1st Draft of The Post-2020 Global Biodiversity Framework (2021), (available at  
374 <https://www.cbd.int/doc/c/abb5/591f/2e46096d3f0330b08ce87a45/wg2020-03-03-en.pdf>).
- 375 14. D. M. Leigh, A. P. Hendry, E. Vázquez-Domínguez, V. L. Friesen, Estimated six per cent loss of  
376 genetic variation in wild populations since the industrial revolution. *Evol. Appl.* **12**, 1505–1512  
377 (2019).
- 378 15. K. L. Millette, V. Fugère, C. Debyser, A. Greiner, F. J. J. Chain, A. Gonzalez, No consistent  
379 effects of humans on animal genetic diversity worldwide. *Ecol. Lett.* **23**, 55–67 (2020).
- 380 16. I. G. Alsos, D. Ehrich, W. Thuiller, P. B. Eidesen, A. Tribsch, P. Schönswetter, C. Lagaye, P.  
381 Taberlet, C. Brochmann, Genetic consequences of climate change for northern plants. *Proc. Biol.*  
382 *Sci.* **279**, 2042–2051 (2012).
- 383 17. S. Theodoridis, C. Rahbek, D. Nogues-Bravo, Exposure of mammal genetic diversity to mid-21st  
384 century global change. *Ecography (Cop.)* (2021), doi:10.1111/ecog.05588.
- 385 18. O. Arrhenius, Species and Area. *J. Ecol.* **9**, 95–99 (1921).
- 386 19. D. Storch, P. Keil, W. Jetz, Universal species-area and endemics-area relationships at continental  
387 scales. *Nature.* **488**, 78–81 (2012).
- 388 20. C. D. Thomas, A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C. Collingham, B. F.  
389 N. Erasmus, M. F. de Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A. S. van  
390 Jaarsveld, G. F. Midgley, L. Miles, M. A. Ortega-Huerta, A. Townsend Peterson, O. L. Phillips,  
391 S. E. Williams, Extinction risk from climate change. *Nature.* **427**, 145–148 (2004).
- 392 21. V. Buffalo, Quantifying the relationship between genetic diversity and population size suggests  
393 natural selection cannot explain Lewontin’s paradox. *Elife.* **10** (2021), doi:10.7554/eLife.67509.
- 394 22. R. A. Fisher, XVII.—The Distribution of Gene Ratios for Rare Mutations. *Proceedings of the*  
395 *Royal Society of Edinburgh.* **50**, 204–219 (1931).
- 396 23. 1001 Genomes Consortium, 1,135 Genomes Reveal the Global Pattern of Polymorphism in  
397 *Arabidopsis thaliana*. *Cell.* **166**, 481–491 (2016).
- 398 24. F. W. Preston, The canonical distribution of commonness and rarity: Part I. *Ecology.* **43**, 185  
399 (1962).
- 400 25. A. Hampe, R. J. Petit, Conserving biodiversity under climate change: the rear edge matters. *Ecol.*  
401 *Lett.* **8**, 461–467 (2005).
- 402 26. V. Buffalo, Why do species get a thin slice of  $\pi$ ? Revisiting Lewontin’s Paradox of Variation.  
403 *bioRxiv* (2021), p. 2021.02.03.429633.
- 404 27. Millennium Ecosystem Assessment, *Millennium ecosystem assessment* (Millennium Ecosystem  
405 Assessment, 2005; <http://chapter.ser.org/europe/files/2012/08/Harris.pdf>).
- 406 28. Ipbes, *Global assessment report of the Intergovernmental Science-Policy Platform on*  
407 *Biodiversity and Ecosystem Services* (IPBES Secretariat, Bonn, Germany, 2019;  
408 <https://www.ipbes.net/news/ipbes-global-assessment-summary-policymakers-pdf>).
- 409 29. G. C. Hurtt, L. Chini, R. Sahajpal, S. Frolking, B. L. Boudirsky, K. Calvin, J. C. Doelman, J. Fisk,  
410 S. Fujimori, K. Klein Goldewijk, T. Hasegawa, P. Havlik, A. Heinemann, F. Humpenöder, J.  
411 Jungclaus, J. O. Kaplan, J. Kennedy, T. Krisztin, D. Lawrence, P. Lawrence, L. Ma, O. Mertz, J.

- 412 Pongratz, A. Popp, B. Poulter, K. Riahi, E. Shevliakova, E. Stehfest, P. Thornton, F. N. Tubiello,  
413 D. P. van Vuuren, X. Zhang, Harmonization of global land use change and management for the  
414 period 850–2100 (LUH2) for CMIP6. *Geosci. Model Dev.* **13**, 5425–5464 (2020).
- 415 30. G. Gibson, Rare and common variants: twenty arguments. *Nat. Rev. Genet.* **13**, 135–145 (2011).
- 416 31. M. V. Rockman, The QTN program and the alleles that matter for evolution: all that’s gold does  
417 not glitter. *Evolution.* **66**, 1–17 (2012).
- 418 32. M. J. Harms, J. W. Thornton, Evolutionary biochemistry: revealing the historical and physical  
419 causes of protein properties. *Nat. Rev. Genet.* **14**, 559–571 (2013).
- 420 33. M. Exposito-Alonso, 500 Genomes Field Experiment Team, H. A. Burbano, O. Bosssdorf, R.  
421 Nielsen, D. Weigel, Natural selection in the *Arabidopsis thaliana* genome in present and future  
422 climates. *Nature.* **573**, 126–129 (2019).
- 423 34. H. R. Taft, D. A. Roff, Do bottlenecks increase additive genetic variance? *Conserv. Genet.* **13**,  
424 333–342 (2012).
- 425 35. P. Ehrlich, B. Walker, Rivets and redundancy. *Bioscience.* **48**, 387 (1998).
- 426 36. C. C. Kyriazis, R. K. Wayne, K. E. Lohmueller, Strongly deleterious mutations are a primary  
427 determinant of extinction risk due to inbreeding depression. *Evolution Letters.* **n/a** (2020),  
428 doi:10.1002/evl3.209.
- 429 37. M. Kardos, E. Armstrong, S. W. Fitzpatrick, S. Hauser, P. Hedrick, J. Miller, D. Tallmon, W.  
430 Chris Funk, The crucial role of genome-wide genetic variation in conservation. *bioRxiv* (2021),  
431 p. 2021.07.05.451163.
- 432 38. R. Lande, Risks of Population Extinction from Demographic and Environmental Stochasticity  
433 and Random Catastrophes. *Am. Nat.* **142**, 911–927 (1993).
- 434 39. M. Lynch, R. Lande, The critical effective size for a genetically secure population. *Anim.*  
435 *Conserv.* **1**, 70–72 (1998).

436

#### 437 SUPPLEMENTAL REFERENCES

- 438 1. S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography* (Monographs in  
439 Population Biology, 2001).
- 440 2. M. Tokeshi, Niche Apportionment or Random Assortment: Species Abundance Patterns  
441 Revisited. *J. Anim. Ecol.* **59**, 1129–1146 (1990).
- 442 3. M. Tokeshi, in *Advances in Ecological Research*, M. Begon, A. H. Fitter, Eds. (Academic Press,  
443 1993; <https://www.sciencedirect.com/science/article/pii/S0065250408600422>), vol. 24, pp. 111–  
444 186.
- 445 4. MOTOMURA, I, A statistical treatment of ecological communities. *Zoological Magazine.* **44**,  
446 379–383 (1932).
- 447 5. R. H. MacArthur, On the relative abundance of bird species. *Proc. Natl. Acad. Sci. U. S. A.* **43**,  
448 293–295 (1957).

- 449 6. R. A. Fisher, A. S. Corbet, C. B. Williams, The Relation Between the Number of Species and the  
450 Number of Individuals in a Random Sample of an Animal Population. *J. Anim. Ecol.* **12**, 42–58  
451 (1943).
- 452 7. F. W. Preston, The canonical distribution of commonness and rarity: Part I. *Ecology.* **43**, 185  
453 (1962).
- 454 8. F. W. Preston, The commonness, and rarity, of species. *Ecology.* **29**, 254–283 (1948).
- 455 9. D. Alonso, A. J. McKane, Sampling Hubbell’s neutral theory of biodiversity. *Ecol. Lett.* **7**, 901–  
456 910 (2004).
- 457 10. D. Storch, P. Keil, W. Jetz, Universal species-area and endemics-area relationships at continental  
458 scales. *Nature.* **488**, 78–81 (2012).
- 459 11. S. L. Pimm, G. J. Russell, J. L. Gittleman, T. M. Brooks, The future of biodiversity. *Science.*  
460 **269**, 347–350 (1995).
- 461 12. C. D. Thomas, A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C. Collingham, B. F.  
462 N. Erasmus, M. F. de Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A. S. van  
463 Jaarsveld, G. F. Midgley, L. Miles, M. A. Ortega-Huerta, A. Townsend Peterson, O. L. Phillips,  
464 S. E. Williams, Extinction risk from climate change. *Nature.* **427**, 145–148 (2004).
- 465 13. J. F. C. Kingman, The coalescent. *Stochastic Process. Appl.* **13**, 235–248 (1982).
- 466 14. R. C. Griffiths, S. Tavaré, The age of a mutation in a general coalescent tree. *Communications in*  
467 *Statistics. Stochastic Models.* **14**, 273–295 (1998).
- 468 15. R. A. Fisher, XVII.—The Distribution of Gene Ratios for Rare Mutations. *Proceedings of the*  
469 *Royal Society of Edinburgh.* **50**, 204–219 (1931).
- 470 16. M. W. Hahn, *Molecular Population Genetics* (Oxford University Press, 2018);  
471 <https://play.google.com/store/books/details?id=3BDkswEACAAJ>.
- 472 17. V. Buffalo, Quantifying the relationship between genetic diversity and population size suggests  
473 natural selection cannot explain Lewontin’s paradox. *Elife.* **10** (2021), doi:10.7554/eLife.67509.
- 474 18. D. R. Marshal, A. D. H. Brown, in *Crop genetic resources for today and tomorrow*, O. H.  
475 Frankel, J. G. Hawkes, Eds. (Cambridge University Press, 1975);  
476 <https://www.researchgate.net/publication/280057199>, vol. 2.
- 477 19. S. Wright, Isolation by Distance. *Genetics.* **28**, 114–138 (1943).
- 478 20. J. Novembre, M. Stephens, Interpreting principal component analyses of spatial population  
479 genetic variation. *Nat. Genet.* **40**, 646–649 (2008).
- 480 21. H. Fan 樊海英, Q. Zhang 张清臣, J. Rao 饶娟娟, J. Cao 曹静文, X. Lu 卢欣, Genetic Diversity-  
481 Area Relationships across Bird Species. *Am. Nat.* **194**, 736–740 (2019).
- 482 22. P. R. A. Campos, V. M. de Oliveira, A. Rosas, Epistasis and environmental heterogeneity in the  
483 speciation process. *Ecol. Modell.* **221**, 2546–2554 (2010).

- 484 23. J. Kelleher, A. M. Etheridge, G. McVean, Efficient Coalescent Simulation and Genealogical  
485 Analysis for Large Sample Sizes. *PLoS Comput. Biol.* **12**, e1004842 (2016).
- 486 24. B. C. Haller, P. W. Messer, SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher  
487 Model. *Mol. Biol. Evol.* **36**, 632–637 (2019).
- 488 25. T. R. Booker, S. Yeaman, M. C. Whitlock, The WZA: A window-based method for  
489 characterizing genotype–environment association. *bioRxiv* (2021), p. 2021.06.25.449972.
- 490 26. J. Kelleher, K. R. Thornton, J. Ashander, P. L. Ralph, Efficient pedigree recording for fast  
491 population genetics simulation. *PLoS Comput. Biol.* (2018), p. e1006581.
- 492 27. B. C. Haller, J. Galloway, J. Kelleher, P. W. Messer, P. L. Ralph, Tree–sequence recording in  
493 SLiM opens new horizons for forward–time simulation of whole genomes. *Mol. Ecol. Resour.*  
494 **19**, 552–566 (2019).
- 495 28. R. R. Hudson, M. Slatkin, W. P. Maddison, Estimation of levels of gene flow from DNA  
496 sequence data. *Genetics*. **132**, 583–589 (1992).
- 497 29. 1001 Genomes Consortium, 1,135 Genomes Reveal the Global Pattern of Polymorphism in  
498 *Arabidopsis thaliana*. *Cell*. **166**, 481–491 (2016).
- 499 30. B. J. McGill, B. J. Enquist, E. Weiher, M. Westoby, Rebuilding community ecology from  
500 functional traits. *Trends Ecol. Evol.* **21**, 178–185 (2006).
- 501 31. P. I. Prado, M. Miranda, A. Chalom, *sads: R package for fitting species abundance distributions*  
502 (Github, 2018; <https://github.com/piLaboratory/sads>).
- 503 32. T. J. Matthews, K. A. Triantis, R. J. Whittaker, F. Guilhaumon, *sars: an R package for fitting,*  
504 *evaluating and comparing species–area relationship models.* *Ecography*. **42**, 1446–1455 (2019).
- 505 33. F. He, S. P. Hubbell, Species–area relationships always overestimate extinction rates from habitat  
506 loss. *Nature*. **473**, 368–371 (2011).
- 507 34. R. J. H. & J. van Etten, *raster: Geographic analysis and modeling with raster data* (2012),  
508 (available at <http://CRAN.R-project.org/package=raster>).
- 509 35. M. Exposito-Alonso, 500 Genomes Field Experiment Team, H. A. Burbano, O. Bossdorf, R.  
510 Nielsen, D. Weigel, Natural selection in the *Arabidopsis thaliana* genome in present and future  
511 climates. *Nature*. **573**, 126–129 (2019).
- 512 36. Y. B. Simons, K. Bullaughey, R. R. Hudson, G. Sella, A population genetic interpretation of  
513 GWAS findings for human quantitative traits. *PLoS Biol.* **16**, e2002985 (2018).
- 514 37. M. Exposito-Alonso, C. Becker, V. J. Schuenemann, E. Reiter, C. Setzer, R. Slovak, B. Brachi, J.  
515 Hagemann, D. G. Grimm, J. Chen, W. Busch, J. Bergelson, R. W. Ness, J. Krause, H. A. Burbano,  
516 D. Weigel, The rate and potential relevance of new mutations in a colonizing plant lineage. *PLoS*  
517 *Genet.* **14**, e1007155 (2018).
- 518 38. H. Seebens, T. M. Blackburn, E. E. Dyer, P. Genovesi, P. E. Hulme, J. M. Jeschke, S. Pagad, P.  
519 Pyšek, M. Winter, M. Arianoutsou, S. Bacher, B. Blasius, G. Brundu, C. Capinha, L. Celesti-

- 520 Grapow, W. Dawson, S. Dullinger, N. Fuentes, H. Jäger, J. Kartesz, M. Kenis, H. Kreft, I. Kühn,  
521 B. Lenzner, A. Liebhold, A. Mosen, D. Moser, M. Nishino, D. Pearman, J. Pergl, W. Rabitsch,  
522 J. Rojas-Sandoval, A. Roques, S. Rorke, S. Rossinelli, H. E. Roy, R. Scalera, S. Schindler, K.  
523 Štajerová, B. Tokarska-Guzik, M. van Kleunen, K. Walker, P. Weigelt, T. Yamanaka, F. Essl,  
524 No saturation in the accumulation of alien species worldwide. *Nat. Commun.* **8**, 14435 (2017).
- 525 39. H. Seebens, F. Essl, W. Dawson, N. Fuentes, D. Moser, J. Pergl, P. Pyšek, M. van Kleunen, E.  
526 Weber, M. Winter, B. Blasius, Global trade will accelerate plant invasions in emerging  
527 economies under climate change. *Glob. Chang. Biol.* **21**, 4128–4140 (2015).
- 528 40. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P.  
529 Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: a tool set for whole-genome  
530 association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- 531 41. C. Mérot, R. A. Oomen, A. Tigano, M. Wellenreuther, A Roadmap for Understanding the  
532 Evolutionary Significance of Structural Genomic Variation. *Trends Ecol. Evol.* **35**, 561–572  
533 (2020).
- 534 42. K. Lucek, Y. Willi, Drivers of linkage disequilibrium across a species' geographic range. *PLoS*  
535 *Genet.* **17**, e1009477 (2021).
- 536 43. J. M. Kreiner, D. A. Giacomini, F. Bemm, B. Waithaka, J. Regalado, C. Lanz, J. Hildebrandt, P.  
537 H. Sikkema, P. J. Tranel, D. Weigel, J. R. Stinchcombe, S. I. Wright, Multiple modes of  
538 convergent adaptation in the spread of glyphosate-resistant *Amaranthus tuberculatus*. *Proc. Natl.*  
539 *Acad. Sci. U. S. A.* **116**, 21076–21084 (2019).
- 540 44. M. A. Supple, J. G. Bragg, L. M. Broadhurst, A. B. Nicotra, M. Byrne, R. L. Andrew, A.  
541 Widdup, N. C. Aitken, J. O. Borevitz, Landscape genomic prediction for restoration of a  
542 *Eucalyptus* foundation species under climate change. *Elife.* **7** (2018), doi:10.7554/eLife.31835.
- 543 45. M. Vallejo-Marín, J. Friedman, A. D. Twyford, O. Lepais, S. M. Ickert-Bond, M. A. Streisfeld,  
544 L. Yant, M. van Kleunen, M. C. Rotter, J. R. Puzey, Population genomic and historical analysis  
545 suggests a global invasion by bridgehead processes in *Mimulus guttatus*. *Commun Biol.* **4**, 327  
546 (2021).
- 547 46. J. T. Lovell, A. H. MacQueen, S. Mamidi, J. Bonnette, J. Jenkins, J. D. Napier, A. Sreedasyam,  
548 A. Healey, A. Session, S. Shu, K. Barry, S. Bonos, L. Boston, C. Daum, S. Deshpande, A.  
549 Ewing, P. P. Grabowski, T. Haque, M. Harrison, J. Jiang, D. Kudrna, A. Lipzen, T. H.  
550 Pendergast 4th, C. Plott, P. Qi, C. A. Sasaki, E. V. Shakirov, D. Sims, M. Sharma, R. Sharma, A.  
551 Stewart, V. R. Singan, Y. Tang, S. Thibivillier, J. Webber, X. Weng, M. Williams, G. A. Wu, Y.  
552 Yoshinaga, M. Zane, L. Zhang, J. Zhang, K. D. Behrman, A. R. Boe, P. A. Fay, F. B. Fritschi, J.  
553 D. Jastrow, J. Lloyd-Reilley, J. M. Martínez-Reyna, R. Matamala, R. B. Mitchell, F. M.  
554 Rouquette Jr, P. Ronald, M. Saha, C. M. Tobias, M. Udvardi, R. A. Wing, Y. Wu, L. E. Bartley,  
555 M. Casler, K. M. Devos, D. B. Lowry, D. S. Rokhsar, J. Grimwood, T. E. Juenger, J. Schmutz,  
556 Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature* (2021),  
557 doi:10.1038/s41586-020-03127-1.
- 558 47. I. R. MacLachlan, T. K. McDonald, B. M. Lind, L. H. Rieseberg, S. Yeaman, S. N. Aitken,  
559 Genome-wide shifts in climate-related variation underpin responses to selective breeding in a  
560 widespread conifer. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021), doi:10.1073/pnas.2016900118.

- 561 48. G. Tuskan, W. Muchero, J.-G. Chen, D. Jacobson, T. Tschaplinski, D. Rokhsar, W. Schackwitz,  
562 J. Schmutz, S. DiFazio, Populus Trichocarpa Genome-Wide Association Study (GWAS)  
563 Population SNP Dataset Released, (available at <https://doi.ccs.ornl.gov/ui/doi/55>).
- 564 49. Anopheles gambiae 1000 Genomes Consortium, Data analysis group, Partner working group,  
565 Sample collections—Angola:, Burkina Faso:, Cameroon:, Gabon:, Guinea:, Guinea-Bissau:,  
566 Kenya:, Uganda:, Crosses:, Sequencing and data production, Web application development,  
567 Project coordination, Genetic diversity of the African malaria vector Anopheles gambiae.  
568 *Nature*. **552**, 96–100 (2017).
- 569 50. Z. L. Fuller, V. J. L. Mocellin, L. A. Morris, N. Cantin, J. Shepherd, L. Sarre, J. Peng, Y. Liao, J.  
570 Pickrell, P. Andolfatto, M. Matz, L. K. Bay, M. Przeworski, Population genetics of the coral  
571 Acropora millepora: Toward genomic prediction of bleaching. *Science*. **369** (2020),  
572 doi:10.1126/science.aba4674.
- 573 51. K. Ruegg, E. C. Anderson, M. Somveille, R. A. Bay, M. Whitfield, E. H. Paxton, T. B. Smith,  
574 Linking climate niches across seasons to assess population vulnerability in a migratory bird.  
575 *Glob. Chang. Biol.* **27**, 3519–3531 (2021).
- 576 52. R. A. Bay, R. J. Harrigan, V. Le Underwood, H. Lisle Gibbs, T. B. Smith, K. Ruegg, Genomic  
577 signals of selection predict climate-driven population declines in a migratory bird. *Science*. **359**,  
578 83–86 (2018).
- 579 53. E. P. Kingsley, K. M. Kozak, S. P. Pfeifer, D.-S. Yang, H. E. Hoekstra, The ultimate and  
580 proximate mechanisms driving the evolution of long tails in forest deer mice. *Evolution*. **71**, 261–  
581 273 (2017).
- 582 54. L. Smeds, J. Aspi, J. Berglund, I. Kojola, K. Tirronen, H. Ellegren, Whole-genome analyses  
583 provide no evidence for dog introgression in Fennoscandian wolf populations. *Evol. Appl.* **14**,  
584 721–734 (2021).
- 585 55. R. M. Schweizer, J. Robinson, R. Harrigan, P. Silva, M. Galverni, M. Musiani, R. E. Green, J.  
586 Novembre, R. K. Wayne, Targeted capture and resequencing of 1040 genes reveal  
587 environmentally driven functional variation in grey wolves. *Mol. Ecol.* **25**, 357–379 (2016).
- 588 56. 1000 Genomes Project Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M.  
589 Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis, A global  
590 reference for human genetic variation. *Nature*. **526**, 68–74 (2015).
- 591 57. J. D. Palacio-Mejía, P. P. Grabowski, E. M. Ortiz, G. A. Silva-Arias, T. Haque, D. L. Des  
592 Marais, J. Bonnette, D. B. Lowry, T. E. Juenger, Geographic patterns of genomic diversity and  
593 structure in the C4 grass *Panicum hallii* across its natural distribution. *AoB Plants*. **13**, lab002  
594 (2021).
- 595 58. A. M. Royer, M. A. Streisfeld, C. I. Smith, Population genomics of divergence within an obligate  
596 pollination mutualism: Selection maintains differences between Joshua tree species. *Am. J. Bot.*  
597 **103**, 1730–1741 (2016).
- 598 59. M. Kapun, J. C. B. Nunez, M. Bogaerts-Márquez, J. Murga-Moreno, M. Paris, J. Outten, M.  
599 Coronado-Zamora, C. Tern, O. Rota-Stabelli, M. P. García Guerreiro, S. Casillas, D. J. Orengo,



- 600 E. Puerma, M. Kankare, L. Ometto, V. Loeschcke, B. S. Onder, J. K. Abbott, S. W. Schaeffer, S.  
601 Rajpurohit, E. L. Behrman, M. F. Schou, T. J. S. Merritt, B. P. Lazzaro, A. Glaser-Schmitt, E.  
602 Argyridou, F. Staubach, Y. Wang, E. Tauber, S. V. Serga, D. K. Fabian, K. A. Dyer, C. W.  
603 Wheat, J. Parsch, S. Grath, M. S. Veselinovic, M. Stamenkovic-Radak, M. Jelic, A. J. Buendía-  
604 Ruíz, M. Josefa Gómez-Julián, M. Luisa Espinosa-Jimenez, F. D. Gallardo-Jiménez, A.  
605 Patenkovic, K. Eric, M. Tanaskovic, A. Ullastres, L. Guio, M. Merenciano, S. Guirao-Rico, V.  
606 Horváth, D. J. Obbard, E. Pasyukova, V. E. Alatortsev, C. P. Vieira, J. Vieira, J. Roberto Torres,  
607 I. Kozeretska, O. M. Maistrenko, C. Montchamp-Moreau, D. V. Mukha, H. E. Machado, A.  
608 Barbadilla, D. Petrov, P. Schmidt, J. Gonzalez, T. Flatt, A. O. Bergland, *Drosophila Evolution*  
609 *over Space and Time (DEST) - A New Population Genomics Resource. bioRxiv* (2021), p.  
610 2021.02.01.428994.
- 611 60. L. N. Di Santo, S. Hoban, T. L. Parchman, J. W. Wright, J. A. Hamilton, Reduced representation  
612 sequencing to understand the evolutionary history of Torrey pine (*Pinus torreyana* Parry) with  
613 implications for rare species conservation. *bioRxiv* (2021), p. 2021.07.02.450939.
- 614 61. J. von Seth, N. Dussex, D. Díez-Del-Molino, T. van der Valk, V. E. Kutschera, M. Kierczak, C.  
615 C. Steiner, S. Liu, M. T. P. Gilbert, M.-H. S. Sinding, S. Prost, K. Guschanski, S. K. S. S.  
616 Nathan, S. Brace, Y. L. Chan, C. W. Wheat, P. Skoglund, O. A. Ryder, B. Goossens, A.  
617 Götherström, L. Dalén, Genomic insights into the conservation status of the world's last  
618 remaining Sumatran rhinoceros populations. *Nat. Commun.* **12**, 2393 (2021).
- 619 62. Millennium Ecosystem Assessment, *Millennium ecosystem assessment* (Millennium Ecosystem  
620 Assessment, 2005; <http://chapter.ser.org/europe/files/2012/08/Harris.pdf>).
- 621 63. G. C. Hurtt, L. Chini, R. Sahajpal, S. Frolking, B. L. Bodirsky, K. Calvin, J. C. Doelman, J. Fisk,  
622 S. Fujimori, K. Klein Goldewijk, T. Hasegawa, P. Havlik, A. Heinemann, F. Humpenöder, J.  
623 Jungclaus, J. O. Kaplan, J. Kennedy, T. Krisztin, D. Lawrence, P. Lawrence, L. Ma, O. Mertz, J.  
624 Pongratz, A. Popp, B. Poulter, K. Riahi, E. Shevliakova, E. Stehfest, P. Thornton, F. N. Tubiello,  
625 D. P. van Vuuren, X. Zhang, Harmonization of global land use change and management for the  
626 period 850–2100 (LUH2) for CMIP6. *Geosci. Model Dev.* **13**, 5425–5464 (2020).
- 627 64. S. Theodoridis, C. Rahbek, D. Nogues-Bravo, Exposure of mammal genetic diversity to mid-21st  
628 century global change. *Ecography (Cop.)* (2021), doi:10.1111/ecog.05588.
- 629 65. I. Overcast, M. Ruffley, J. Rosindell, L. Harmon, P. A. V. Borges, B. C. Emerson, R. S. Etienne,  
630 R. Gillespie, H. Krehenwinkel, D. Luke Mahler, F. Massol, C. E. Parent, J. Patiño, B. Peter, B.  
631 Week, C. Wagner, M. J. Hickerson, A. Rominger, A unified model of species abundance, genetic  
632 diversity, and functional diversity reveals the mechanisms structuring ecological communities.  
633 *bioRxiv* (2020), p. 2020.01.30.927236.
- 634 66. S. Díaz, N. Zafra-Calvo, A. Purvis, P. H. Verburg, D. Obura, P. Leadley, R. Chaplin-Kramer, L.  
635 De Meester, E. Dulloo, B. Martín-López, M. R. Shaw, P. Visconti, W. Broadgate, M. W.  
636 Bruford, N. D. Burgess, J. Cavender-Bares, F. DeClerck, J. M. Fernández-Palacios, L. A.  
637 Garibaldi, S. L. L. Hill, F. Isbell, C. K. Khoury, C. B. Krug, J. Liu, M. Maron, P. J. K.  
638 McGowan, H. M. Pereira, V. Reyes-García, J. Rocha, C. Rondinini, L. Shannon, Y.-J. Shin, P.  
639 V. R. Snelgrove, E. M. Spehn, B. Strassburg, S. M. Subramanian, J. J. Tewksbury, J. E. M.  
640 Watson, A. E. Zanne, Set ambitious goals for biodiversity and sustainability. *Science.* **370**, 411–  
641 413 (2020).

- 642 67. C. Rahbek, R. K. Colwell, Biodiversity: Species loss revisited. *Nature*. **473** (2011), pp. 288–289.
- 643 68. H. M. Pereira, G. C. Daily, Modeling biodiversity dynamics in countryside landscapes. *Ecology*.  
644 **87**, 1877–1885 (2006).
- 645 69. J. Harte, A. B. Smith, D. Storch, Biodiversity scales from plots to biomes with a universal  
646 species-area curve. *Ecol. Lett.* **12**, 789–797 (2009).
- 647 70. D. R. Lockwood, C. M. Richards, G. M. Volk, Probabilistic models for collecting genetic  
648 diversity: Comparisons, caveats, and limitations. *Crop Sci.* **47**, 861–866 (2007).
- 649 71. L. Czech, M. Exposito-Alonso, grenepipe: A flexible, scalable, and reproducible pipeline to  
650 automate variant and frequency calling from sequence reads. *arXiv* (2021), (available at  
651 <http://arxiv.org/abs/2103.15167>).
- 652 72. A. Miraldo, S. Li, M. K. Borregaard, A. Flórez-Rodríguez, S. Gopalakrishnan, M. Rizvanovic, Z.  
653 Wang, C. Rahbek, K. A. Marske, D. Nogués-Bravo, An Anthropocene map of genetic diversity.  
654 *Science*. **353**, 1532–1535 (2016).
- 655 73. D. H. Parks, T. Mankowski, S. Zangooei, M. S. Porter, D. G. Armanini, D. J. Baird, M. G. I.  
656 Langille, R. G. Beiko, GenGIS 2: Geospatial Analysis of Traditional and Genetic Biodiversity,  
657 with New Gradient Algorithms and an Extensible Plugin Framework. *PLoS One*. **8**, e69885  
658 (2013).
- 659 74. C. Li, D. Tian, B. Tang, X. Liu, X. Teng, W. Zhao, Z. Zhang, S. Song, Genome Variation Map: a  
660 worldwide collection of genome variations across multiple species. *Nucleic Acids Res.* **49**,  
661 D1186–D1191 (2021).
- 662 75. M. Kardos, E. Armstrong, S. W. Fitzpatrick, S. Hauser, P. Hedrick, J. Miller, D. Tallmon, W.  
663 Chris Funk, The crucial role of genome-wide genetic variation in conservation. *bioRxiv* (2021),  
664 p. 2021.07.05.451163.
- 665 76. D. Spielman, B. W. Brook, R. Frankham, Most species are not driven to extinction before  
666 genetic factors impact them. *Proceedings of the National Academy of Sciences*. **101** (2004), pp.  
667 15261–15264.
- 668 77. Ipbes, *Global assessment report of the Intergovernmental Science-Policy Platform on*  
669 *Biodiversity and Ecosystem Services* (IPBES Secretariat, Bonn, Germany, 2019;  
670 <https://www.ipbes.net/news/ipbes-global-assessment-summary-policymakers-pdf>).

671



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26

## Supplementary Materials for

### **Genetic diversity loss in the Anthropocene**

Moises Exposito-Alonso, Tom R. Booker, Lucas Czech, Tadashi Fukami, Lauren Gillespie, Shannon Hateley, Christopher C. Kyriazis, Patricia L. M. Lang, Laura Leventhal, David Nogues-Bravo, Veronica Pagowski, Megan Ruffley, Jeffrey P. Spence, Sebastian E. Toro Arana, Clemens L. Weiß, Erin Zess.

Correspondence to: [moisesexpositoalonso@gmail.com](mailto:moisesexpositoalonso@gmail.com)

#### **This PDF file includes:**

Supplementary Methods  
Supplementary Results  
Figs. S1 to S27  
Tables S1 to S18

27		
28	<b><u>Table of contents</u></b>	
29		
30	<b>SUPPLEMENTAL METHODS</b>	5
31	<b>I. Background on species biodiversity and biogeography</b>	5
32	I.1 Theoretical models of biodiversity	5
33	Fig. S1   Example of typical plots used for species abundance curve studies	5
34	I.1.2. Niche apportionment approaches	5
35	I.1.2. Niche statistical approaches of species sampling	6
36	Fig. S2   Summary of theoretical models of Species Abundance Curves.	7
37	I.2 Metric of species diversity	7
38	I.3 Biogeography of species and extinction.	7
39	Fig. S3   Example of a Species-Area Relationship in Galapagos Islands	8
40	I.4 Estimating extinction of species from the species area relationship	8
41	<b>II. Population genetics models and the site frequency spectrum.</b>	8
42	II.1 The Wright-Fisher model and the site frequency spectrum	8
43	Fig. S4   Similarity between the Species Abundance Distribution and the Site Frequency Spectrum	10
44	II.2 Metrics of genetic diversity	10
45	II.3 Spatial genetics and the mutations-area relationship (MAR)	11
46	II.3.1 Panmictic population	11
47	Fig. S5   Expected ranges of z MAR given sample sizes.	14
48	II.3.2 Scaling zMAR for low sampling and low census size	14
49	II.3.3 Meta-populations in space	14
50	Fig. S6   msprime 2D deme simulations and the mutations-area relationship	15
51	Table S1   msprime population genetic simulations in 2D	15
52	II.3.4 Meta-populations in space with local adaptation	15
53	Fig. S7   SLiM population genetic simulations in 2D with selection and local adaptation	16
54	Table S2   Linear model explaining zMAR by migration rate and natural selection	16
55	II.3.5. Meta-populations in space with purifying selection	16
56	Fig. S8   SLiM population genetic simulations in 2D with purifying selection	17
57	II.3.6 Continuous-space non-Wright-Fisher models	17
58	Fig. S9   Continuous space SLiM population genetic simulations	18
59	II.3.7 Connection of zMAR with the isolation-by-distance pattern	18
60	Fig. S10   SLiM population genetic simulations in 2D comparing FST and zMAR	19
61	II.4 The loss of mutations (genetic diversity) in space	19
62	II.5 Recovery of genetic diversity after a bottleneck or local extinction	20
63	Fig. S11   2D stepping-stone msprime simulations with extinction and recovery	20
64	<b>SUPPLEMENTAL RESULTS</b>	21
65	<b>III. The mutations-area relationship with the 1001 Arabidopsis Genomes</b>	21
66	III.1 The Site Frequency Spectrum of the 1001 Arabidopsis Genomes	21
67	Fig. S12   Mutation abundance study in <i>A. thaliana</i>	21

68	Fig. S13   Fit of mutation abundance study in <i>A. thaliana</i> with different SAD models	22
69	Table S3   AIC values for model fit of common species distribution curves.	22
70	III.2 Building the Mutations-Area Relationship	23
71	Table S4   Different SAR curves fit to mutations.	23
72	Table S5   The mutations-area relationship (MAR).	24
73	Table S6   The endemic-mutations-area relationship (EMAR).	24
74	Fig. S14   The mutations-area and endemic-mutations-area relationships in <i>A. thaliana</i> .	24
75	III.3 Testing numerical artefacts	24
76	Table S7   MAR built with different area calculations and grid sizes	25
77	Fig. S15   Cartoon of raster sampling to build the MAR	26
78	Fig. S16   MAR comparison with different area calculations.	26
79	Fig. S17   MAR and EMAR in <i>Arabidopsis thaliana</i> using outward and inward sampling.	27
80	Table S8   Outward and inward MAR and EMAR	27
81	Table S9   MAR for putatively neutral, deleterious, and locally adaptive alleles in <i>Arabidopsis thaliana</i>	27
82		28
83	III.4 Local population extinction in <i>Arabidopsis</i>	28
84	Fig. S18   Loss of mutations with habitat loss in <i>A. thaliana</i> .	29
85	III.5 Potential impacts of genetic loss in adaptability	29
86	Fig. S19   Bias of low frequency mutations and effect size for fitness traits in <i>A. thaliana</i> .	30
87	Fig. S20   Simulations illustrating the potential loss of locally-adaptive mutations in <i>A. thaliana</i> .	31
88	Fig. S21   Extinction simulations showing proxies of adaptive capacity of <i>A. thaliana</i> .	31
89	III.6 Case study of a massive natural bottleneck	31
90	<b>IV. The mutations-area relationship in diverse species</b>	33
91	Fig. S22   MAR summaries across species.	43
92	Table S10   The mutations-area relationship across species. Extended Table 1	43
93	IV.1 Exclusion of species from global averages	43
94	Table S11   Mean zMAR and other summary statistics across species.	44
95	Table S12   Traits, life history, and other characteristics of the analyzed species.	44
96	Table S13   Association of traits, life history, and other characteristics with zMAR.	45
97	<b>V. An estimate of global genetic diversity loss</b>	46
98	V.1 Estimates of ecosystem area losses	46
99	Table S14   Millennium Ecosystem Assessment land cover transformation.	46
100	Table S15   IPBES land cover transformation,	46
101	Table S16   Land Use Harmonization 2 from 1850 to 2015	47
102	Table S17   IUCN Red List categories of extinction risk and number of species.	47
103	V.2 A global estimate of genetic loss	48
104	Fig. S23   The parameter space of genetic diversity loss, extended	49
105	Table S18   Estimates of average expected genetic loss for different ecosystems.	49
106	V.3 Community ecology simulations and MAR	51
107	Fig. S24   zMAR calculated from MESS eco-evolutionary simulations	51
108	V.4 The nested species extinction and genetic diversity loss processes	52
109	Fig. S25   Cartoon of nested extinction of species and genetic diversity loss.	52

110	Fig. S26   The distribution of per-species area lost and total ecosystem extinction with 1000 species	53
111	Fig. S27   Numeric simulation of nested species and genetic diversity loss.	54
112	<b>VI. Limitations and outlook</b>	55
113	VI.1 Reasons for overestimations	55
114	VI.2 Reasons for underestimations	55
115	VI.3 Final notes	56
116	<b>SUPPLEMENTAL REFERENCES</b>	57
117	<hr/>	
118		
119		

## 120 SUPPLEMENTAL METHODS

121

### 122 I. Background on species biodiversity and biogeography

#### 123 I.1 Theoretical models of biodiversity

124

125 Studies in biogeography have modelled the species-area relationship with several functions.

126 Below we summarise the different approaches using an example of richness of  $S = 100$

127 species, with variable abundance or area,  $A$ .

128

129 We may visualise the different areas or abundances of species as a frequency

130 histogram (Fig. S1, Preston plot), with x-axis: logarithm of abundance bins (historically  $\log_2$

131 as a rough approximation to the natural logarithm), and y-axis: number of species at given

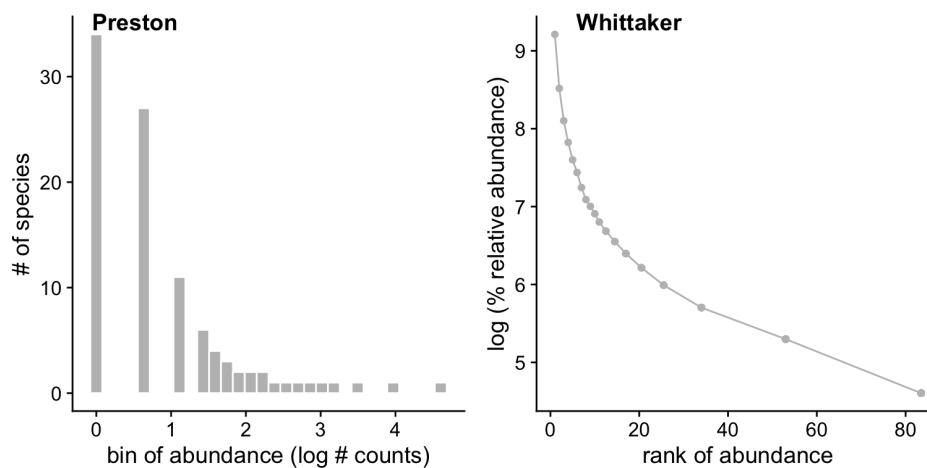
132 abundance. Alternatively, as a rank-abundance diagram (Fig. S1, Whittaker plot): x-axis:

133 species list, ranked in order of descending abundance (i.e. from common to rare), and y-axis:

134 logarithm of % relative abundance.

135

136



137

138 ***Fig. S1 | Example of typical plots used for species abundance curve studies***

139 *Due to their strong skew, Species Abundance Curves are often plotted using the Preston plot (left) where the x axis*  
140 *represents bins of  $\log_2$  abundances (also referred to as octaves), or using the Whittaker plot (right) where the x axis is the*  
141 *rank of each species in a dataset and y axis the species' relative abundance.*

142

143

144

#### 145 I.1.2. Niche apportionment approaches

146

147 A series of theoretical deterministic and stochastic "niche apportionment models" have been  
148 put forward (summarised in (1) or (2, 3)).

149

150 The Motomura (4) geometric series suggests that each species that arrives takes half  
151 the area. The first would take 50%, the second 50% of 50%, and so forth, which can be  
152 expressed as:

153

$$154 P_i = 0.5^i$$

155

156 Similarly, one can imagine that as a species colonises a habitat, it takes up a fraction different  
157 than 50%. This gives a geometric series with parameters  $k$  which can be written as

158

159  $P_i = k(1 - k)^{i-1}$ .

160

161 Other geometric series-related models include stochasticity, where  $k$  instead of being a fixed  
162 parameter is a random uniform variable and there is a  $k_i$  each time  $i$  a new species arrives to  
163 the ecosystem. The "dominance preemption" model draws from 50-100% at any new arrival  
164 of a species, the random fraction model draws from 0-100%. Then the abundance of a species  
165 depends on the stochastic process of previous  $f = 1 \dots i-1$  species arriving first:

166

$$E[P_i | P_1, \dots, P_{i-1}, k_i] = k_i \times \left(1 - \sum_{f=1}^{i-1} P_f\right)$$

167

168

169 Another approach is the broken stick by MacArthur (5), which theorised a habitat is broken  
170 into  $S-1$  places at random, which creates  $S$  fractions of an area. Then the relative area of a  
171 species is:

172

$$E[P_i] = \left(\frac{1}{S}\right) \sum_{w=i}^S \frac{1}{w}$$

173

174

175

### 176 **I.1.2. Niche statistical approaches of species sampling**

177

178 Differently from niche partitioning functions, statistical approaches such as the log-series  
179 from Fisher (6) and log-normal from Preston (7) are probability distributions, and approach  
180 modelling in a conceptually different way: they model the sampling process of species  
181 collections given an underlying relative abundance (see below).

182

183

184 Statistical-based derivations probably began with Fisher (6), with the log-series  
185 distribution. It assumes that species abundances in the community are independent identically  
186 distributed variables, sampling is a Poisson process, sampling is done with replacement, or  
187 the fraction sampled is small enough to approximate a sample with replacement. Here,

188

$$189 S_n = \frac{\alpha x^n}{n},$$

190

191 where  $x$  is a constant  $x \in [0, 1]$  related to the sample dataset (typically close to 1),

192  $x = \frac{N}{\alpha + N}$ , and  $\alpha$  is a new constant term (ecosystem-specific) that is used as a measure of  
193 biodiversity. Fisher proposed the number of species could be estimated as:

194

$$195 S = \alpha \times \log\left(1 + \frac{N}{\alpha}\right).$$

196

197 Finally, Preston (8) posed that the skewness of previous proposals is due to lack of  
198 sampling. With little data, common species are collected sooner, but with more abundant  
199 sampling, the rarest species are also well-sampled and have abundances well above 0. Preston  
200 then proposed that the octaves (bins of doubling abundance) follow a normal distribution,  
201 making the raw abundance log-normal distributed. Given  $S_0$  is the number of species in the  
202 model octave of abundance and a variance composite of the log-Normal  $\sigma^2$ , the number of  
203 species per abundance (octave) bin  $R (= \log(n))$  is:

204



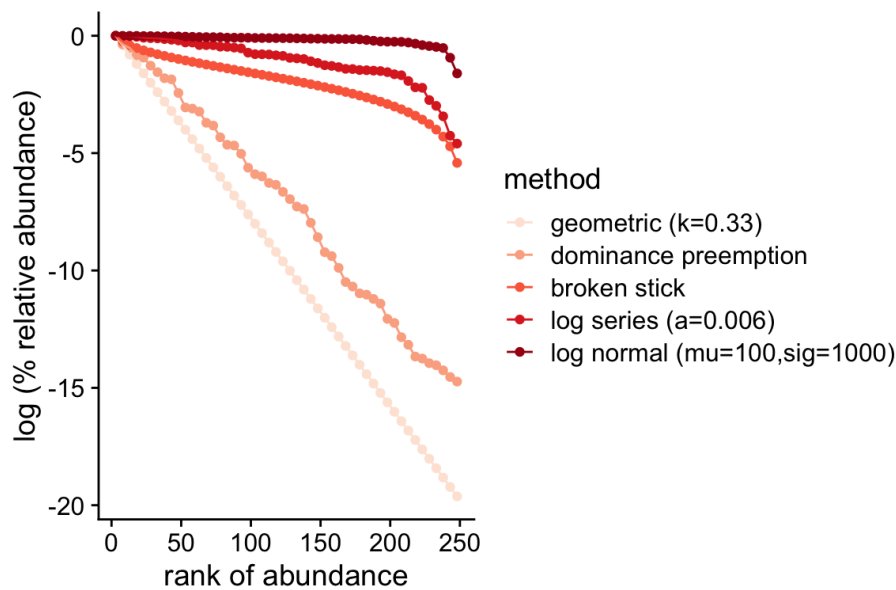
205  $S_R = S_0 e^{-R^2/2\sigma^2}$ .

206

207 The Unified Neutral Theory of Biodiversity (UNTB) by Hubbell (1) takes a stochastic  
208 approach of a community with immigrants, extinctions, and speciation in continuous  
209 dynamics. Interestingly, the UNTB's key parameter,  $\theta$ , coincides with Fisher's  $\alpha$ , as the log-  
210 series is a limiting case of UNTB. Hubbell's discovery was that  $\alpha=2J_m v$ , where  $J_m$  is the size  
211 of the external metacommunity that provides migrants of species to the focal community, and  
212  $v$  is the speciation rate. Alonso and McKane (9) derived the so-called Metacommunity Zero-  
213 Sum Multinomial (MZSM) distribution from the UNTB. In practice, both distributions have  
214 almost-identical fits (lines completely overlapping in Fig. S2 below).

215

216



217

218 ***Fig. S2 | Summary of theoretical models of Species Abundance Curves.***

219 *Five niche partitioning or statistical models shown in a Whittaker plot. The different models expect different levels of*  
220 *evenness in abundance across the species in the community, from the lowest (geometric series) to the highest (log-normal).*

221

222

## 223 I.2 Metric of species diversity

224

225 Although a number of metrics exist to measure species diversity, such as the Shannon index,  
226  $H' = -\sum_{i=1}^S P_i \log P_i$  (with  $P_i$  the relative proportions of species abundances) or Fisher's non-  
227 dimensional  $\alpha$  parameter, the study of species abundances and area relationships has focused  
228 on species richness  $S$ , that is, the total number of species in a given location or area. Below  
229 we therefore focus on species richness.

230

231

## 232 I.3 Biogeography of species and extinction.

233

### 234 SAD and SAR connection

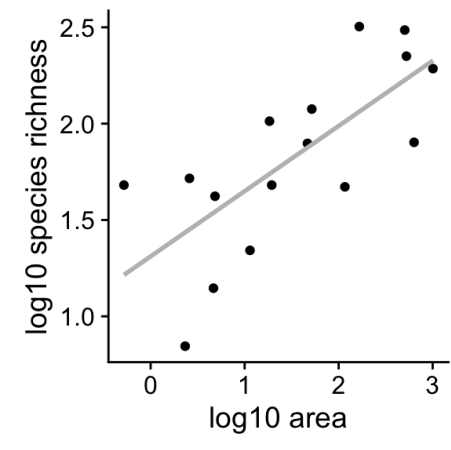
235

236 Due to many species being rare, it is expected that as researchers sample an area, the most  
237 common species will be sampled first, and as the area studied increases, more and more  
238 species will be discovered. This is thought to happen following a power law relationship,

239 where the number of species in that area  $S_A$  increases with the sampled area  $A$ , with scaling  $z$   
 240 (slope in a log-log plot), and with a constant  $c$ :

241  
 242  $SAR(A) = cA^z$ .

243  
 244 Preston (7) derived theoretically that from a log-normal series, one would expect  
 245  $z=0.27$ , under a number of assumptions (Fig. S3). This has been empirically shown to be  
 246 close to reality (7, 10), although there is some variation across ecosystems and spatial scales.  
 247



248  
 249 **Fig. S3 | Example of a Species-Area Relationship in Galapagos Islands**  
 250 *Classic species richness dataset from the Galapagos Islands (Preston, 1962). It depicts species richness as a function of*  
 251 *island area in a log-log plot.*

252  
 253  
 254 **I.4 Estimating extinction of species from the species area relationship**

255  
 256 The first estimates of species extinction used the SAR relationship. Given a reduction of  
 257 ecosystem area,  $A$ , by an area of  $a$  (11, 12). If these areas, as well as the SAR scaling,  $z$ , are  
 258 known, then one can predict the number of species in the future as:

259  
 260  $S_{\text{now}} - S_{\text{fut}} = cA_{\text{now}}^z - cA_{\text{fut}}^z$ ,

261  
 262 However, we are normally interested in the fraction of species that will go extinct  $X_s$ , so we  
 263 can take the ratio:

264  
 265  $X_s = \frac{S_{\text{now}} - S_{\text{fut}}}{S_{\text{now}}} = 1 - \frac{cA_{\text{fut}}^z}{cA_{\text{now}}^z} = 1 - \left(\frac{A_{\text{fut}}}{A_{\text{now}}}\right)^z$ .

266  
 267  
 268 **II. Population genetics models and the site frequency spectrum.**

269 **II.1 The Wright-Fisher model and the site frequency spectrum**

270  
 271 Statisticians and population geneticists from the 20<sup>th</sup> century, Wright and Fisher, built a  
 272 simple statistical model of evolution of a population. It assumes that each generation a  
 273 population of  $N$  monoecious (hermaphrodite) individuals mate randomly to create a new  
 274 generation of  $N$  individuals and then immediately die so that only  $N$  individuals remain in the  
 275 population at any given time. This random sampling process causes the frequency of a variant

276 in one generation to possibly differ from its frequency in the previous generation—a process  
277 known as genetic drift.

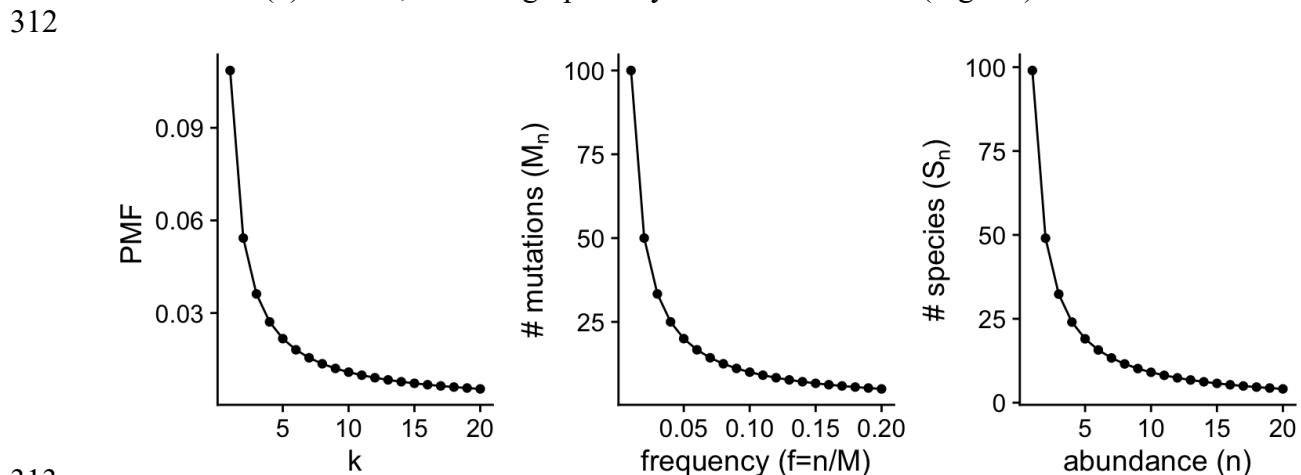
278  
279 When a nucleotide mutation or variant (e.g. ACGAA → ACGTA ) emerges by a  
280 random process of, for instance, DNA replication error, it will first be in  $1/N$  individuals (if  
281 we consider these diploid,  $1/2N$  chromosomes). Through random sampling that T mutation  
282 may be lost, stay at the same frequency, or randomly move to higher frequency. Although  
283 rarely, just by chance, the mutation may reach 100% frequency. This results in a  
284 “commonness of rarity” when looking at mutations in a population, as we have seen in  
285 previous sections for species. Since these genetic drift dynamics affect all mutations genome-  
286 wide, we therefore expect the majority of mutations to be absent, or rare, and only a much  
287 smaller proportion of variants to be at moderate or high frequencies.

288  
289 The site frequency spectrum (SFS) refers to the distribution of frequencies of variants  
290 in a population. This is the number of sites at which we observe a variant at frequency  $q$  in a  
291 sample of  $n$  individuals. To derive the expected SFS distribution, we turn to Kingman’s  
292 Coalescent (13). Both models describe the same ideal population of random mating, constant  
293 population size, and mutations emerging at a low rate and drifting in frequency. But while the  
294 Wright-Fisher model describes the dynamics of a whole population forward-in-time, the  
295 Kingman’s Coalescent describes the genealogy of a sample of individuals from a population,  
296 going backward in time. By building a model around the individuals that are sampled or that  
297 survived, rather than of an entire population, the Coalescent provides a simpler way to derive  
298 expectations in small populations or in cases, for example here, where a limited sample of  
299 genomes are sequenced. Using the Coalescent (see (14) for details), one obtains that the  
300 expected number of mutations of a given abundance,  $n$ , is inversely related to their frequency,  
301  $q$ :

302  
303 
$$M_n = c \frac{1}{q}$$

304  
305 for some constant  $c$  that depends on the mutation rate and the population size. This SFS from  
306 population genetics theory is remarkably similar to the Species Abundance Relationship. In  
307 fact, Fisher himself (15) derived an expression similar to the above.

308  
309 Rearranging terms, one can see this is a constrained version of the log-series  
310 Probability Mass Function (PMF), which Fisher also proposed for the distribution of species  
311 abundances (6). Below, one can graphically see the similarities (Fig. S4):



313

314 **Fig. S4 | Similarity between the Species Abundance Distribution and the Site Frequency Spectrum**

315 *Left is the Probability Mass Function of the log-series ( $p=0.999$ ), center is the SFS ( $N=100$ ,  $c=1$ ), and right is*  
316 *the log-series-based abundance of species ( $\alpha=100$ ,  $N=10000$ ).*

317

318 Keeping the abundance,  $n$ , constant (and low), when the number of individuals

319  $N \rightarrow \infty$ , we know that the constant  $x$  from Fisher's SAD approaches 1,  $x = \frac{N}{N+\alpha} \rightarrow 1$ . Then,  
320 we can rewrite the number of species at any given abundance ( $S_n$ ) as:

321

$$322 S_n = \alpha \frac{\left(\frac{N}{\alpha+N}\right)^n}{n} = \alpha \frac{1^n}{n} = c \frac{1}{n} = M_n$$

323

324 So both have the same form as the log series PMF:  $f(k) = \frac{-1}{\ln(1-p)} \frac{p^k}{k}$  when  $p \rightarrow 1$ . In  
325 the next section we will see that the constants of the SAD and the SFS are proportional to  
326 species and mutation diversity, although the Site Frequency Spectrum (SFS) is a specific case  
327 of SAD. One can also see that because the constant in the SFS is the population scaled  
328 mutation rate,  $c = \theta = Ne\mu$ , and Fisher's  $\alpha \approx \theta$  for large  $N$ .

329

330 **II.2 Metrics of genetic diversity**

331

332 In population genetics, multiple measurements of genetic diversity have been put forward.  
333 The most straightforward is the allelic richness, also number of mutations, or also called the  
334 number of segregating sites. Segregating sites,  $M$ , is the direct equivalent of the species  
335 richness,  $S$ , and it depends on the number of samples used and length of DNA sequence  
336 explored (Note: we use the non-standard notation,  $M$ , as the standard in population genetics is  
337  $S$  [for segregating sites] but this is already in use for species richness. We then use  $M$  for  
338 mutations and  $S$  for species). This metric can also be thought of as the area under the curve of  
339 the SFS. Two other metrics that describe the SFS but that aim to be sequence-length- and  
340 individual independent are Watterson's Theta,  $\theta_W$ , and Nucleotide diversity,  $\pi$ , (also called  $\theta_\pi$   
341 ). These two metrics of diversity are identical at population equilibrium and are estimates of  
342  $4Ne\mu$  (when the SFS follows a  $1/q$  relationship), with effective population size  $N_e$  and per-  
343 generation mutation rate  $\mu$ , whereas they differ in non-equilibrium demographics, under  
344 natural selection, or under other behaviors not considered in the Wright-Fisher neutral model,  
345 such as different mating systems (16).

346

347 First,  $\pi$  is described as:

348

$$349 \pi = \frac{\sum_{i=1}^{n-1} i(n-i)M_i}{n(n-1)/2},$$

350

351 and  $\theta_W$  as:

352

$$353 \theta_W = \frac{\sum_{i=1}^{n-1} M_i}{\sum_{i=1}^{n-1} 1/i},$$

354

355 where  $\sum_{i=1}^{n-1} 1/i$  is the  $n-1^{th}$  Harmonic number, which serves to scale the segregating  
356 sites based on the assumption that the abundance of mutations follows a  $1/q$  SFS. The  
357 diversity metrics  $\pi$  and  $\theta_W$  are both functions of the SFS, as opposed to Fisher's  $\alpha$  from the  
358 Species Abundance Distribution, which is a parameter that changes the shape of the  
359 distribution.

360

361 Although often nucleotide diversity  $\pi$  is reported as a typical measure of genetic  
362 diversity of a species, since it can be calculated for a single genome and it captures the  
363 process of inbreeding of a population (17), classic literature relating germplasm management  
364 for conservation and breeding has advocated for allelic richness (18).

365  
366

### 367 **II.3 Spatial genetics and the mutations-area relationship (MAR)**

368

369 Since its inception, a number of concepts in population genetics have dealt with genetic  
370 variation in populations of different sizes, or populations separated in space. For instance, one  
371 classic result in population genetics is the relationship of  $\pi \approx 4N_e \mu$ , which relates genetic  
372 diversity  $\pi$  with the effective population size  $N_e$  and the mutation rate of the species  $\mu$ . A  
373 relationship which is still studied nowadays in an effort to reconcile data with theory (17).

374 In 1943, Sewall Wright turned to study the genetics of multiple populations within a  
375 species. He proposed that populations sampled further apart geographically must differ more  
376 in allele frequency due to more independent drift (19), leading to the commonly used  
377 correlation between geographic distance and the metric of differentiation  $F_{ST}$ . Most  
378 prominently, the use of correlation in the accumulation of mutations of populations that are  
379 geographically close or share evolutionary history has been uncovered using dimensionality  
380 reduction approaches such as PCA (20).

381 Despite these enormous advances in understanding spatial genetic structures,  
382 surprisingly little quantitative work has been done to parametrize the loss of genetic diversity  
383 by direct loss of habitat.

384 Because of the abundance of rare mutations in populations, it is straightforward to  
385 think that the more area and individuals sampled, the more segregating sites will be found.  
386 Analogous to the Species Area Relationship (SAR),  $S=cA^z$ , we should thus be able to  
387 estimate the equivalent scaling for a mutations-area relationship (MAR):

388

$$389 \quad M=cA^z,$$

390

391 with a scaling  $z = z_{MAR}$ , which corresponds to the slope of best fit in a log-log-plot of  
392  $A$  and  $M$  for a given species. (Other functions are often fit empirically for SAR datasets,  
393 which we explore later in section III.3. We work with the power law because of its historical  
394 use, mathematical convenience, and because other more complicated functions only  
395 improved fitting marginally, see Table S4).

396

397 This differs from other efforts to understand the number of segregating sites or  
398 heterozygosity differences across species that differ in their total census size or geographic  
399 distribution (21, 22). The MAR instead is built within a species, as its ultimate aim is to relate  
400 the number of mutations left in a species as it loses spatial populations.

401

402 Below we derive what are the expectations of MAR taking two opposite scenarios of  
403 neutral population evolution, and study how many segregating sites or mutations  $M$  are  
404 discovered with increasing area in the simulations. We further test the scenario of meta-  
405 populations in space with varying migration rates and neutral or natural selection processes.

406

#### 407 **II.3.1 Panmictic population**

408

409 The expected number of mutations,  $M$ , is a constant that depends on the mutation rate,  $\mu$ , and  
410 the expected total branch length of the population genealogy,  $L$ , with  $M=\mu L$ . Under the

411 coalescent, the total branch length is equal to the number of lineages or individuals sampled  
 412 from the population,  $n$ , times the time of the genealogy during which there are such lineages,  
 413  $T_n$ , plus  $n-1$  times the time in the genealogy with such number of lineages, and so forth:

414  
 415 
$$L = nT_n + (n - 1)T_{n-1} + \dots + 2T_2.$$

416  
 417 Under the coalescent,

418  
 419 
$$E[T_n] = \frac{2N_e}{n(n-1)},$$

420  
 421 and thus:

422  
 423 
$$E[L] = n \frac{2N_e}{n(n-1)} + (n-1) \frac{2N_e}{(n-1)(n-2)} + \dots,$$

424  
 425 which simplifies to

426  
 427 
$$E[L] = 2N_e \left( \frac{1}{n-1} + \frac{1}{n-2} + \dots + 1 \right) = 2N_e H_{n-1},$$

428  
 429 where  $H_{n-1}$  is the  $(n-1)$ th harmonic number. This is of course related to one of the  
 430 diversity metrics (section II.2), where Watterson's  $\Theta_W$  scales the number of segregating sites  
 431 ( $M$ ) by the harmonic number of sampled individuals. This is based on the expectation that as  
 432 more individuals are sampled, we expect to discover more mutations proportional to the  
 433 above harmonic number. Because such number is not so easy to work with to create an  
 434 expectation for  $z_{MAR}$ , we further simplify this expectation following the Taylor expansion  
 435 approximation of the harmonic number:

436  
 437 
$$H_n = \gamma + \log(n) + \frac{1}{2n} + O\left(\frac{1}{n^2}\right) \simeq \gamma + \log(n) + \frac{1}{2n},$$

438  
 439 which we can further approximate as:

440  
 441 
$$E[L] \approx 2N_e \log(n-1) + c.$$

442  
 443 Therefore, assuming a constant mutation rate and effective population size ( $N_e$ ) under  
 444 panmixia,  $M$  grows following  $\log(n)$ . In such a case, a log-log plot (typical power law plot)  
 445 does not display a linear relationship, and the slope is asymptotic to  $z \rightarrow 0$  for  $N \rightarrow \infty$ . On the  
 446 other hand, with low values of  $x$  (area or individuals sampled close to 0), the slope  $z_{MAR}$  will  
 447 be incorrectly high. We can show this effect trivially by studying the local derivative of the  
 448 function  $\log_{10}(M) = \log_{10}(\log(N))$ . The local slope of that function is an approximation of our  
 449  $z_{MAR}$  parameter. This can be locally estimated at any given point  $N$  by taking the derivative:

450  
 451 
$$\frac{d \log_{10}(\log(N))}{d(\log_{10}(N))} = \frac{1}{\log_{10}(N) \log(10)}.$$

452  
 453 The implication of this nonlinear function is that if we sampled only few individuals  
 454 or areas of a species (e.g.,  $n=100$ ), even if this species was completely panmictic we would  
 455 expect a non-zero  $z_{MAR}$ , a value that will change with sampling effort. We can roughly approximate  
 456  $z_{MAR}$  by the local slope of the number in the midpoint of the graph, e.g., for  $n=100$  we look at  
 457 the slope at  $n=50$ , and obtain  $1/(\log_{10}(50) \times \log(10)) \cong 0.256$ . Therefore, with small sample  
 458 sizes, this parameter will not be helpful to understand whether a species behaves

459 panmictically or is limited by migration, which may be problematic for estimates of genetic  
460 diversity loss later. We can visualise our expectation of the  $z_{MAR}$  under panmixia plotting the  
461 first derivative above (Fig. S5). Because—as we will show below—we do expect a power  
462 law relationship under a migration-limited scenario,  $z_{MAR}$  should theoretically not change with  
463 sample size. The graphical study of the (non-)linearity of the log-log plots between the  
464 number of mutations and area sampled should be diagnostic to this problem (We see for  
465 instance that *Pinus contorta* has a highly nonlinear relationship, likely due to the use of  
466 ascertained intermediate frequency markers instead of genome-wide data, Fig. S22).

467

468 Finally, we used msprime (23) to corroborate this finding ( $z_{MAR}$  being constant with  
469 respect to sample size) with simulations, simulating 1600 demes in a 40x40 grid of demes or  
470 populations of  $N=N_e=1000$  that are completely panmictic (universal gene flow or dispersal,  
471 so this is equivalent to a single panmictic deme). We observed the  $z_{MAR}$  for  $t=100\dots 10,000$   
472 generations in  $\log_{10}$  increments. After this time, we sample  $n=1\dots 100$  individuals in  
473 increasingly large groups of adjacent demes. The range of estimates of  $z_{MAR}$  in these  
474 simulations was 0.07-0.15.

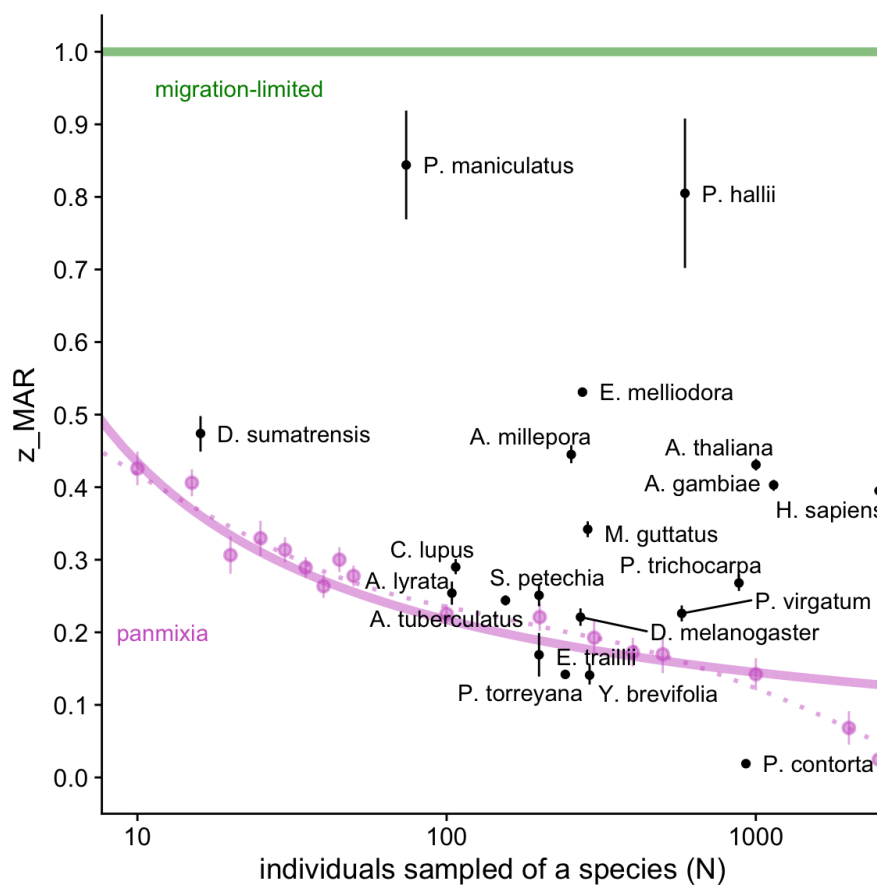
475

476 Fig. S5 indicates that the minimum average  $z_{MAR}$  even under panmixia would  
477 continuously increase with lower numbers of individuals of a species sampled. This is due to  
478 the fact that the site frequency spectrum is not fully sampled with small numbers of  
479 individuals. Therefore, we devised an approach to rescale  $z_{MAR}$ .

480

481

482



483

484 **Fig. S5 | Expected ranges of  $z_{MAR}$  given sample sizes.**

485 For increasing numbers of individuals sampled, we plot the expected mean  $z_{MAR}$  under two theoretical trends of a migration-  
486 limited (green) and a panmictic (purple) species (Purple dots indicate averages from SLiM simulations under panmixia to  
487 confirm the theoretical trend based on the derivative approach above). In black,  $z_{MAR}$  and 95% Confidence Interval of  
488 species analyzed in section IV are plotted (see section for details).

489

490

491 **II.3.2 Scaling  $z_{MAR}$  for low sampling and low census size**

492

493 Let  $z_{pan-n} = E[z_{MAR} | n, panmixia]$ , be the expected value of  $z_{MAR}$  of a panmictic  
494 species given that we only have small sampling of  $n$ . Although theoretically  $z_{MAR}$  should  
495 approach 0, with small samples it can be upwardly biased. In order to force the possible  
496 values of  $z_{MAR}$  to range 0-1 despite small sample sizes, we can scale it as:

497

$$498 z_{naive\ scaled} = (z_{MAR} - z_{pan-n}) / (1 - z_{pan-n}).$$

499

500 In words, this moves the purple line in Fig. S5 to zero, stretching the space above it  
501 accordingly.

502

503 Most species have census sizes so large that  $z_{MAR}$  should indeed approach 0 under  
504 panmixia, so we should correct the sample estimate  $z_{MAR}$  to range 0-1. However, some  
505 species have such low census size  $N$  that even if we sample all individuals of a species, the  
506 sample size will still be small. In those cases, we should not scale  $z_{MAR}$  to range 0-1, but  
507 rather scale it from  $z_{pan-N} - 1$ , where  $z_{pan-N} = E[z_{MAR} | N, panmixia]$  is the expected value of  
508  $z_{MAR}$  given a census size  $N$  (plants or animals living in the wild). The updated scaling  
509 approach for both census and sample size would then be:

510

$$511 z^*_{scaled} = (1 - z_{pan-N}) (z_{MAR} - z_{pan-n}) / (1 - z_{pan-n}) + z_{pan-N}.$$

512

513

514 Note that this scaled estimate must be conservative because while we adjust the  
515 minimum  $z$  for the average value expected for low sample sizes, we do not adjust for the  
516 maximum possible  $z$ , which only under very extraordinary theoretical conditions can be  $z=1$ ,  
517 namely under an unrealistic complete disconnection of populations by gene flow (see below).  
518 Because deriving the maximum  $z$  would require more biological knowledge of the species'  
519 demography, landscape connectivity, genome structure, etc., and because we rather create  
520 conservative estimates, we do not create further scaling approaches.

521

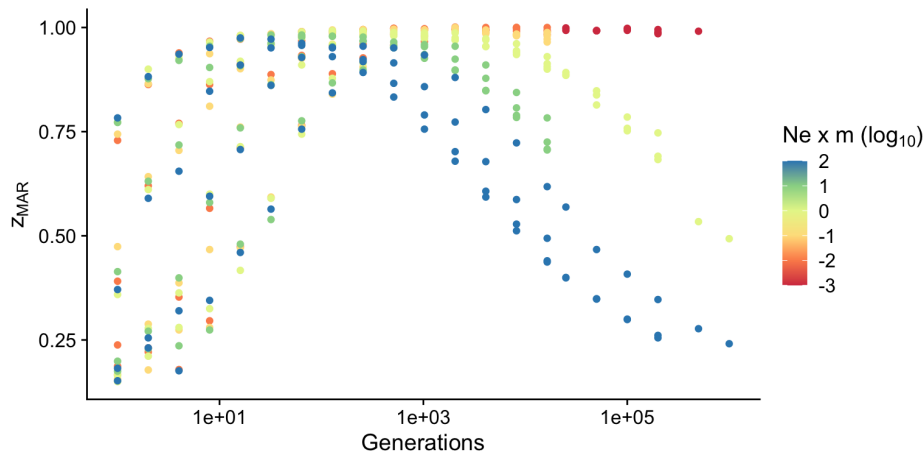
522 **II.3.3 Meta-populations in space**

523

524 A more realistic simulation than a panmictic population is that of the same 40x40 deme grid  
525 where migration can happen between adjacent demes. This migration rate can be changed to  
526 understand the effect of population structure and migration on  $z_{MAR}$ . Under no migration (or  
527 very low migration), we expect the mutations in two distinct populations (and thus their SFS)  
528 to be (almost) completely independent. Hence, when explored demes are doubled ( $N_e$   
529 doubles), we discover twice as many mutations. In this case, the number of mutations should  
530 scale linearly with the area, so we expect the following to be true:  $M=A$ ,  $\log(M) = \log(A)$ ,  
531 and  $z_{MAR}=1$ . Our analyses under different sampling schemes, and with different numbers of  
532 "burn-in generations" (generations since a single deme colonised the full 40x40 space)  
533 confirm that  $z_{MAR}$  approaches 1 in the limit of low migration (see Table S1 and Fig. S6).  
534 Different from the panmictic situation, as we increase the sampled area, we not only increase  
535  $n$ , which would lead to a  $\log(A)$  in mutations, but also increase  $N_e$ .



536  
537  
538



539

540 **Fig. S6 | msprime 2D deme simulations and the mutations-area relationship**

541 Simulations with different burn-in and migration rates under neutrality, and their corresponding  $z_{MAR}$ .  
542

543 **Table S1 | msprime population genetic simulations in 2D**

544 Simulations summarised by grouping ranges of the resulting  $z_{MAR}$  parameters. The average parameters of the simulations  
545 with similar  $z_{MAR}$  EW provided. (Acronyms:  $N_{emt}$  = product of effective population size, migration rate, and simulated  
546 generations).  
547

$z_{MAR}$	Samples/deme	Generations	Migration rate	$N_{emt}$
0.2 +/- 0.05	2.4	50001.7	0.0271675	5000044.23
0.3 +/- 0.05	20.25	70003	0.0561655	7000075.77
0.4 +/- 0.05	26.5714286	13057.4286	0.04450857	1305497.96
0.5 +/- 0.05	12.9230769	121759.462	0.04017769	752221.743
0.6 +/- 0.05	15.6111111	3218.77778	0.045735	321174.768
0.7 +/- 0.05	35.6842105	35034.8421	0.03395895	143791.614
0.8 +/- 0.05	35.030303	15655.1212	0.03055818	58023.5539
0.9 +/- 0.05	36.5806452	3057.12903	0.0253029	15290.4081
1 +/- 0.05	42.0140845	13625.4085	0.00861178	1798.36141

548

549

550 These simulations corroborated that we can recover  $z_{MAR}$  values ranging between 0-1  
551 just varying migration and burn-in generation parameters. We found that it was both the time  
552 of the system to reach an equilibrium as well as the migration rate that determined  $z_{MAR}$ . In  
553 the future, it will be interesting to study different non-equilibrium scenarios to better  
554 understand how genetic drift, gene flow, and different landscape structures may shape the  
555  $z_{MAR}$ .

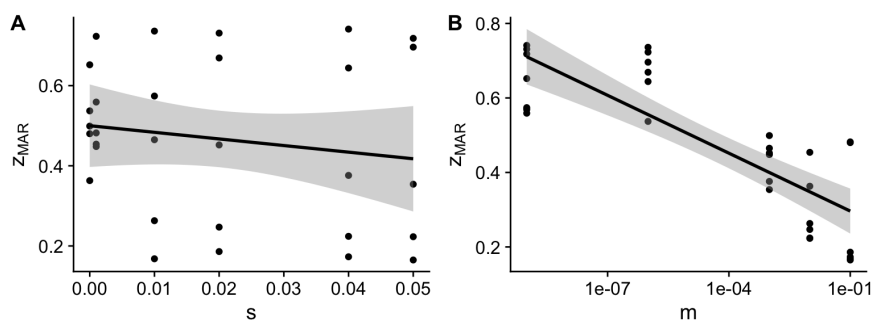
556

### 557 II.3.4 Metapopulations in space with local adaptation

558

559 In order to simulate local adaptation, we use the individual-based simulation software SLiM  
560 (24) following the approach of (25). These simulations were set up for 196 demes arranged in  
561 a 14 x 14 grid. Each grid cell contains a population of  $N=1000$  and has an environment  
562 attribute,  $e$ , which varied spatially from the lower-left to the upper-right corners (approx.  $-7 <$   
563  $e < 7$ ). 12 locations in the genome were allowed to be under directional natural selection. The  
564 selection coefficient was fixed for a simulation, and grid runs were conducted with  
565  $0 < s < 0.05$ , but this selection would vary based on the environmental selection value of a grid  
566 cell, according to  $e \times s$ . Therefore, these alleles are antagonistic pleiotropic. Selected  
567 mutations across the 12 loci in the genome behaved additively (e.g. if an individual in grid

568 cell  $i$  had two of the selected mutations, fitness would be  $w=1+2s \times e_i$ . The migration rate  
 569 varied from one individual in a billion ( $1 \times 10^{-9}$ ), to one individual every ten ( $1 \times 10^{-1}$ ). Finally,  
 570 the mutation rate was set to  $10^{-8}$  mutations/bp/generation and the recombination rate to  $10^{-7}$   
 571 crossovers/bp/generation.  
 572  
 573  
 574



575  
 576 **Fig. S7 | SLiM population genetic simulations in 2D with selection and local adaptation**  
 577 Simulations were carried out with different combinations of migration rates and strength of antagonistic  
 578 pleiotropic selection at 12 QTLs. (A) Marginal relationship between  $z_{MAR}$  with the strength of spatially-varying  
 579 selection  $s$ . (B) Marginal relationship between  $z_{MAR}$  with the migration rate  $m$ .

580  
 581 These results, together with individual-based simulations, corroborate what we had  
 582 observed with coalescent simulations, i.e. that  $z_{MAR}$  is lowest with a high migration rate. The  
 583 simulations also appear to show a negative effect of selection on  $z_{MAR}$ . Generating a linear  
 584 model fitting migration rate and selection and their interaction to understand what factors  
 585 explain the scaling coefficient:  $z_{MAR} \sim \log_{10}(m) + s + \log_{10}(m) s$ ; we confirm that both had a  
 586 significant effect, and that selection significantly reduces  $z_{MAR}$  (Fig. S7, see below summary  
 587 Table S2). This may seem counterintuitive, as one may expect that locally-adaptive mutations  
 588 are rare and will be localised only to where they are adaptive. More work is necessary to  
 589 understand the signatures that spatially-varying natural selection (and its different types)  
 590 create on  $z_{MAR}$ , but we can think that under migration limited scenarios (where  $z$  approaches  
 591 1) adaptive alleles and their linked mutations permeate faster to similar neighbour  
 592 environments than neutral alleles.  
 593

594 **Table S2 | Linear model explaining  $z_{MAR}$  by migration rate and natural selection**  
 595 Summary table of the linear model  $z_{MAR} \sim mig + s + mig:s$

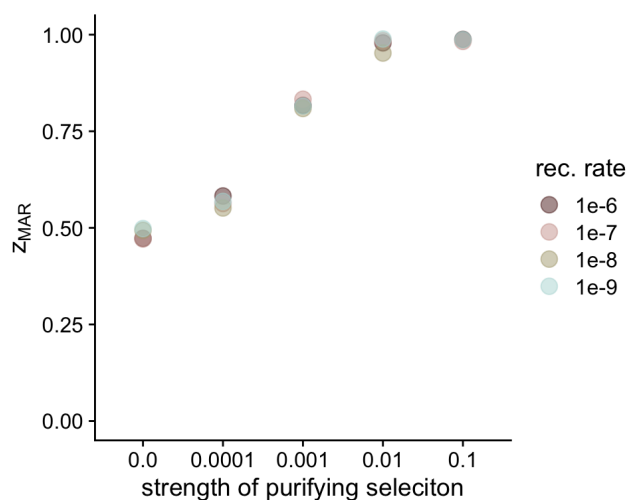
	Estimate	SE	t-value	P-value
intercept	0.3385022	0.0469174	7.214859	0.0000001
mig	-0.0419733	0.0085804	-4.891792	0.0000407
s	-4.693492	1.6290184	-2.881178	0.0076725
mig : s	-0.4998393	0.2426463	-2.059950	0.0491621

596  
 597  
 598 **II.3.5. Metapopulations in space with purifying selection**  
 599

600 To understand the effect of purifying selection on  $z_{MAR}$  we also ran 2D simulations with a  
 601 fraction of the genome allowed to be globally-deleterious (i.e. independent of the spatially-  
 602 varying environment). We simulated an increasingly strong purifying selection ( $|s|$  range  
 603 from 0.0 to 0.1), simulating roughly that 29% of the genome of Arabidopsis is coding  
 604 (arabidopsis.org) and mutations can be deleterious. We also varied the degree of  
 605 recombination. Following our expectation, with stronger purifying selection deleterious

606 mutations are pushed to lower allele frequencies, stopping their geographic spread, which  
607 increases  $z_{MAR}$ . Recombination rate appears to have a minor role on  $z_{MAR}$  (Fig. S8).

608  
609



610

611 **Fig. S8 | SLiM population genetic simulations in 2D with purifying selection**

612 Simulations were carried out with varying strengths of purifying selection ( $|s|$  range from 0.0 to 0.1) at coding positions,  
613 representing about 29% of the genome. Different values of recombination rate were also used in all pairwise combinations  
614 with  $|s|$ .

615

616 **II.3.6 Continuous-space non-Wright-Fisher models**

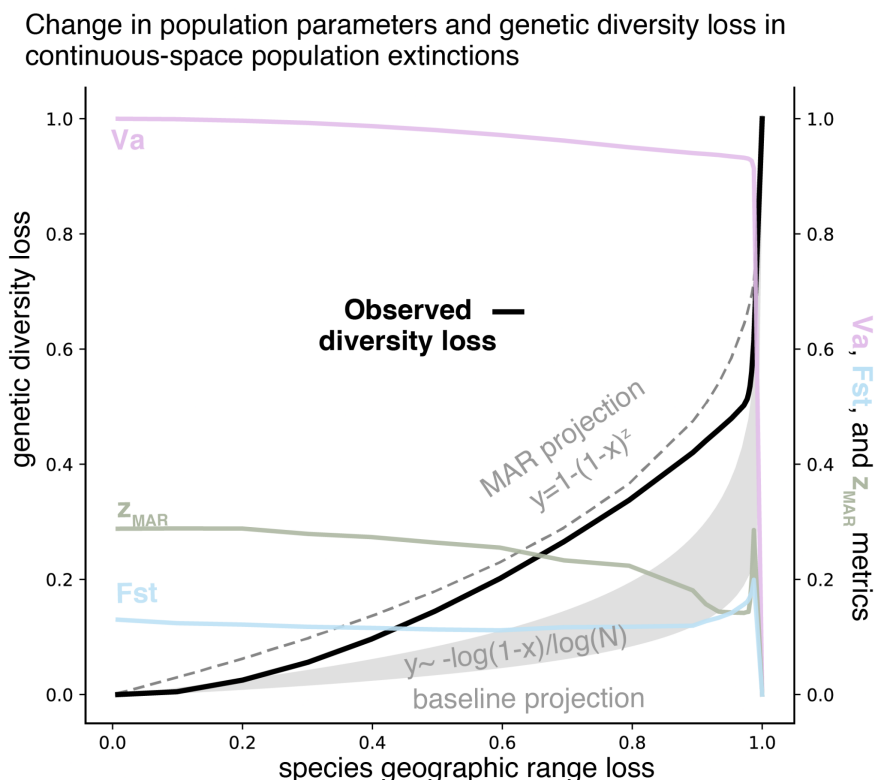
617

618 In order to confirm  $z_{MAR}$  generality in highly realistic conditions and its behavior through the  
619 population extinction process (II.4), we set up SLiM simulations in continuous space using  
620 non-Wright-Fisher dynamics (24). Spatial population structure in these simulations was  
621 established through individual dispersal, local mate choice and spatial competition, which we  
622 chose to lead to realistic values of  $F_{ST}$  across space. Spatial competition also acted as  
623 population control, by keeping the total population size below a target carrying capacity  
624 through direct effects on individual fitness. In addition to competition, fitness was also  
625 affected by individual age as well as by a polygenic trait under stabilising selection. A subset  
626 of variants (final proportion  $\sim 10\%$ ) directly affected this trait with effect sizes drawn from a  
627 Gaussian distribution with mean = 0.0 and standard deviation = 0.1, and a fitness penalty was  
628 incurred by deviating from the optimal trait value using a Gaussian fitness function centered  
629 at the optimum and with a standard deviation = 5.0. We initialised functional variation for  
630 SLiM using neutral coalescent simulations with msprime (23) to reduce the computational  
631 burden of burn-in, and loaded the resulting tree sequences into SLiM (26, 27). We drew  
632 functional effect sizes for these variants, placed individuals into continuous space, and ran  
633 simulations forward-in-time for 5,000 generations. After that, the geographic distribution of  
634 the species experienced impacts as expected during global change: every generation, 0.001 of  
635 one edge of the species distribution got its carrying capacity reduced to 0. This meant that  
636 over 1,000 generations the whole species would disappear (note that this is a reasonable  
637 fraction of area reduction given the estimates of yearly deforestation and habitat change in  
638 section V). We subsequently overlaid neutral mutations on the tree sequence using  
639 msprime, and analysed genomes sampled throughout the extinction process (by tracking them  
640 in the tree sequence output) and extracted using tskit.

641

642

643



644

645 **Fig. S9 | Continuous space SLiM population genetic simulations**

646 *At 19 timepoints leading up to extinction, 1,000 individuals were sampled randomly in continuous space to quantify diversity*  
 647 *loss (black line). The prediction of MAR (dashed line) using the starting  $z_{MAR}$  seemed to follow the real trend better than the*  
 648 *baseline of just loss of individuals (dashed line). This suggests that even if  $z_{MAR}$  varies during the population extinction*  
 649 *process, it is relevant to understand genetic loss by area reduction. We also tracked metrics of population structure ( $z_{MAR}$ ,*  
 650  *$F_{ST}$ ) and a proxy of adaptive capacity ( $V_a$ ), which showed qualitatively similar patterns as the GWA-based trends (Fig S21).*  
 651

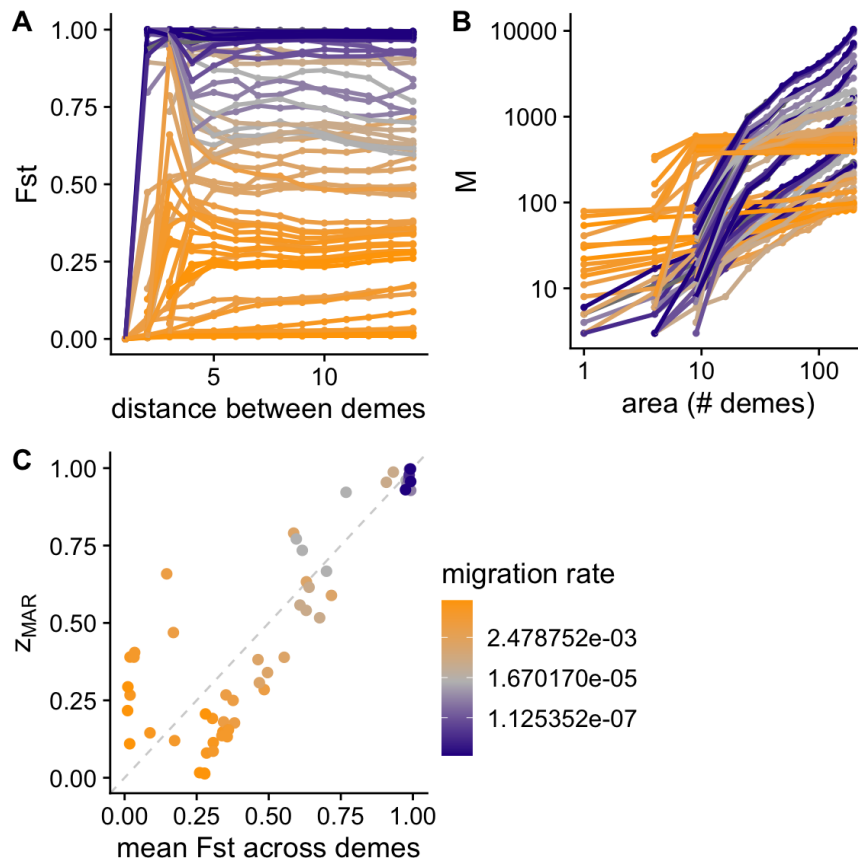
652

653 **II.3.7 Connection of  $z_{MAR}$  with the isolation-by-distance pattern**

654

655 Ultimately,  $z_{MAR}$  is a complex integrator of evolutionary forces acting in space (mutation,  
 656 migration, drift, selection) and captures how structured the distribution of a species'  
 657 mutations is. Although the isolation-by-distance pattern conceptually resembles  $z_{MAR}$ , we have  
 658 found no obvious analytical expression that relates both. Note that  $F_{ST}$  is defined based on  
 659 heterozygosity or  $\pi$ , instead of the number of segregating sites (i.e., mutations  $M$ ). For  
 660 instance, using Hudson's estimator (28) to compute  $F_{ST}$  across a set of populations we  
 661 calculate  $F_{ST} = 1 - (\pi_w / \pi_b)$ , where  $\pi_w$  is the diversity or heterozygosity within a population  
 662 and  $\pi_b$  is the same parameter calculated for the meta-population. Plotting  $F_{ST}$  of a  
 663 metapopulation by the distance of the farthest demes shows the typical non-linear trend of  
 664 isolation-by-distance, which shows that very close populations have similar allele frequencies  
 665 whereas populations further away drift apart. A challenge of  $F_{ST}$  is that it requires pre-  
 666 defining discrete populations, which is straightforward in stepping-stone simulations but hard  
 667 in real data. Comparing average  $F_{ST}$  of our 14x14 spatial demes and  $z_{MAR}$ , we see that the two  
 668 parameters correlate (Fig. S10C). However, it appears that for low values of  $F_{ST}$ ,  $z_{MAR}$  captures  
 669 more variation across the simulations (Fig. S10). These patterns were also confirmed in  
 670 continuous space simulations (not shown).  
 671

671



672

673 **Fig. S10 | SLiM population genetic simulations in 2D comparing  $F_{ST}$  and  $z_{MAR}$**

674 Neutral SLiM simulations with different degrees of migration. (A) Hudson's  $F_{ST}$  across populations with different area  
 675 subsamples. Following the expectation of the isolation-by-distance pattern, as the distance between the farthest demes in the  
 676 subsample increases,  $F_{ST}$  becomes larger and saturates at large distances. (B) The mutations-area relationship. (C)  
 677 Comparison between  $F_{ST}$  and  $z_{MAR}$ .

678

679

## 680 II.4 The loss of mutations (genetic diversity) in space

681

682 The aim is to predict the fraction of genetic diversity loss,  $x_M$ , from shrinking of an ecosystem  
 683 by an area  $a$ . To define all terms, we then have a past area  $A_{t-1}$  and a present reduced area  
 684  $A_t = A_{t-1} - a$ , and a fraction of area extinct  $x = a/A_{t-1}$

685

686 We first think of the loss of genetic diversity  $x_M$  through the basic process of losing  
 687 individuals. From the population genetics's coalescent theory derivation of the number of  
 688 mutations or segregating sites from individuals we got the approximation  $M \sim \log(N)$ .

689 Assuming the loss of area is simply the loss of individuals ( $A=N$ ), we can derive the fraction  
 690 of genetic diversity loss as:

691

$$\begin{aligned}
 x_M &= 1 - \frac{M_t}{M_{t-1}} = 1 - \frac{\log(N_t)}{\log(N_{t-1})} = 1 - \frac{\log(N_{t-1}(1-x))}{\log(N_{t-1})} = 1 - \frac{\log(N_{t-1}) + \log(1-x)}{\log(N_{t-1})} \\
 &= -\frac{\log(1-x)}{\log(N_{t-1})}
 \end{aligned}$$

693

694

695 The loss of mutations is then in the scale of:  $\log(1-x)$ ; which is very slow, as we  
 696 expected from having derived the trend that under panmixia  $z_{MAR} \approx 0$ . A substantial loss of  
 697 genetic diversity in this case only happens when population extinction is almost complete.

698  
699  
700  
701  
702  
703

Species do not typically behave perfectly panmictic given different  $z_{MAR}$  values. Under population structure, we can use our relationship to project the number of mutations (genetic diversity) lost as the geographic distribution due to habitat loss or climate change following equation:

$$x_M = 1 - \frac{M_t}{M_{t-1}} = 1 - \frac{MAR(A_t)}{MAR(A_{t-1})} = 1 - \frac{A_t^z}{A_{t-1}^z} = 1 - \left(\frac{A_t}{A_{t-1}}\right)^z = 1 - (1-x)^z$$

704  
705  
706  
707  
708

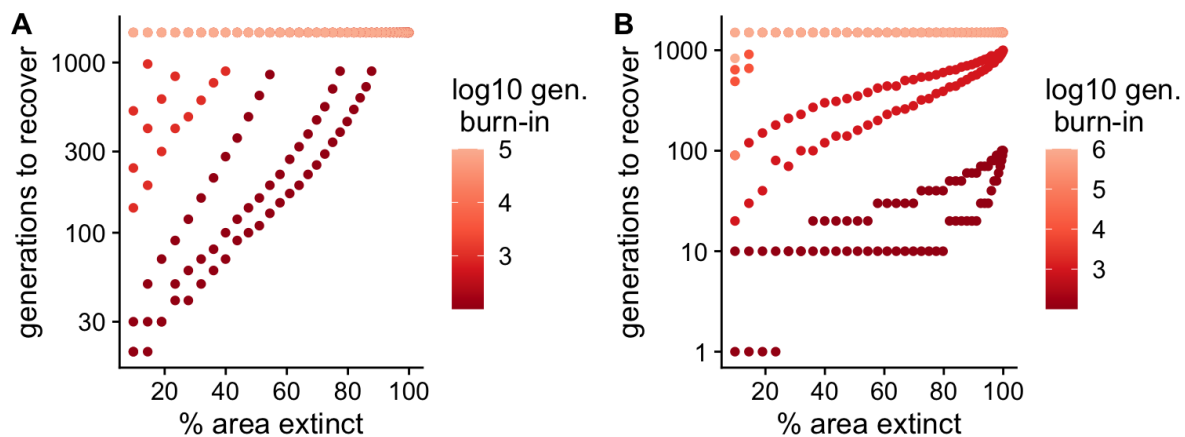
In the most extreme scenario of  $z_{MAR} \approx 1$ , the fraction loss of geographic area directly translates to the same fraction loss of genetic diversity.

709  
710  
711  
712

Reality should be in between the panmictic and fully-migration-limited cases. With combinations of environmental selection, non-equilibrium demography, and long-range dispersal, we may get intermediate  $z_{MAR}$  values, and it will be empirical estimates that can inform us how much may be lost (Section III).

713  
714  
715

## II.5 Recovery of genetic diversity after a bottleneck or local extinction



716

717 **Fig. S11 | 2D stepping-stone msprime simulations with extinction and recovery**

718 (A) Recovery of genetic diversity (number mutations) after loss of a fraction of the population. (B) Recovery of genetic  
719 diversity after instantaneous loss of a fraction of the population and consecutive repopulation.

720 \*Simulations with number of generations until recovery that are exceedingly large are assigned a value of 1,500, as none  
721 are realistic for current conservation timelines.

722  
723

724 The intuition that rapid recovery of genetic diversity may be possible is likely flawed.  
725 While genetic recovery may be faster than speciation rates, which are on the order of millions  
726 of years, the time for a set of populations that went through a simulation burn-in of 1,000  
727 generations (not yet in diversity equilibrium), and that suffer an instantaneous 5% reduction  
728 of area and an instantaneous recovery (e.g., through reforestation) would range from 20-90  
729 generations. This number of generations for long-lived species would translate into centuries  
730 or millennia of recovery without further impacts. About 49% of simulations – including every  
731 simulation that reached equilibrium (burn-in generations >10,000) – have a recovery time of  
732 more than a thousand generations (Fig. S11).

733  
734

## 735 SUPPLEMENTAL RESULTS

736

### 737 III. The mutations-area relationship with the 1001 Arabidopsis Genomes

738

739 We begin testing the idea of a general mutations-area relationship using the extensive  
740 sampling of the model plant species *Arabidopsis thaliana* and the 1001 Arabidopsis Genomes  
741 Project (29). This section will serve as a case study to explore different approaches and biases  
742 when building MAR to then apply the learned lessons across species (section IV).

743

744

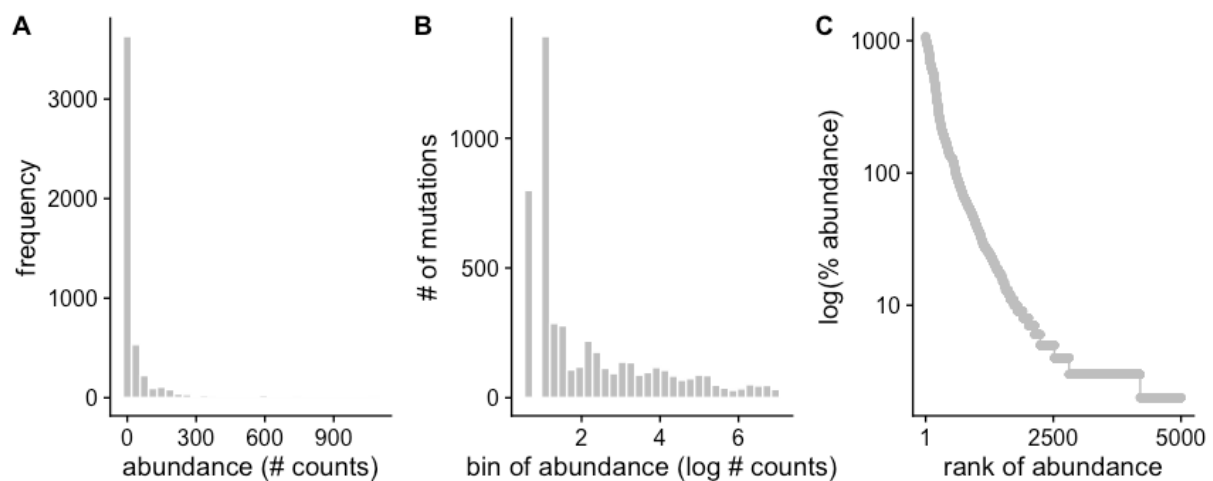
#### 745 III.1 The Site Frequency Spectrum of the 1001 Arabidopsis Genomes

746

747 We began analyzing the frequency distribution of 11,769,920 biallelic genetic variants (i.e.,  
748 mutations), which is typically called the Site Frequency Spectrum (SFS) in population  
749 genetics.

750

751



752

#### 753 **Fig. S12 | Mutation abundance study in *A. thaliana***

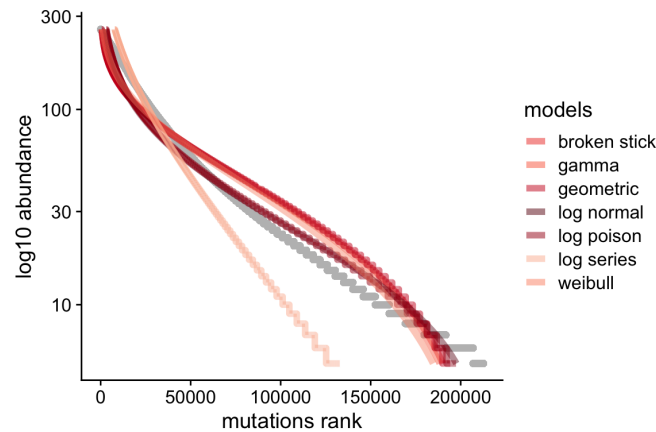
754 (A) Site Frequency Spectrum (SFS). (B) Preston plot of mutation abundances. (C) Whittaker plot of mutation rank  
755 abundances.

756

757 To showcase the similarities to the Species Abundance Distributions (SAD), we use  
758 the Whittaker plot of mutation rank abundance (Fig. S12) that suggests a log-normal of S-  
759 shape may be the best fitting model (Table S3). For a review listing many popular models,  
760 see (30), and for implementation details of 13 SAD models see the thorough manual of R  
761 package SADS (31). As we shall see later, the log-normal distribution seems to be the best fit  
762 across species.

763

764



765

766 **Fig. S13 | Fit of mutation abundance study in *A. thaliana* with different SAD models**

767 Representative models from Table S3 are plotted along with the observed frequency of 11,769,920 mutations.

768

769 Although model AIC captures best the fit of a curve accounting for the difference in  
770 parameter complexity of each model and the statistical distributions behind, we often are  
771 interested in the variance explained. We then calculated a proxy of predictive accuracy using  
772 a pseudo- $R^2$  approach of the difference between the model fit and the observed data as:

773  $R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$ . For *A. thaliana*, we used 10,000 SNPs sampled at random to an accuracy of  
774 over  $R^2 > 0.999$  for both the top log-Normal model and the bottom log-Series model,  
775 indicating that all “commonness of rarity” models must have a pretty good fit of mutation  
776 frequency data.

777

778 **Table S3 | AIC values for model fit of common species distribution curves.**

779 For each SAD model, the degrees of freedom and the delta AIC compared to the top model are reported.

780

Model	dAIC	df
log-Normal	0	2
Poisson	7204.37509	2
Geometric	44267.5475	1
Weibull	45872.3678	2
Gamma	48805.6065	2
Broken Stick	49076.4368	0
UNTB (MTZSM)	168434.181	1
log-Series	168434.726	1

781

782

783 The typical SFS from population genetics is of course not implemented in current packages for Species  
784 Abundance Distributions like R sads. For comparison, in the main text we also calculate the log  
785 likelihood and AIC of this following the standard population genetics likelihood:

786

$$\log L = \sum_i \log\left(\frac{1}{Nq_i}\right) - \log(H_n(N-1))$$

787

788

789 where  $N$  represents the number of individuals in a sample, and  $q_i$  is the minor allele frequency of a  
790 SNP in the sample, in the main text calculated for  $i=1 \dots 10000$  random SNPs (see main text). As  
791 before,  $H_n$  is the harmonic number function.

792

793



794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815

### III.2 Building the Mutations-Area Relationship

In the following, we explain how the area was estimated that was used to compute  $z_{MAR}$  on real world data. In short, we used a grid on the world map, with samples placed on the map based on their geo-coordinates of origin (Fig. 1). We first create square spatial subsamples of the *Arabidopsis thaliana* geographic distribution (Fig. 1, Fig. S15) and quantify diversity  $M$  as the total segregating sites. Excluding zeros, these two variables are fed to the `sars_power` function from the R SARS package (32).

Although the power law mutations-area relationship was already theoretically motivated (II.3), here we also fit different types of functions typically applied to the Species-Area Relationship. Doing this, we reach the conclusion that multiple models perform very similarly, and the classic power law is among the top models, see Table S4. Although small marginal fitting accuracy could be achieved with other models, for mathematical convenience and historical continuity, we use the power law for later sections and the study of MAR across species (Sections IV and V).

**Table S4 | Different SAR curves fit to mutations.**

We fit 20 different functions and calculated the variance explained (R<sup>2</sup>), Pearson's r, and Spearman's rho.

Model	R <sup>2</sup>	r	rho
Asymptotic regression	0.21825683	0.46717965	0.53510077
Beta-P cumulative	0.22012799	0.46917799	0.53374757
Chapman Richards	0	NA	NA
Cumulative Weibull 3 par.	0.21929646	0.468291	0.53374757
Cumulative Weibull 4 par.	0.21930145	0.46829633	0.53374757
Extended Power model 1	0.21833611	0.46726449	0.53026812
Extended Power model 2	0.21682584	0.46564561	0.53462775
Gompertz	0.16393078	0.40488366	0.45964364
Heleg(Logistic)	0.21929721	0.4682918	0.53531975
Kobayashi	0.22228406	0.47147011	0.53526975
Linear model	0.19579007	0.44248171	0.53510077
Logarithmic	0.20280401	0.45033767	0.53430311
Logistic(Standard)	0.22536996	0.47473146	0.53549765
Monod	0.22500999	0.47435217	0.53579276
Negative exponential	0.22801633	0.47751055	0.53447179
Persistence function 1	0.21929612	0.46829063	0.53501182
Persistence function 2	0.21760028	0.46647645	0.53409266
<b>Power</b>	<b>0.21929556</b>	<b>0.46829004</b>	<b>0.53543785</b>
PowerR	0.21753225	0.46640353	0.53493321
Rational function	0.22072491	0.46981369	0.53451874

816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826

Because in the species literature it is recommended to only quantify richness of endemic species (33), we also count segregating sites that are private to the area subsample, creating the equivalent endemic-mutations-area relationship (EMAR) (33). The MAR slope and 95% Confidence Interval was  $z = 0.324$  (0.238 - 0.41) (Table S5, Fig. S14 A), while the EMAR was  $z = 1.241$  (1.208 - 1.274) (Table S6, Fig. S14 B). Interestingly, the endemics-area relationship of  $z \approx 1$  resembles that of endemic species, whereas the total mutation relationship with area is above that of species relationships, which typically follows the canonical  $z \approx 0.2 - 0.4$ .

827 We must note that EMAR, the genetic analogy of the Endemic-(species)-Area  
 828 Relationship (EAR) may not be that meaningful when analyzing genomic data (we did not  
 829 find a way to theoretically motivate it in section II), and later we see it overestimates loss in  
 830 our simulations (Fig. S18)

831

832 **Table S5 | The mutations-area relationship (MAR).**

833 Fitted values in a log-log power function between area sampled and mutations discovered.

	Estimate	Std. Error	t value	P	2.5%	97.5%	nls.Est.	nls.2.5%	nls.97.5%
c	494.565432	135.6314588	3.646392	0.0003138	223.3025141	765.8283493	494.5531270	278.1107276	822.829918
z	0.323727	0.0430277	7.523681	0.0000000	0.2376715	0.4097824	0.3237367	0.2430303	0.413162

834

835

836 **Table S6 | The endemic-mutations-area relationship (EMAR).**

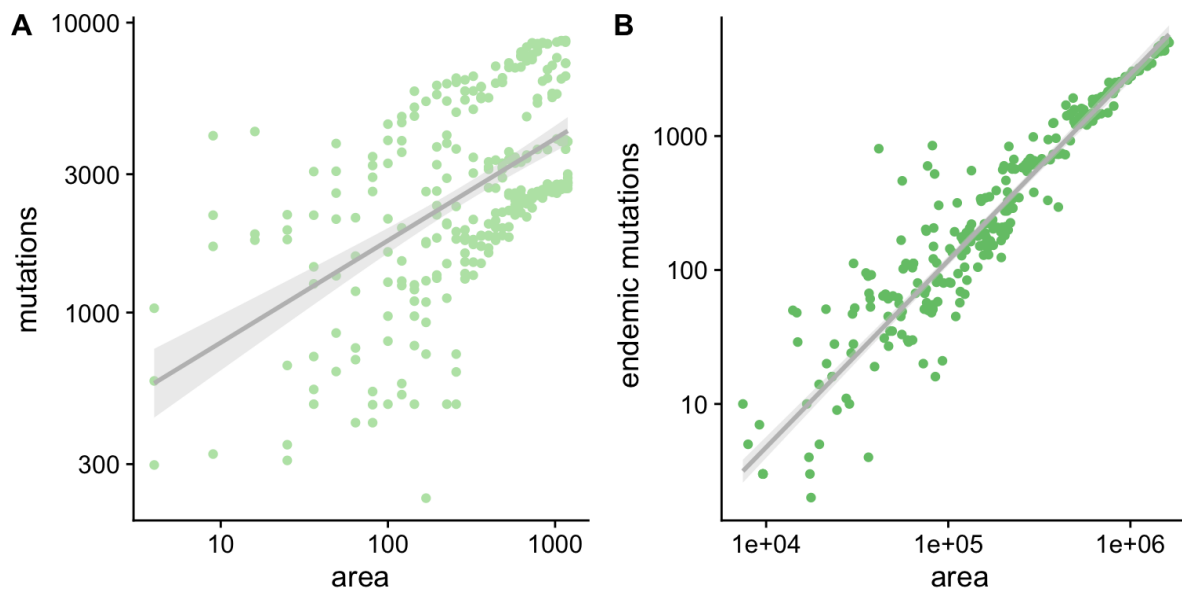
837 Fitted values in a log-log power function between area sampled and endemic mutations discovered.

	Estimate	Std. Error	t value	P	2.5%	97.5%	nls.Est.	nls.2.5%	nls.97.5%
c	0.0001001	0.0000231	4.337758	1.98e-05	0.0000539	0.0001463	0.0001001	0.0000635	0.0001555
z	1.2411831	0.0165268	75.101442	0.00e+00	1.2081296	1.2742366	1.2412125	1.2096087	1.2737927

838

839

840



841

842 **Fig. S14 | The mutations-area and endemic-mutations-area relationships in *A. thaliana*.**

843 Dividing *A. thaliana* native geographic distribution into a 1 degree lat/long grid, square areas with 1 degree side-length to  
 844 36 degrees side-length were randomly placed ( $n=100$  for each size) across the distribution, and genetic diversity metrics  
 845 were computed to produce the (A) Mutations-Area Relationship and (B) Endemic-Mutations Area relationship.

846

847

### 848 III.3 Testing for potential numerical artefacts

849

850 We wondered whether MAR estimates may be affected by some numerical artefacts in our  
 851 software pipeline (available at <https://github.com/moiepositoalonsolab/mar>). For instance,  
 852 real world data may have uneven sampling in space, the spatial resolution of georeferenced  
 853 samples may vary, projection of samples into gridded maps may have limited resolution,  
 854 software pipelines may produce biased estimates, etc. To test this, we conducted several  
 855 experiments:

856

857 **Lower bound of the method for  $z_{MAR}$ .** Our first experiment when building the MAR  
 858 aimed to make sure that spatial sampling, or some unknown bias in genome sequencing, or  
 859 the number of samples used, are not creating artificially large  $z_{MAR}$ . We then simulated a mock  
 860 dataset of *A. thaliana* with the same number of mutations, samples, and using the original  
 861 geographic locations. The number of SNPs were also sampled in a way that we created a  
 862 canonical  $1/q$  SFS for the whole species. Under no biases, we then expect the MAR to follow  
 863 the theoretical derivation under panmixia with a  $z \sim 0$ . This exercise confirmed we get a value  
 864 approaching zero:  $z = 0.033$ ,  $(-0.095 - 0.162)$ .  
 865

866 **Table S7 | MAR built with different area calculations and grid sizes**  
 867

Grid resolution (deg.)	$z_{MAR}$ [CI95%] (cell area)	$z_{MAR}$ [CI95%] (total area)
A=N	0.431 (0.423 - 0.439)	NA
0.1	0.435 (0.424 - 0.446)	0.367 (0.281 - 0.454)
0.25	0.454 (0.449 - 0.459)	0.422 (0.376 - 0.467)
0.5	0.488 (0.465 - 0.511)	0.352 (0.152 - 0.551)
1	0.543 (0.529 - 0.558)	0.389 (0.295 - 0.483)
2.5	0.644 (0.6 - 0.688)	0.388 (0.251 - 0.526)
5	0.617 (0.205 - 1.029)	0.403 (-0.204 - 1.011)

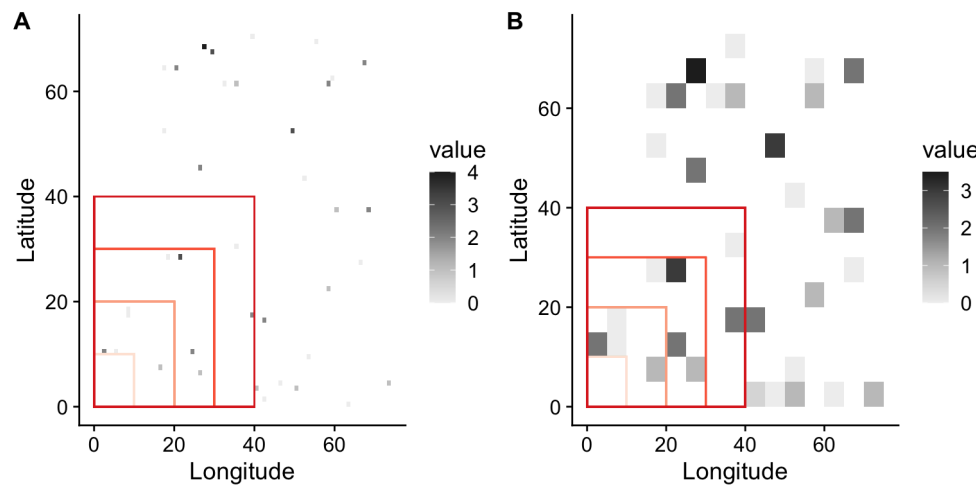
868  
 869  
 870  
 871 **Grid sizes, area calculations, and non-random spatial sampling.** In order to  
 872 streamline geospatial operations, we implemented the MAR relationship calculations in this  
 873 project using R raster objects (34). This required projecting the collected samples of a species  
 874 and the observations of any given mutation into a world map (i.e., each mutation's geographic  
 875 distribution). Necessarily, in order to be able to assign areas to sets of samples or mutations  
 876 on the map, the projection requires the choice of a grid size. The larger the grid size (e.g.,  
 877 lower spatial resolution), the faster the spatial operations can be performed. Further, for larger  
 878 grid sizes, we expect the slope of MAR to be more influenced by larger-scale patterns, while  
 879 for smaller grid sizes, the MAR will be influenced by smaller-scale patterns. To test this, we  
 880 repeated the subsampling of *A. thaliana* distribution with grid sizes ranging 0.1 degrees  
 881 latitude/longitude (roughly 10km side-length in temperate regions) to 10 degrees (roughly  
 882 1,000 km side-length). The estimates were roughly consistent between 0.4-0.6, but increases  
 883 in value at larger grid sizes (row in Table S7 for large grid size values), a scale-dependent  
 884 pattern that resembles results of SAR of species in ecosystems fitted at different scales (10).  
 885

886 Because we often have sparse samples of individuals in space, we devised two  
 887 strategies to calculate areas during the subsampling of MAR (see cartoon in Fig. S15): (A)  
 888 the total square area of the minimum and maximum latitude/longitude values of all the  
 889 samples analyzed. That is, simply the area of the red box in the figure. (B) the sum of areas of  
 890 grid cells that contain at least one sample. That is, the sum of the grey squares within the red  
 891 box in the figure. In addition, we also calculated the MAR relationship assuming the total  
 892 area is equal to the number of individuals ( $A=N$ ) (which should be theoretically equivalent to  
 893 a grid of very high resolution where we end up with a maximum of one individual sampled at  
 894 any grid cell).  
 895

896 Table S7 values suggest there is a dependency of  $z_{MAR}$  with the grid size when areas  
 897 are calculated as the sum of grid cells with at least one sample. Our intuition for this pattern  
 898 is that lower resolution grids (e.g., 5 degrees side) lead to some grid cells having many  
 899 samples, which would increase the number of mutations discovered when discovering the

900 area. On the other hand, the calculation of  $z_{MAR}$  using the total area does not seem to affect the  
901  $z_{MAR}$  estimate; however, because large areas often do not have samples (limiting the potential  
902 to find new mutations), it creates a higher variance in the estimate of  $z_{MAR}$  (see confidence  
903 intervals in Table S7 and Fig. S16). Here, we favored consistency of  $z$  at the expense of  
904 broader, more conservative confidence intervals. All the estimates reported below and in the  
905 main text therefore use the total area approach.

906  
907  
908

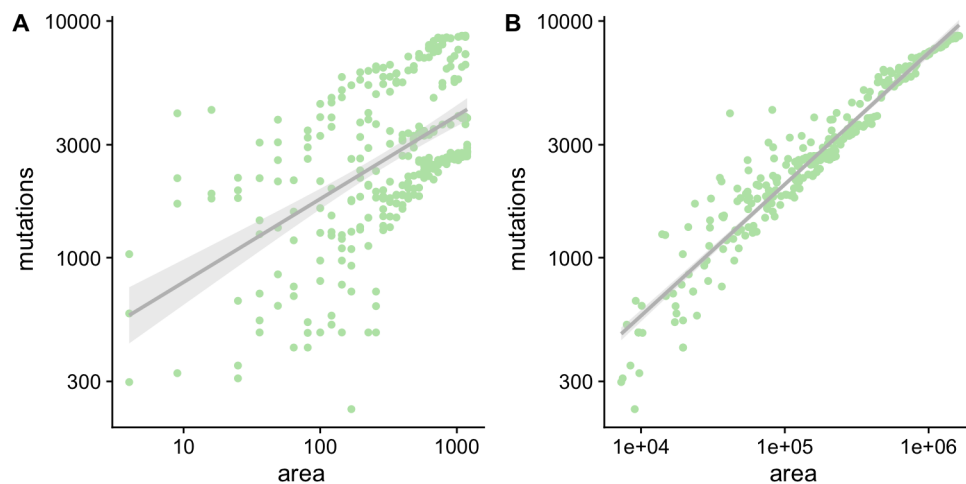


909

910 **Fig. S15 | Cartoon of raster sampling to build the MAR**

911 *Map of mock samples of a species projected into a raster. Grey scale indicates the number of samples per grid cell. Red*  
912 *boxes exemplify the process of spatial subsampling of increasing area to build the MAR relationship. Two example grid sizes*  
913 *were created for illustrative purposes: (A) Small grid size or high spatial resolution. (B) Large grid size or low spatial*  
914 *resolution.*

915  
916



917

918 **Fig. S16 | MAR comparison with different area calculations.**

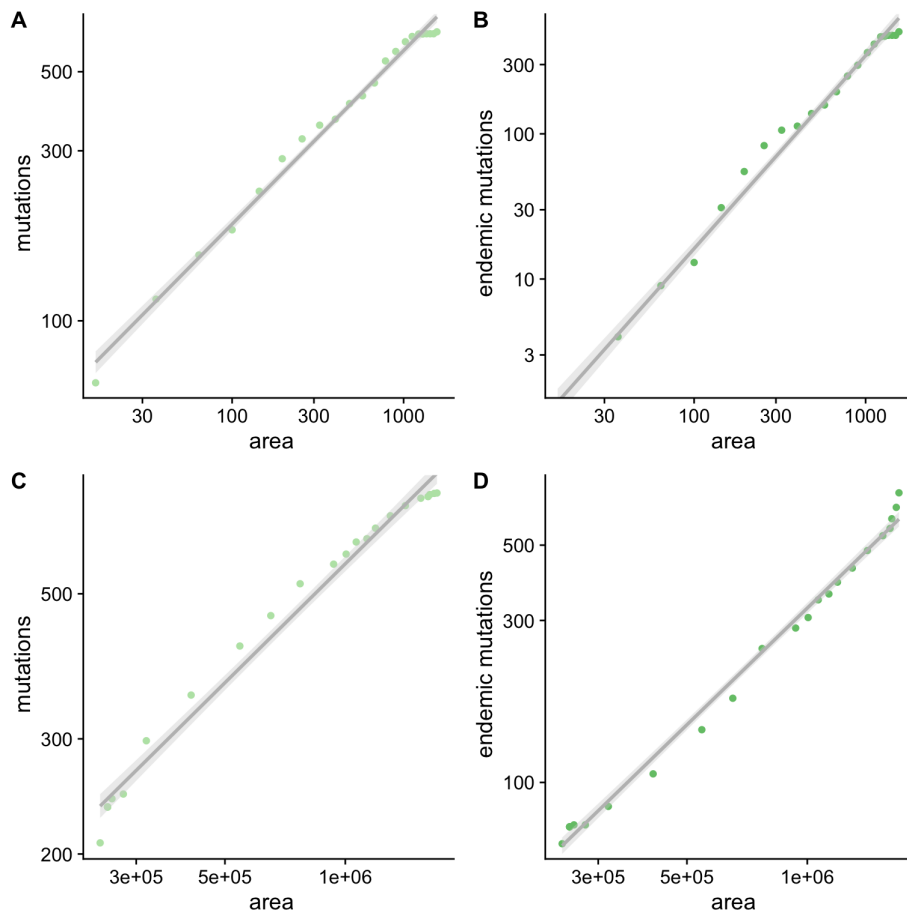
919 *(A) Using total area, (B) using grid cell sum with at least one sample. For 1 degree latitude/longitude grid cell.*

920

921 **Geographic subsampling strategy (inwards, outwards, random).** It has been  
922 indicated that the way the Species-Area Relationship (SAR) and Endemics-Area Relationship  
923 (EAR) are created may create differences in the scaling parameter  $z$ . The plots and estimates  
924 above were produced by randomly placing boxes of different size or area across the  
925 distribution of the species. Often, however, either discovery of species or extinction happen

926 in certain patterns. For instance, we often imagine sampling an ecosystem concentrically  
 927 outwards from a focal point, whereas we may think of the extinction process of species  
 928 reductions being concentrically inwards (33). Because these patterns seem of importance, we  
 929 also calculated the MAR and EMAR outwards from the latitude and longitude median of all  
 930 the samples in the map, moving outwardly until the map is filled (Fig. S17, Table S8).  
 931 Likewise, the inward pattern is conducted in an inverse manner.

932  
 933  
 934



935

936 **Fig. S17 | MAR and EMAR in *Arabidopsis thaliana* using outward and inward sampling.**

937 *Dividing *A. thaliana* native distribution in 1 degree lat/long grid, a square area of 1 degree was placed at the median of the*  
 938 *sampling range and was expanded iteratively by 1 degree lat/long until all the area of the distribution was covered. (A-B)*  
 939 *MAR and EMAR using a typical outward sampling. (C-D) MAR and EMAR using an inward sampling. The latter may not be*  
 940 *a common process of sample collection, but it is common for extinction progress.*

941

942 **Table S8 | Outward and inward MAR and EMAR**

943 *The MAR and EMAR relationship computed with inward or outward nested subsampling, calculating area only as those*  
 944 *cells with samples.*

945

Relationship	$z$
MAR outwards	0.444 (0.412 - 0.476)
EMAR outwards	1.086 (0.982 - 1.189)
MAR inwards	0.561 (0.524 - 0.597)
EMAR inwards	1.295 (1.192 - 1.399)

946

947

948 **Incomplete sampling of the species.** To check whether the relationship holds with  
949 few individuals of a species or limited geographic distributions, we compared the species-  
950 wide MAR with that of subset populations. Downsampling the native distribution of *A.*  
951 *thaliana* to a region within North-East Spain (-2.00–4.25 degrees East, 36.52–42.97 degrees  
952 North), or to a region within Germany (2.69–13.73 degrees East, 50.0–52.0 degrees North),  
953 and using only 1,000 SNPs, we recovered  $z_{MAR} = 0.423(0.233-0.614)$  for Spain and  
954  $0.525(0.242-0.807)$  for Germany, which were close to the estimate based on the whole  
955 distribution (Table 1). This result is reassuring in that if migratory patterns are relatively  
956 homogeneous, one may be able to estimate this parameter from a subset of the species  
957 distribution. For heterogeneous population structure cases, we expect incomplete sampling to  
958 produce unreliable estimates.

959  
960 **Number of genome-wide SNPs used.** To check whether different numbers of SNPs  
961 used for the analyses would lead to different  $z_{MAR}$ , we conducted analyses with random  
962 subsets consisting of 100, 1,000, and 10,000 SNPs, replicated 3 times. Estimates had a  
963 coefficient of variation of 4.7%, which is way below the standard error of typical estimates  
964 (Table 1).

965  
966 **Locally-adaptive variants.** We then aimed to understand the effect of utilizing SNPs  
967 that appear to be related to adaptation. To study this, we utilized an outdoor climate-  
968 manipulated experiment that recorded fitness data (survivorship and reproduction output of  
969 seeds) for 515 *Arabidopsis thaliana* ecotypes part of the 1001 Genomes set in 8 environments  
970 (Exposito-Alonso, 2019). We devised two sets of alleles: 10,000 that were negatively  
971 correlated with fitness in a Genome-Wide Association across 8 different environments, and  
972 10,000 alleles that were associated positively with fitness in one environment but negatively  
973 in another (antagonistic pleiotropic). The MAR relationship was computed as before and  
974 compared to the original random (putatively neutral) set of alleles from the previous sections  
975 (Table S9). Although we see a trend that locally-adaptive alleles have a slightly higher  $z$ ,  
976 estimates overlap. The effects seen here of having smaller  $z$  for adaptive alleles than neutral  
977 variation could, however, be due to top GWA SNPs often being ascertained to higher  
978 frequency than background SNPs.

979  
980

981 Table S9 | MAR for putatively neutral, deleterious, and locally adaptive alleles in *Arabidopsis thaliana*  
982

SNP set	$z$
neutral	0.324 (0.238 - 0.41)
globally deleterious	0.209 (0.13 - 0.288)
locally adaptive	0.291 (0.217 - 0.365)
globally positive	0.234 (0.137 - 0.332)

983

984

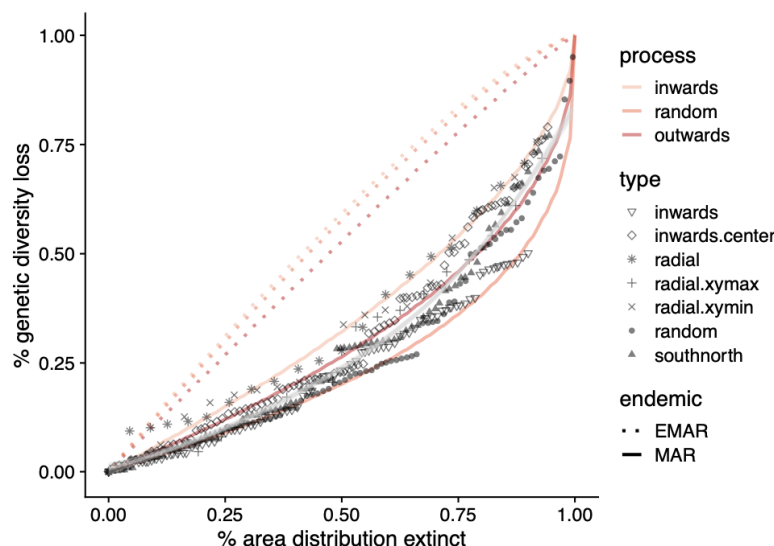
### 985 III.4 Local population extinction in *Arabidopsis*

986

987 Using the MAR framework, we can make projections of loss of mutations (or its inverse, the  
988 remaining genetic diversity. By doing this, the known intuition is that with  $z > 1$  (as from  
989 EMAR) the decrease of diversity is much faster than the decrease of habitat, but with  $z < 1$   
990 (as from MAR), there is a (desirable) slower dynamics of genetic loss. In the latter, despite  
991 habitats disappearing, reservoirs of mutations distributed across different locations enable  
992 conservation of certain variation. To study which one is more likely and to observe the

993 stochastic nature of genetic diversity loss, we simulated in silico population extinctions of  
 994 map cells from the Arabidopsis map (Fig. 1) and directly estimated from the genome matrix  
 995 of remaining individuals the remaining genetic diversity. These simulations were  
 996 implemented to capture different hypothesised patterns of extinction (see main text). All,  
 997 however, agree with the more hopeful estimate of  $z_{MAR} \approx 0.3$ .

998  
 999  
 1000



1001

1002 **Fig. S18 | Loss of mutations with habitat loss in *A. thaliana*.**

1003 Predictions based on MAR and EMAR functions and in silico extinction stochastic simulations in *A. thaliana*.

1004

1005 To study the fit of the genetic loss predictions based on MAR relationships and the  
 1006 results from computer simulations, we calculated a pseudo- $R^2$  based on the squared  
 1007 differences between the predicted line and the “observed” genetic loss as:  $R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$ .  
 1008 This results in a high fit  $R^2=0.872$  of the MAR, built from random samples of distribution  
 1009 areas, while the EMAR had a poor fit due to overestimation of genetic loss:  $R^2=-0.710$   
 1010 (negative values indicate predictions are worse than the mean of the data).

1011

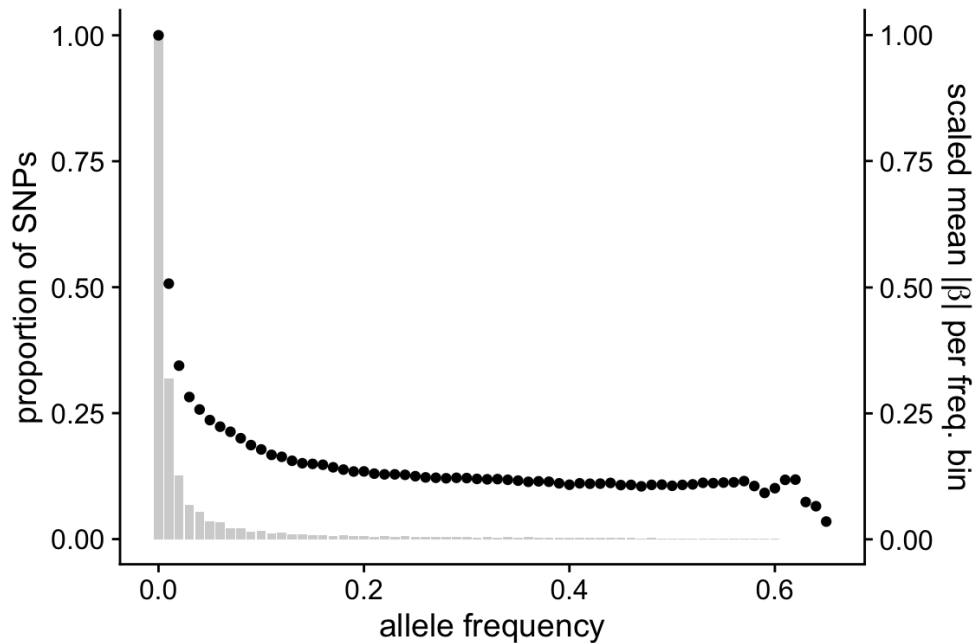
1012

### 1013 **III.5 Potential impacts of genetic loss in adaptability**

1014

1015 Although likely imperfect, Genome-Wide Associations could help to understand the  
 1016 relevance of mutations in different frequency classes in model organisms such as *Arabidopsis*  
 1017 *thaliana*. Fig. S19 shows the site frequency spectrum and a metric of the "total accumulated  
 1018 effect in fitness" of the alleles in every bin. Effect sizes were retrieved from GWA on lifetime  
 1019 fitness of 515 ecotypes in outdoor experiments (35). The average effect size across 8 fitness  
 1020 GWA from 8 experimental combinations were used: high/low precipitation, high/low latitude  
 1021 of outdoor stations, and high/low plant density. This exercise showcases the phenomenon that  
 1022 low frequency variants often have strong effect sizes, which is expected under a stabilising  
 1023 selection quantitative model (36). Because low frequency alleles will be the first to be lost  
 1024 during a bottleneck (as would happen with the rapid extinction of populations of a species),  
 1025 we may expect to lose variants that are related to fitness and thus potentially lose diversity  
 1026 that could be advantageous in some environments. Alternatively, deleterious mutations are  
 1027 also expected to be at low frequency, in which case would also make them more easily lost.

1028  
1029



1030

1031 **Fig. S19 | Bias of low frequency mutations and effect size for fitness traits in *A. thaliana*.**

1032 *Grey bars represent the site frequency spectrum (scaled for visualisation purposes). The black dots represent the mean*  
1033 *absolute effects of alleles as estimated from GWAs with 515 accessions scored for fitness traits in 8 outdoor experiments.*

1034

1035

1036

1037

1038

1039

1040

1041

1042

1043

1044

1045

1046

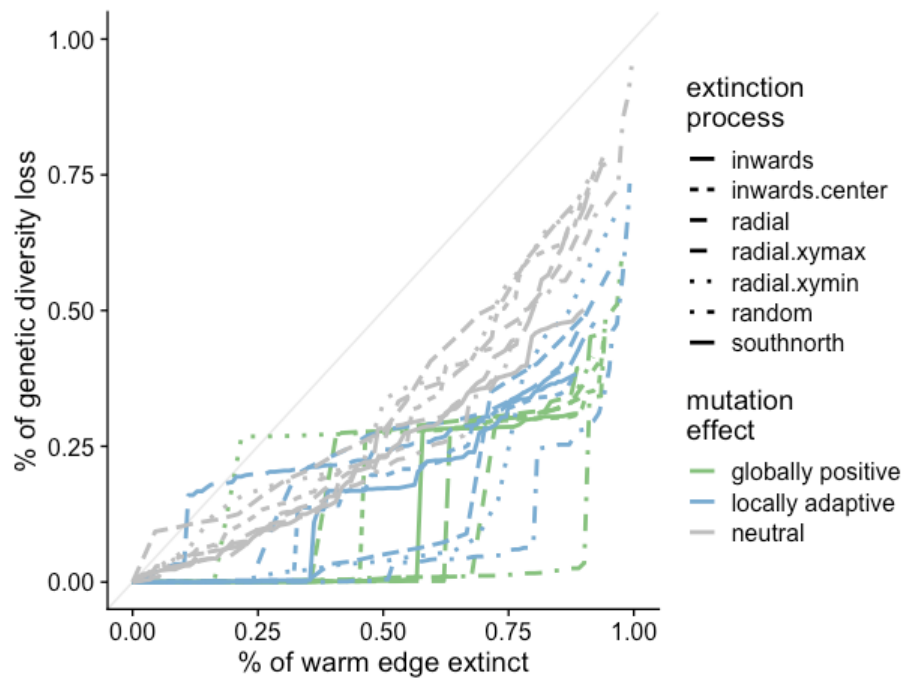
1047

1048

1049

To further build intuition on the progress of extinction in relation to loss of genetic diversity that is not neutral, we repeated warm edge extinction simulations with several subsets of alleles: randomly selected SNPs, SNPs that were associated positively in 2 environments (low precipitation Spain and high precipitation Germany) (labelled globally positive), and SNPs that were associated positively in one environment and negatively in the other (labelled antagonistic pleiotropic or putatively locally-adaptive). This (Fig. S20) supports our intuition that although putatively functional alleles (or alleles tightly linked to such functional ones) may have slower loss dynamics than neutral variants due to a high frequency and  $z_{MAR}$ , certain population extinction patterns may actually lead to rapid loss of potentially-adaptive genetic diversity. The complexity of these patterns, together with the evolutionary feedback created by lowering genetic standing variation that affects fitness, make the inference of adaptive capacity loss even more difficult than just inferring the loss of genetic diversity itself.





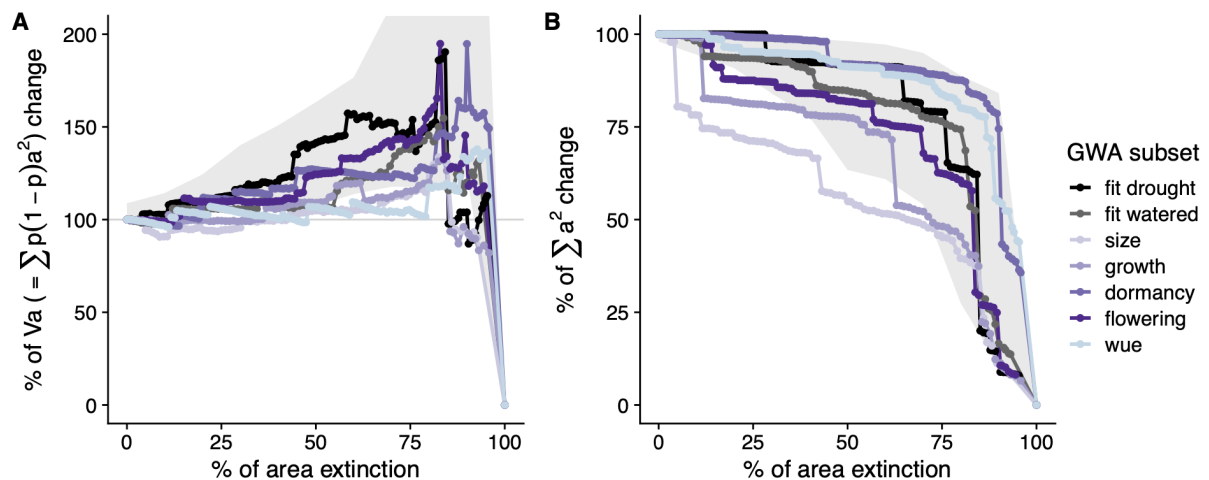
1050

1051 **Fig. S20 | Simulations illustrating the potential loss of locally-adaptive mutations in *A. thaliana*.**

1052 Simulations of extinction using multiple patterns of population losses with different subsets of alleles ascertained to show  
 1053 positive associations in fitness GWA in two outdoor experiments (green), positive associations in one environment (e.g. low  
 1054 precipitation) but negative in a second environment (e.g. high precipitation) or vice versa (green). These were compared to  
 1055 a random set (grey).

1056

1057



1058

1059 **Fig. S21 | Extinction simulations showing proxies of adaptive capacity of *A. thaliana*.**

1060 Using estimated allele effect sizes from 10,000 SNPs in the 1% P-value tails of several Genome-Wide Associations, we show  
 1061 (A) Percentage of change of  $V_a$  as a proxy of adaptive potential and (B) raw square sum of allele effects to showcase the  
 1062 inflating effect of intermediate frequency alleles. Grey background shape indicates the minimum and maximum boundaries  
 1063 of trajectories created by replicated frequency-matched non-effect sets of SNPs (one per GWA). The trajectories of some  
 1064 effect alleles appear to show faster loss than the non-effect background trajectories.

1065

1066

### 1067 III.6 Case study of a massive natural bottleneck

1068

1069 A recent colonisation of North America by *Arabidopsis thaliana* can help us understand the  
 1070 recovery of genetic variation. Whole-genome sequencing of 100 specimens of North  
 1071 American *A. thaliana* indicates that it migrated from its native range of Europe to North

1072 America in the 17th century, and began spreading across the continent from a genetically-  
1073 homogeneous population (37). Despite ideal conditions to re-gain genetic diversity—a  
1074 continental population expansion aided by human travel (38, 39)—only ~8,000 new  
1075 mutations were detected through spontaneous accumulation, equivalent to only ~0.067% of  
1076 the species-wide native genetic diversity. Because most of these mutations are at very low  
1077 frequency, as expected during population expansion, the scaling of genetic diversity with area  
1078 is approximately 1 ( $z_{MAR} = 1.025$  [CI95%: 0.878 - 1.173]).  
1079  
1080  
1081

#### 1082 **IV. The mutations-area relationship in diverse species**

1083

1084 Every dataset was retrieved online either from the published article in the form of VCF or  
1085 fastq files, or provided by the study authors upon request. All datasets were first transformed  
1086 into PLINK files using PLINK v1.9 (40). For computational efficiency, and since we showed  
1087 random subsampling does not appear to affect calculations of  $z_{MAR}$  (Section III.3), we  
1088 conducted all analyses with up to 10,000 randomly selected SNPs for each species sampled  
1089 genome-wide, or in the largest chromosome for those species with large genomes. We aim to  
1090 use mostly unfiltered SNP datasets to avoid ascertainment biased toward intermediate  
1091 frequency SNPs, and therefore we did not apply a MAF filter for any analyses. By default,  
1092 PLINK transforms SNP matrices into biallelic (if multiallelic, it takes the two most common  
1093 alleles). Although the preservation of structural genetic variation may also be relevant and  
1094 may have important consequences in adaptation (41), we do not expect dramatic differences  
1095 in their scaling relationship compared to biallelic SNPs, as their SFS are relatively similar  
1096 (Structural variants may show a skew to lower frequency, resulting in steeper  $z_{MAR}$ . By  
1097 excluding those, our analyses may be conservative). In order to properly characterise the  
1098 geographic distribution of a mutation using all available geo-tagged individuals, we filtered  
1099 for genotyping rate (plink --geno), and the final value is reported per dataset.

1100

1101 Details for dataset processing or homogenization are described below.

1102

- 1103 - The 1001 Arabidopsis Genomes Consortium (29) generated a WGS Illumina  
1104 sequencing dataset of *Arabidopsis thaliana* comprising 1,135 individuals and  
1105 11,769,920 SNPs. The VCF with the data is available at: <https://1001genomes.org>.  
1106 The raw sequencing data is available at  
1107 <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA273563>. These included recently  
1108 colonised regions such as North America or Japan. Analyses of  $z_{MAR}$  were calculated  
1109 only for the native range, which comprises most of the species diversity (>99%) and  
1110 1001 individuals. For computational efficiency, we conducted analyses using  
1111 randomly sampled SNPs from chromosome 1, as we did not observe any difference  
1112 when sampling from other chromosomes. A number of MAR approaches were tested  
1113 in this species (section III). For homogeneity, the final reported estimate (Table 1)  
1114 was conducted following the same procedures as other species with a random sample  
1115 of 10,000 SNPs.
- 1116
- 1117 - Lucek & Willi (42) recently published a dataset of WGS Illumina sequencing 108  
1118 *Arabidopsis lyrata* individuals from North America, which the authors directly shared  
1119 as a VCF. The raw data is available at  
1120 <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJEB30473>. We retrieved the  
1121 latitude/longitude data from the supplemental material. We applied a genotyping rate  
1122 filter ending with a dataset of 0.955431 genotyping rate. 10,000 SNPs were subsetted  
1123 at random from the genome-wide data.
- 1124
- 1125 - Kreiner et al. (43) WGS Illumina sequenced 165 individuals of *Amaranthus*  
1126 *tuberculatus*. The raw data is available in the link  
1127 <https://www.ebi.ac.uk/ena/browser/view/PRJEB31711>. The authors provided a VCF.  
1128 Overall, 155 individuals contained latitude and longitude information and were kept  
1129 for the analyses. The genotyping rate was 0.98162 and we subsetted randomly 10,000  
1130 SNPs.

1131

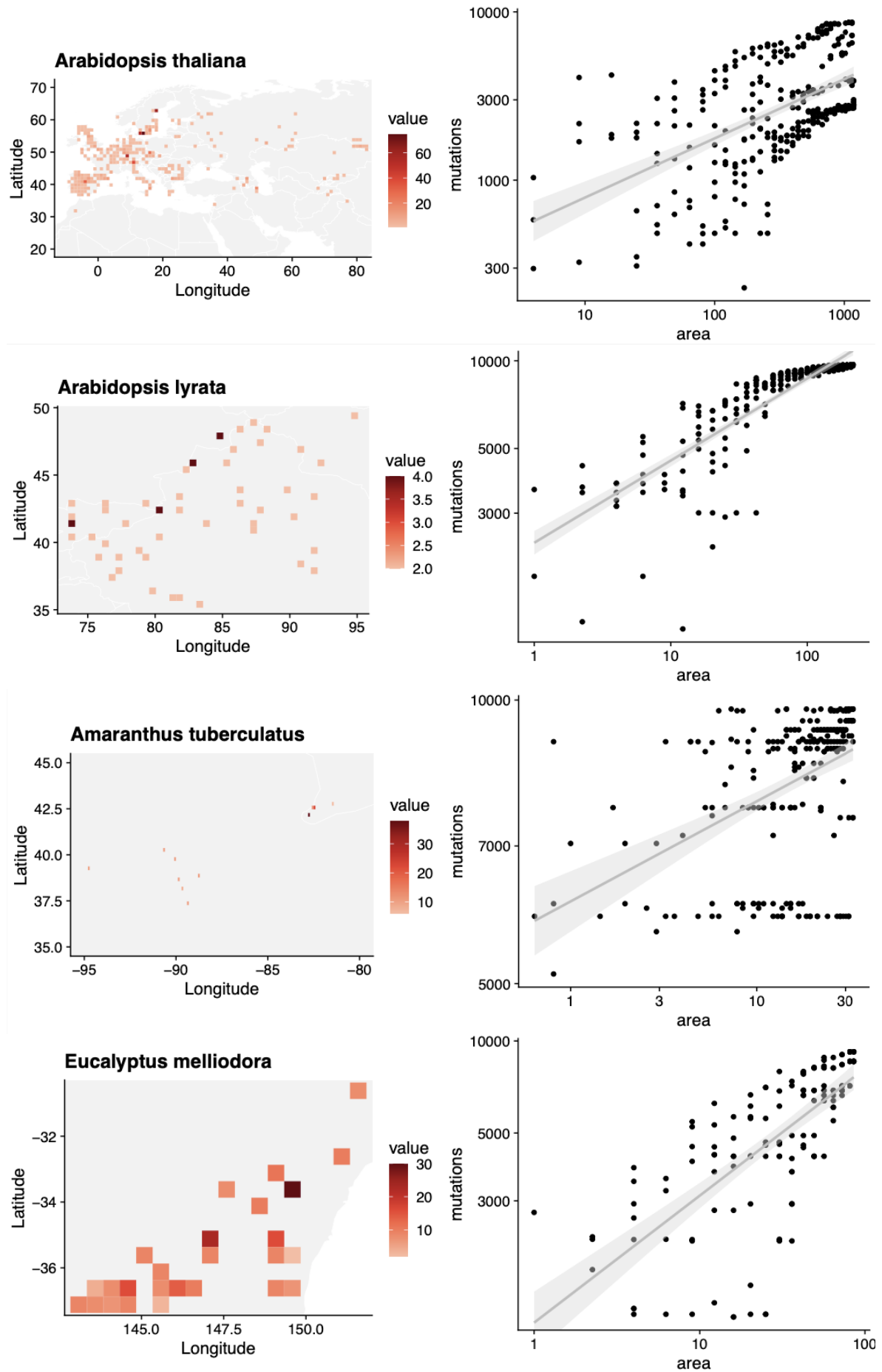
- 1132 - Supple et al. (44) generated a dataset of *Eucalyptus melliodora* of 275 individuals  
1133 from 36 broadly distributed populations. The dataset was produced by Illumina  
1134 sequence Genotyping-by-Sequencing (GBS) libraries digested with ApeKI as in  
1135 Elshire et al. (2011). The raw data is available at  
1136 <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA413429/>. The authors provided the  
1137 dataset in PLINK format. Genotyping rate was 0.769807 but we did not apply a  
1138 further filter to avoid reducing the total number of variants. We conducted analyses  
1139 with all 9378 SNPs. The genotyping rate in this dataset is likely not problematic as the  
1140 total number of GPS locations is 36, with multiple individuals sampled closely. This  
1141 sampling scheme probably allows to characterise an allele's distribution correctly  
1142 despite the lower genotyping rate.  
1143
- 1144 - Vallejo-Marin et al. (45) generated a GBS dataset of 521 *Mimulus* plants, with 286  
1145 samples being *Mimulus guttatus* from its native distribution. Libraries for  
1146 Genotyping-By-Sequencing were prepared with PstI enzyme as described in Twyford  
1147 & Friedman (2015) and sequenced using Illumina. The VCF of this dataset is  
1148 available at <http://hdl.handle.net/11667/168> and was also directly shared by the  
1149 authors. After applying a filtering for missingness, we ended up with a genotyping  
1150 rate of 0.904192 and 1,498 SNPs, which were used for the analyses.  
1151
- 1152 - Lovell & MacQueen (46) generated a WGS Illumina sequencing dataset of  
1153 Switchgrass, *Panicum virgatum*, of a collection of 732 individuals and 33,905,044  
1154 variants. The raw data is available at:  
1155 <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA622568>. The authors provided a  
1156 VCF file and latitude/longitude tables. 576 individuals were from natural collections.  
1157 The dataset contains also other collections such as cultivars, which were not used to  
1158 build the MAR. The genotyping rate was 0.976393 and analyses were conducted with  
1159 10,000 SNPs drawn from the largest chromosome.  
1160
- 1161 - MacLachlan et al. (47) generated a SNP chip dataset of *Pinus contorta* comprising  
1162 929 trees with latitude and longitude information and 32,449 SNPs. Genotyping was  
1163 conducted with the AdapTree lodgepole pine Affymetrix Axiom 50,298 SNP array  
1164 and data was provided in the supplemental material of the paper along with custom  
1165 scripts to parse the data. The database is available at  
1166 <https://datadryad.org/stash/dataset/doi:10.5061/dryad.ncjsxkstp>. The genome matrix  
1167 was transformed into PLINK. The genotyping rate was 0.959146, and analyses were  
1168 conducted with 10,000 randomly drawn SNPs. The fact that this dataset was created  
1169 with ascertained SNPs likely generates a frequency bias. In Fig. S22, one can see that  
1170 this may be a problem to calculate  $z_{MAR}$ , as the mutations~area graph appears  
1171 nonlinear and rapidly saturates. This confirms the expectation that SNPs are  
1172 ascertained to be common, as they are discovered immediately with very few samples.  
1173
- 1174 - Tuskan et al. (48) WGS Illumina sequenced 882 *Populus trichocarpa* trees. The  
1175 dataset includes 28,342,826 SNPs. The data is available under this DOI  
1176 <https://doi.ccs.ornl.gov/ui/doi/55> which redirects to a globus data sharing platform.  
1177 The authors provided the dataset as a VCF along with latitude/longitude coordinates.  
1178 This dataset was downsampled to the first chromosome. The genotyping rate was  
1179 0.921191, and 10,000 SNPs were randomly sampled for analyses.  
1180

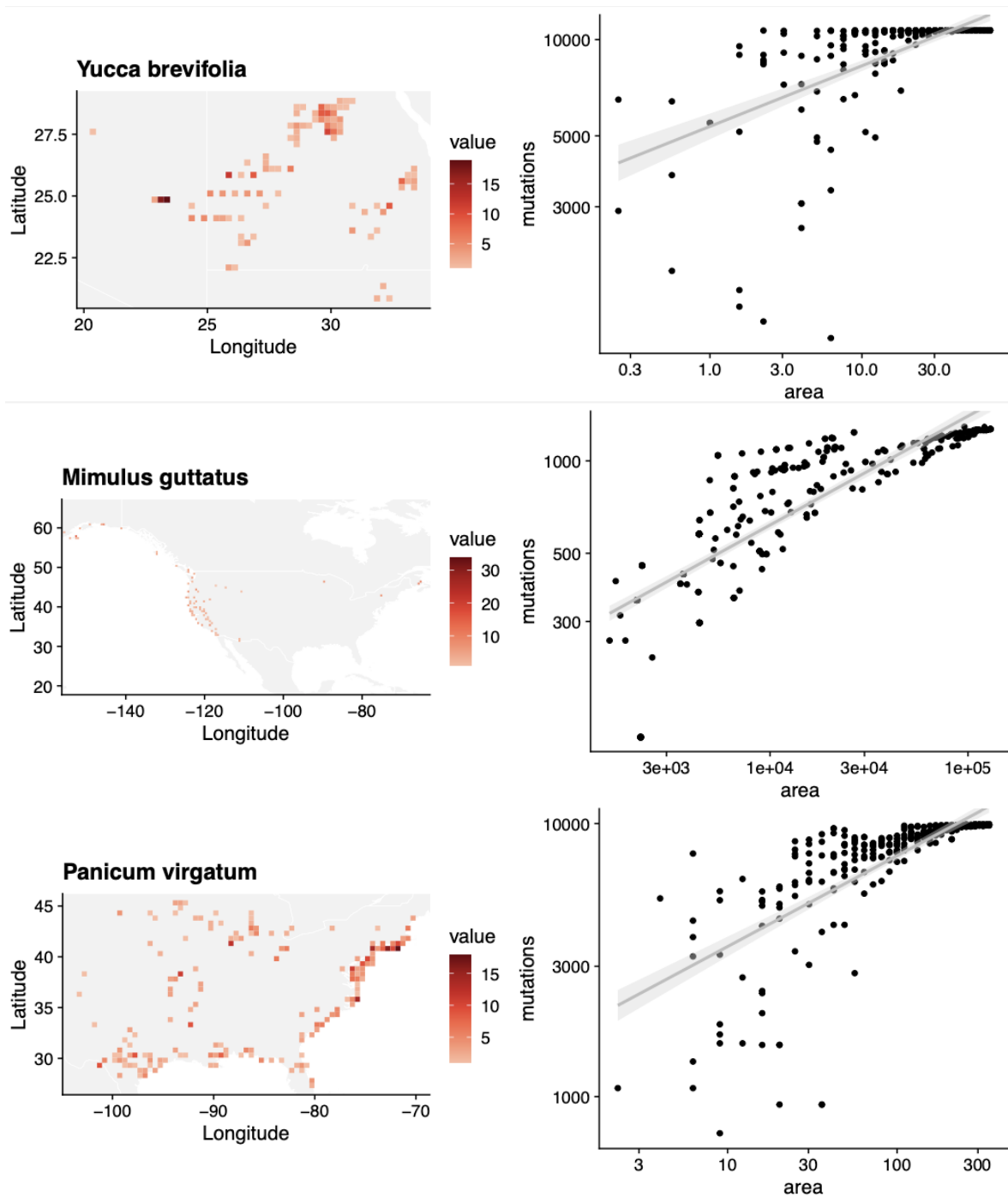
- 1181 - The Anopheles gambiae 1000 Genomes Consortium (49) (Phase 2) produced Whole-  
1182 Genome Illumina sequencing data for 1142 wild-caught mosquitoes of *Anopheles*  
1183 *gambiae*. All raw and processed data are available through  
1184 <https://www.malariagen.net/data>. We downloaded a VCF and latitude/longitude  
1185 coordinate files. The VCF was filtered for genotyping rate ending up at a 0.998895  
1186 rate. For efficiency, 10,000 randomly-selected SNPs from the VCF of the largest  
1187 chromosome 2L were used for analyses downstream.  
1188
- 1189 - Fuller et al. (50) WGS Illumina sequenced 253 coral individuals of *Acropora*  
1190 *millepora* in 12 reefs. The dataset was downloaded as fastq files from the published  
1191 online material from <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA593014>,  
1192 and SNPs were called as described in the supplemental material ending with  
1193 17,931,448, which were filtered to achieve a genotyping rate of 0.935709 for a total of  
1194 2,512 SNPs, which were used in the analyses.  
1195
- 1196 - Ruegg et al. (51) generated a dataset of 219 birds *Empidonax traillii*, for which 199  
1197 could be matched with geographic coordinates. SNPs were ascertained from several  
1198 publications using RAD seq and Fluidigm 96.96 IFC described and available in their  
1199 repository <https://github.com/eriqande/ruegg-et-al-wifl-genoscape>. A total of 349,014  
1200 SNPs were parsed using their custom scripts and we transformed them into PLINK  
1201 files. A genotyping rate filter was applied ending with a 0.96061 rate and 195,700  
1202 SNPs. 10,000 SNPs were selected at random for downstream analyses. Similarly, as  
1203 with the *Pinus contorta*, the incorporation of some ascertained SNPs in the dataset  
1204 based on Fluidigm technology could lead to quick saturation of the MAR curve (Fig.  
1205 S22).  
1206
- 1207 - Bay et al. (52) generated a dataset of 199 *Setophaga petechia* birds using a Restriction  
1208 site-associated DNA sequencing (RAD-Seq). The raw data is available at  
1209 <https://www.ncbi.nlm.nih.gov/bioproject/421926>. The authors shared a VCF file, with  
1210 a genotyping rate of 0.962419 and a total of 104,711 SNPs. 10,000 SNPs were  
1211 selected at random for downstream analyses.  
1212
- 1213 - Kingsley et al. (53) produced a dataset of 80 *Peromyscus maniculatus* deermice, for  
1214 which 78 could be matched with geographic locations. The SNP dataset was produced  
1215 using MY-select capture followed by Illumina sequencing. The VCF and PLINK files  
1216 are available via Figshare at <https://doi.org/10.6084/m9.figshare.1541235>. The dataset  
1217 included a total of 14,076 variants which were filtered to achieve a genotyping rate of  
1218 0.940411 for 2,946 SNPs, which were used in subsequent analyses.  
1219
- 1220 - We identified two published datasets for wolves. Smeds et al. (54) produced a WGS  
1221 Illumina sequencing dataset and combined it with pre-existing datasets for a total of  
1222 349 local dog breeds and wolves, of which 230 were *Canis lupus* from natural  
1223 populations. However, these samples did not have GPS locations assigned. The  
1224 second dataset we identified was from Schweizer et al. (55), which contained 107  
1225 geo-tagged grey wolves from North America using a capture and resequencing  
1226 approach for 1040 genes. The raw data is available at  
1227 <https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP065570>, and meta-data along  
1228 with a VCF area available at <https://doi.org/10.1111/mec.13467>. This data contained  
1229 13,092 SNPs at 0.993061 calling rate, and a better geographic resolution. We report  
1230 data for the second dataset.

- 1231  
1232 - The 1000 Genome Consortium (56) created WGS Illumina sequencing for over 2,504  
1233 humans and 24 unique geographic locations. We downloaded chromosome 1 from  
1234 [http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/datacollections/  
1235 1000G2504highcoverage/working/20190425NYGCGATK/](http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/datacollections/1000G2504highcoverage/working/20190425NYGCGATK/) and gathered the  
1236 population locations from [https://www.internationalgenome.org/data-  
1237 portal/population](https://www.internationalgenome.org/data-portal/population). To conduct analyses, we subsampled 10,000 SNPs at genotyping  
1238 rate 0.991069.  
1239  
1240 - Palacio-Mejia (57) used WGS for 591 *Panicum hallii* individuals to sequence at low  
1241 coverage. The raw data is available at  
1242 <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA390994>. The authors shared an  
1243 unfiltered VCF of 45,589 SNPs. Because of the low-coverage, stringent filters of  
1244 calling rates as used for other species would lead to removing all SNPs, and we settled  
1245 on a genotyping rate of 0.825824 for 242 variants, all of which were used for  
1246 downstream analyses.  
1247  
1248 - Royer et al. (58) produced a SNP dataset using RAD-Seq based Genotyping-By-  
1249 Sequencing of 290 *Yucca brevifolia* (Joshua Tree) individuals. A total of 10,695 SNPs  
1250 with a genotyping rate of 0.897501 were used for the analyses. The data was available  
1251 at Dryad <https://datadryad.org/stash/dataset/doi%253A10.5061%252Fdryad.7pj4t>.  
1252  
1253 - Kapun et al. (59) produced a WGS dataset of pooled *Drosophila melanogaster*,  
1254 sequencing ~80 pooled individuals from each of 271 populations as part of the  
1255 European "Drosophila Evolution over Space and Time" (DEST) project. A total of  
1256 5,019 shared SNPs with a genotyping rate of 0.937697 were used for analyses. The  
1257 dataset, both raw and processed, is available through <https://dest.bio>.  
1258  
1259 - Di Santo et al. (60) studied the highly-threatened species *Pinus torreyana*. They used  
1260 Genotyping-by-Sequencing of 242 individuals of the last remaining populations. The  
1261 dataset is not yet available through NCBI but the authors kindly shared a VCF directly  
1262 with us. From a total set of 166,564 SNPs with a genotyping rate of 0.964632, 10,000  
1263 were randomly selected for our analyses.  
1264  
1265 - von Seth et al. (61) studied the highly-threatened species *Dicerorhinus sumatrensis*.  
1266 They used Illumina WGS of 16 individuals of the last remaining populations. The raw  
1267 data is available at <https://www.ebi.ac.uk/ena/browser/view/PRJEB35511>. The  
1268 authors shared a VCF. In total, this comprises a set of 8,870,513 SNPs, with a  
1269 genotyping rate of 0.854862, which we did not further filter due to the small number  
1270 of individuals. For computational efficiency we selected 10,000 SNPs from the largest  
1271 chromosome.  
1272

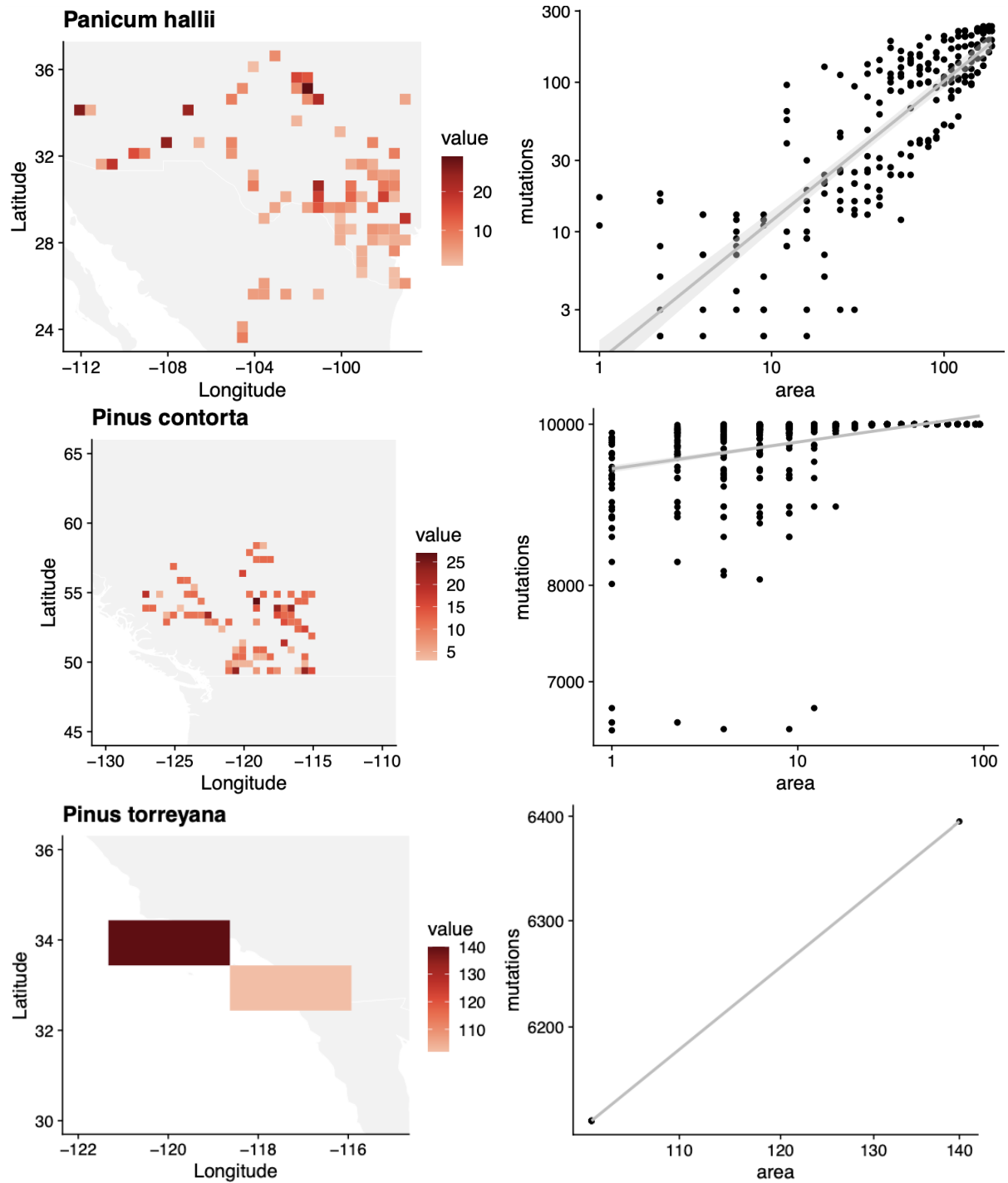
1273 Information and results per species are gathered in Table 1 and its extended version, Table  
1274 S10, and the average  $z_{MAR}$  across species are provided in Table S11.

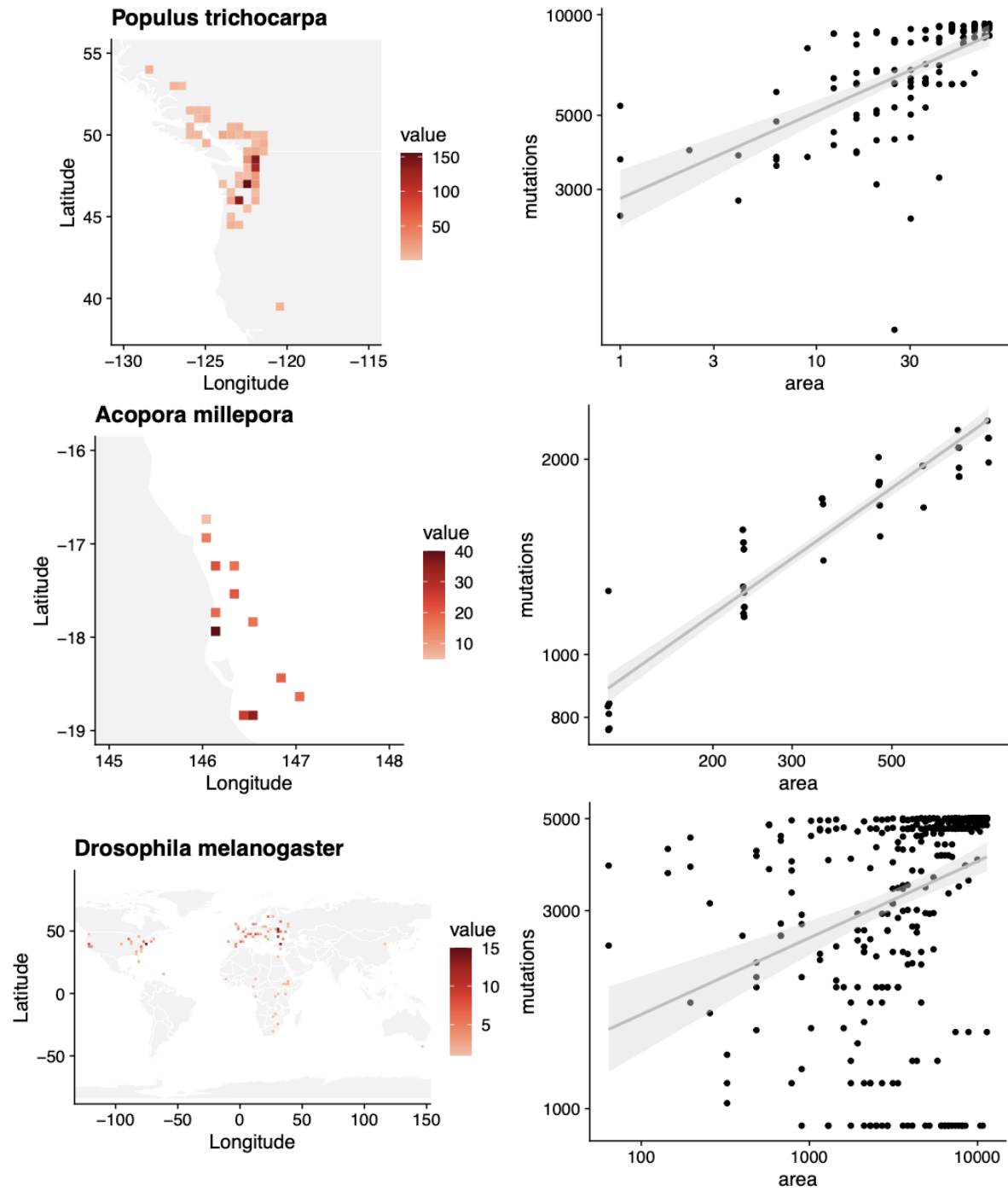
1275  
1276

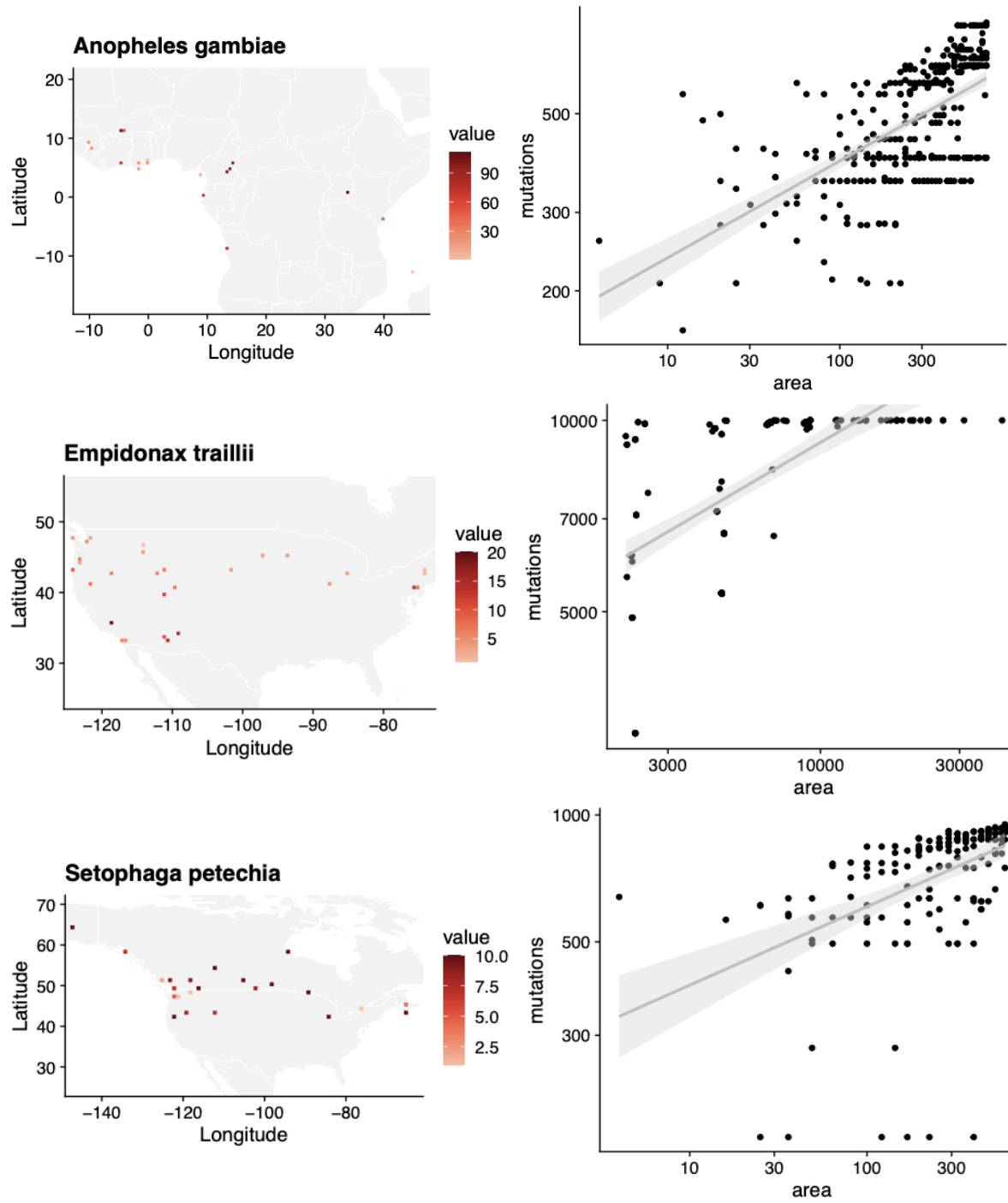


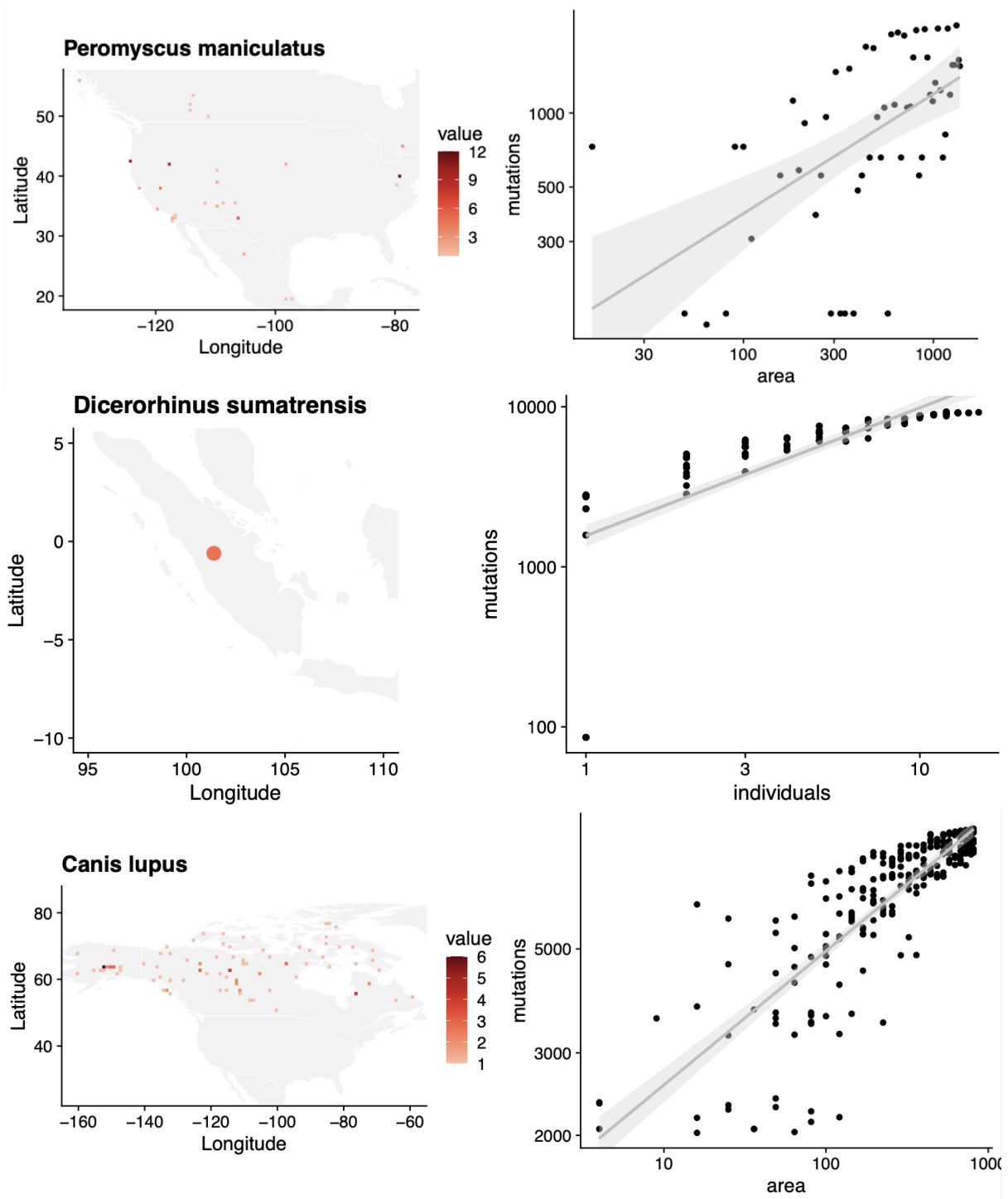


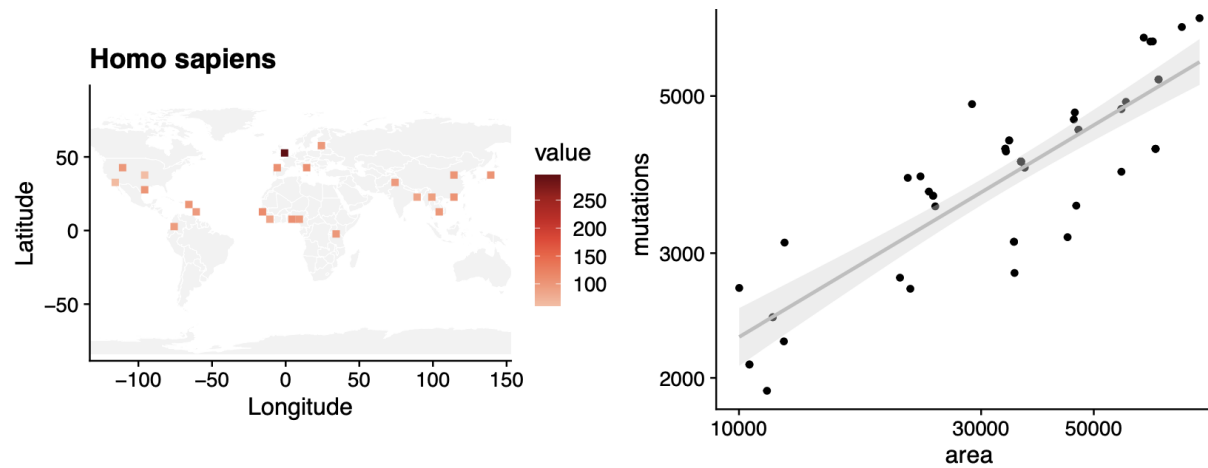












1277 **Fig. S22 | MAR summaries across species.**

1278 For each species we plot (left) the map of sample density in space and (right) the mutations-area relationship. (The locations  
1279 of 16 *Dicerorhinus sumatrensis* are unknown so only Sumatra is shown. *Pinus torreyana* was only found in two extant  
1280 populations.)

1281

1282 **Table S10 | The mutations-area relationship across species. Extended Table 1**

1283 The Mutations-Area Relationship (MAR) fitted with Area = Individuals and the scaled version. In the main text areas to  
1284 protect 90% of genetic diversity per species are provided given the scaled  $z^*$ . Here, we also provide the average estimated  
1285 area based on % of grid cells per species to be transformed from 2015 to 2050 using the LUH<sup>2</sup> dataset, the area where at  
1286 least 10% of grid cells will be transformed, and the genetic loss corresponding to those area transformations (see section  
1287 V.2).  
1288

Species (study)	SFS mod [ΔAIC]	MAR (A=N) $z_N$ [CI95%]	MAR scaled $z^*$ [CI95%]	LUH <sup>2</sup> change '50	LUH <sup>2</sup> >10% change '50	LUH <sup>2</sup> extinct '50	LUH <sup>2</sup> >10% extinct '50
<i>Arabidopsis thaliana</i> (29)	logN (85.8)	0.431 (0.423 - 0.439)	0.312 (0.305–0.32)	4.58	13.54	1.12	3.43
<i>Arabidopsis lyrata</i> (42)	logN (9592.4)	0.254 (0.238 - 0.27)	0.15 (0.136–0.165)	0.79	2.64	0.19	0.64
<i>Amaranthus tuberculatus</i> (43)	logN (7317.5)	0.244 (0.237 - 0.251)	0.142 (0.135–0.148)	4.86	11.13	1.19	2.79
<i>Eucalyptus melliodora</i> (44)	logN (157.5)	0.531 (0.526 - 0.536)	0.402 (0.397–0.406)	3.82	7.77	0.93	1.92
<i>Yucca brevifolia</i> (58)	logN(33300)	0.141 (0.128 - 0.155)	0.049 (0.037–0.062)	0.74	0	0.18	0
<i>Mimulus guttatus</i> (45)	logN (580.8)	0.342 (0.331 - 0.353)	0.231 (0.221–0.241)	3.78	NA	0.92	NA
<i>Panicum virgatum</i> (46)	logN (8345.2)	0.226 (0.215 - 0.237)	0.126 (0.116–0.136)	8.07	27.65	2	7.47
<i>Panicum hallii</i> (57)	logN (86)	0.983 (0.907 - 1.059)	0.814 (0.745 - 0.883)	3.78	11.36	0.92	2.85
<i>Pinus contorta</i> (47)	Wei (19413.7)	0.019 (0.018 - 0.02)	-	1.95	5.54	0.47	1.36
<i>Pinus torreyana</i> (60)	logN(766156)	0.239 (0.232 - 0.245)	0.105 (0.099–0.11)	25.4	NA	6.79	NA
<i>Populus trichocarpa</i> (48)	logS (0)	0.268 (0.257 - 0.28)	0.164 (0.154–0.175)	4.68	17.28	1.14	4.45
<i>Anopheles gambiae</i> (49)	logS (0)	0.221 (0.209 - 0.233)	0.121 (0.11–0.132)	9.95	21.96	2.48	5.78
<i>Acropora millepora</i> (50)	logN (452.3)	0.403 (0.395 - 0.41)	0.287 (0.28–0.293)	72.73	84.69	26.79	36.26
<i>Drosophila melanogaster</i> (59)	logN(33300)	0.445 (0.433 - 0.458)	0.324 (0.313–0.336)	0.95	NA	0.23	NA
<i>Empidonax traillii</i> (51)	Wei (640401.9)	0.169 (0.139 - 0.199)	0.074 (0.047–0.101)	5.55	15.14	1.36	3.86
<i>Setophaga petechia</i> (52)	ln (67138.5)	0.251 (0.236 - 0.267)	0.149 (0.135 - 0.163)	2.83	7.54	0.69	1.86
<i>Peromyscus maniculatus</i> (53)	logN (1449.7)	0.844 (0.769 - 0.919)	0.68 (0.613–0.748)	5.61	13.68	1.38	3.47
<i>Dicerorhinus sumatrensis</i> (61)	w (107864.2)	0.474 (0.449 - 0.498)	0.123 (0.106–0.14)	0.25	NA	0.06	NA
<i>Canis lupus</i> (55)	logN (85.8)	0.29 (0.28 - 0.301)	0.183 (0.174–0.193)	0.23	NA	0.06	NA
<i>Homo sapiens</i> (56)	logN (9592.4)	0.395 (0.339 - 0.451)	0.28 (0.229–0.331)	28.81	40.13	7.83	11.58

1289  
1290  
1291

Extended acronyms:  
logN: log Normal distribution. logS: log Series distribution. Wei: Weibull distribution.

1292 **IV.1 Exclusion of species from global averages**

1293

1294 To avoid contaminating across-species averages of  $z_{MAR}$  with estimates of species whose data  
 1295 we do not fully trust, we conducted global averages excluding species for which we are not  
 1296 confident  $z_{MAR}$  reflects the correct species diversity-area relationships.

1297  
 1298 *Pinus contorta* showed a lower  $z_{MAR}$  than what is expected in a theoretical baseline from  
 1299 individual sampling (section II). This is most likely due to this being the only species for  
 1300 which SNPs were previously ascertained to be intermediate frequency (i.e. the genome  
 1301 technology was a SNP chip). This alters SFS, so we are not confident the  $z_{MAR}$  is the true  
 1302 parameter of the species.

1303  
 1304 *Yucca brevifolia* was a dense sampling of several local populations within a constrained area  
 1305 that is a hybrid zone. Since this species was not sampled range-wide we do not feel confident  
 1306 to include it in downstream analyses. The species also has a lower  $z$  than expected (Fig. S5)

1307  
 1308 *Pinus torreyana* only has two wild populations left, and therefore the MAR is based on two  
 1309 area sizes (Fig. S22). Because this is such a threatened species with already most of its range  
 1310 loss, we do not have confidence in the  $z$  parameter.

1311  
 1312 *Dicerorhinus sumatrensis* has only ~30 estimated adult individuals in the wild. Again we do  
 1313 not have confidence in the  $z$  parameter in such extinction-edge cases.

1314  
 1315 *Homo sapiens*. We exclude our own species.  
 1316

1317 **Table S11 | Mean  $z_{MAR}$  and other summary statistics across species.**

1318 We selected those species that did not show artefacts in Fig. S22 or whose  $z_{MAR}$  overlapped with 0 to calculate a species-  
 1319 wide mean. See section IV.1.  
 1320

	$z_{MAR}$	$z_{MAR} (A=N)$	$z^*_{MAR} \text{ scaled}$
mean	0.31	0.39	0.27
mean se	0.038	0.053	0.048
median	0.25	0.29	0.18
IQR	0.15	0.19	0.17

1321  
 1322  
 1323 Although we could not see any obvious patterns relating  $z_{MAR}$  with certain groups of  
 1324 species (Table 1), we wondered whether any life history trait of the species analysed could  
 1325 explain the variation we observed (see Table S12 of traits). An ANOVA did not show any  
 1326 significant relationship. Because we know theoretically this parameter must be related to the  
 1327 degree of dispersal ability of genotypes of a species relative to the whole species geographic  
 1328 range, we expect traits involved in determining these to be good predictors. Future work will  
 1329 be necessary to validate this, as the sample size (n=19) may not permit enough power to  
 1330 detect these expected patterns.

1331  
 1332 **Table S12 | Traits, life history, and other characteristics of the analyzed species.**

Species	RedList	Known Decline	Kingdom	Reproduction	Pollination	Mobility	AreaRange
<i>Arabidopsis thaliana</i>	NO	NO	Plantae	Selfing	Selfing	Sessile	27337467.4
<i>Arabidopsis lyrata</i>	NO	NO	Plantae	Outcrossing	Vector	Sessile	2791301.4
<i>Amaranthus tuberculatus</i>	LC	NO	Plantae	Outcrossing	Vector	Sessile	804124.8
<i>Eucalyptus melliodora</i>	VU	NO	Plantae	Outcrossing	Wind	Sessile	948699.3
<i>Yucca brevifolia</i>	LC	YES	Plantae	Outcrossing	Vector	Sessile	1213454.4
<i>Mimulus guttatus</i>	LC	NO	Plantae	Outcrossing	Vector	Sessile	25138310.6
<i>Panicum virgatum</i>	LC	NO	Plantae	Outcrossing	Wind	Sessile	6291400.2

<i>Panicum hallii</i>	NO	NO	Plantae	Outcrossing	Wind	Sessile	2188807.4
<i>Pinus contorta</i>	LC	NO	Plantae	Outcrossing	Wind	Sessile	886182.2
<i>Pinus torreyana</i>	CR	YES	Plantae	Outcrossing	Wind	Sessile	30781.95
<i>Populus trichocarpa</i>	LC	NO	Plantae	Outcrossing	Wind	Sessile	1119664.1
<i>Drosophila melanogaster</i>	NO	NO	Animalia	Outcrossing	Activemating	Fly	115208408
<i>Anopheles gambiae</i>	NO	NO	Animalia	Outcrossing	Activemating	Fly	19959809.9
<i>Acropora millepora</i>	NT	YES	Animalia	Outcrossing	Activemating	Fly	26725.9
<i>Empidonax traillii</i>	LC	YES	Animalia	Outcrossing	Activemating	Fly	7027395.2
<i>Setophaga petechia</i>	LC	NO	Animalia	Outcrossing	Activemating	Fly	15172431.15
<i>Peromyscus maniculatus</i>	LC	NO	Animalia	Outcrossing	Activemating	Mobile	22609152.6
<i>Dicerorhinus sumatrensis</i>	CR	YES	Animalia	Outcrossing	Activemating	Mobile	3335605.58
<i>Canis lupus</i>	LC	NO	Animalia	Outcrossing	Activemating	Mobile	19102403.5
<i>Homo sapiens</i>	NA	NA	NA	NA	NA	NA	80763121.8

1333

1334

1335

1336

**Table S13 | Association of traits, life history, and other characteristics with  $z_{MAR}$ .**

Acronyms: NO=not assessed but likely non-threatened, LC=low concern, VU=vulnerable, CR=critically endangered

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
RedList	4	0.0952396	0.0238099	0.5580988	0.7040464
KnownDecline	1	0.0275537	0.0275537	0.6458527	0.4580865
Kingdom	1	0.0011684	0.0011684	0.0273876	0.8750400
Reproduction	1	0.0003238	0.0003238	0.0075890	0.9339612
Pollination	1	0.0375975	0.0375975	0.8812784	0.3909509
Mobility	1	0.1600627	0.1600627	3.7518370	0.1104995
AreaRange	1	0.0174745	0.0174745	0.4095989	0.5503439
Residuals	5	0.2133125	0.0426625	NA	NA

1337

1338

1339

1340

1341

1342

1343

1344

1345

1346

While no association between life history and  $z_{MAR}$  was found (Table S13), this may be due to limited power, as the sample size of species analysed here is still small,  $n=20$ . Further studies expanding the numbers of species will be necessary to confirm or reject this expected association.

1347 **V. An estimate of global genetic diversity loss**

1348

1349 Using the approach described in section II.4, we generated a number of estimates either per  
1350 ecosystem or per species. All estimates below tried to be conservative, and thus we always  
1351 used the scaled  $z_{MAR}$  values (section II.3.2.)

1352

1353 **V.1 Estimates of ecosystem area losses**

1354

1355 **Table S14 | Millennium Ecosystem Assessment land cover transformation.**

1356 *Changes of ecosystem area pre-21<sup>st</sup> century. Ecosystem names are repeated for ecosystem sub-classes.*

1357 Source: <https://www.millenniumassessment.org>

1358

1359

System	Area (km <sup>2</sup> x10 <sup>6</sup> )	Earth % surface	Protected areas (%)	Area transformed (%)
MARINE	349.3	68.6	0.3	NA
COASTAL	17.2	4.1	7	NA
- TERRESTRIAL	6	4.1	4	11
- MARINE	11.2	2.2	9	NA
INLAND WATER	10.3	7	12	11
FOREST/WOODLAND	41.9	28.4	10	42
- TROPICAL	23.3	15.8	11	34
- TEMPERATE	6.2	4.2	16	67
- BOREAL	12.4	8.4	4	25
DRYLAND	59.9	40.6	7	18
- HYPERARID	9.6	6.5	11	1
- ARID	15.3	10.4	6	5
- SEMIARID	22.3	15.3	6	25
- SUBHUMID	12.7	8.6	7	35
ISLAND	7.1	4.8	17	17
- STATES	4.7	3.2	18	21
MOUNTAINS	35.8	24.3	14	12
- 300-1000	13	8.8	11	13
- 1000-2500	11.3	7.7	14	13
- 2500-4500	9.6	6.5	18	6
- 4500+	1.8	1.2	22	0.3
POLAR	23	15.6	42	0.38
CULTIVATED	35.3	23.9	6	47
- PASTURE	0.1	0.1	4	11
- CROPLAND	8.3	5.7	4	62
- MIXED	26.9	18.2	6	43
URBAN	3.6	2.4	0	100
GLOBAL	510	NA	4	38

1360

1361

1362

1363 Ecosystem transformation has been tracked over decades. We extracted ecosystem  
1364 transformations from the Millennium Ecosystem Assessment (62), which estimated  
1365 ecosystem transformations from presumably native systems to cultivated or urban areas by  
1366 GLC2000 land cover dataset (Table S14). The forest/woodland is calculated as percentage  
1367 change between potential vegetation from WWF ecoregions to the current actual  
1368 forest/woodland areas from GLC2000. These provide bulk ecosystem reductions, not for a  
1369 given species, but may be a good proxy for an average across species.

1370

1371 **Table S15 | IPBES land cover transformation.**

1372 Source: <https://ipbes.net>

1373

Region	Area(Mkm2)	MSA 2010	MSA 2050 SSP2	MSA 2050 SSP1	MSA 2050 SSP3
North America	20	65	56	NA	NA
Central and South America	18	65	53	NA	NA
Middle East and Northern Africa	11	81	77	NA	NA
Sub-Saharan Africa	24	70	56	NA	NA
Western and Central Europe	6	37	29	NA	NA
Russian region and Central Asia	21	73	65	NA	NA



South Asia	5	44	35	NA	NA
China region	11	56	49	NA	NA
Southeast Asia	7	55	43	NA	NA
Japan, Korea and Oceania	8	71	57	NA	NA
Polar	2	96	91	NA	NA
World	132	66	56	62	54

1374  
1375  
1376  
1377

1378 The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem  
1379 Services (IPBES) recently used a PBL satellite product from the Netherlands Environmental  
1380 Assessment Agency (<https://www.pbl.nl/en/nature-and-biodiversity>) to study the % of area  
1381 ecosystem transformation in the world (Table S15). This provides an updated estimate to the  
1382 Millennium Assessment as well as projections under several Shared Socioeconomic  
1383 Pathways (1-3) for 2050. These were reported per region as of 2010, and for projections to  
1384 2050 (scenario SSP2). Instead of direct area, the metric is a composite of land use  
1385 information to predict Mean Species Abundance (MSA), a measure of the size of populations  
1386 of wild organisms as a percentage of their inferred abundance in their natural state (% MSA).

1387  
1388 A global transformation metric can also be captured by the most updated land use  
1389 transformation data, the Land Use Harmonization 2 (release v2e for 2015-2011 and release  
1390 v2h for baseline 1850-2015) (63). Baseline transformation of primary ecosystems was  
1391 calculated subtracting the total area covered by primary forest (primf) and primary non-forest  
1392 (primn) variables between year 1850 layer (roughly pre-industrial baseline) and the present,  
1393 2015, as  $1 - A_{2015} / A_{1850}$  (Table S16). Analyses that use projections to mid-21<sup>st</sup> century were  
1394 conducted similarly as in (64), summing over all transitions from primary forest (primf),  
1395 primary non-forest (primn), secondary forest (secdf) and secondary non-forest (secdn) lands  
1396 to any other category for all years within the 2015-2050 period (see Table S10).

1398 **Table S16 | Land Use Harmonization 2 from 1850 to 2015**  
1399 Source: <https://luh.umd.edu/data.shtml>

	Area %
Primary forest transformed	43
Primary non-forest transformed	50

1401  
1402 Finally, we searched for timely estimates of forest reduction (based on vegetation  
1403 cover) reported in the Global Forest Watch website:  
1404 <https://www.globalforestwatch.org/dashboards/global/> (accessed June 2021). From 2002 to  
1405 2020, there has been a global tree cover loss of 10%, with an annual tree cover loss of 0.6-  
1406 1.1%.

1407  
1408 Although these are not direct area transformations, we also used the IUCN Red List  
1409 resource (<https://www.iucnredlist.org>, Table S12 shows status of the species analysed here),  
1410 which includes guides to categorise species as vulnerable, endangered, critically endangered,  
1411 and extinct, and has conducted extensive assessments across thousands of species (Table  
1412 S17).

1413  
1414 **Table S17 | IUCN Red List categories of extinction risk and number of species.**  
1415 Source: [www.iucnredlist.org](http://www.iucnredlist.org), January 2021

IUCN Red List	Description	Criterion of area or pop.	# plant species
---------------	-------------	---------------------------	-----------------

Category		reduction (>%)	
EX	Extinct	100	164
EW	Extinct in the Wild	100	
CR	Critically Endangered	80	4674
EN	Endangered	50	8593
VU	Vulnerable	30	8459
NC, LR, NT, DD,LC	No Concern, Low Risk, Near Threatened, Data Deficient, Least Concern, Other	0	32237

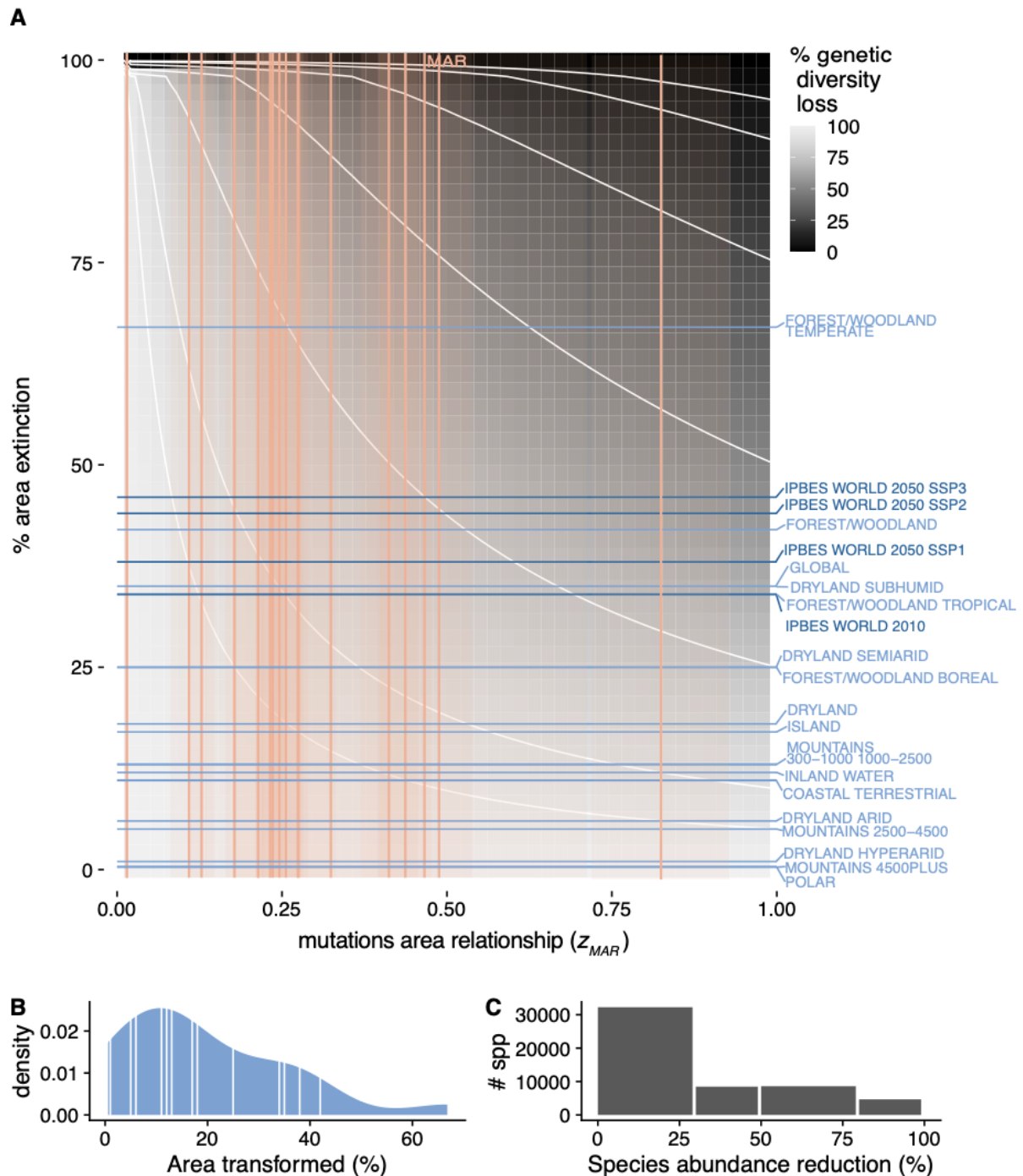
1417  
1418

## 1419 **V.2 A global estimate of genetic loss**

1420

1421 Taking the estimates and standard error of  $z_{MAR}$  across species, and the world's reduction of  
1422 ecosystems we can calculate the fraction of genetic diversity reduction following the MAR  
1423 equation (section II.4), giving a range of estimates (Table S18).

1424



1425

1426 **Fig. S23 | The parameter space of genetic diversity loss, extended**

1427 (A) The theoretical space of genetic diversity loss.  $z_{MAR}$  values (using area, unscaled for samples, differently from Fig. 32)  
 1428 computed for species analyzed here are marked as orange vertical lines, with confidence intervals as orange shading. Blue  
 1429 horizontal lines correspond to ecosystem transformations from the Millennium Assessment (light blue) and IPBES  
 1430 Assessment (dark blue) (B) Density histogram of percentages of area transformed across ecosystems from the MA, with  
 1431 averages per ecosystem marked in the distribution as well as horizontal lines in (A). (C) The number of species of each of the  
 1432 IUCN categories and the most optimistic range of area or abundance reduction for each of the category brackets.

1433

1434

1435 **Table S18 | Estimates of average expected genetic loss for different ecosystems.**

1436 Assuming ecosystem transformation approximately translates into average species distribution reduction, and using the  
 1437 ranges of  $z_{MAR}$  from Table 1 of the main text, we project the average genetic loss using the Mutations Area Relationship.

1438

System	Area transformed (%)	Genetic loss % (mean z based)	Genetic loss % (min z based)	Genetic loss % (max z based)
--------	----------------------	-------------------------------	------------------------------	------------------------------

COASTAL TERRESTRIAL	11	3.2	0.9	9
INLAND WATER	11	3.2	0.9	79.7
FOREST/WOODLAND	42	14.0	4	35.8
FOREST/WOODLAND TROPICAL	34	10.5	3	28.7
FOREST/WOODLAND TEMPERATE	67	26.5	7.9	59.4
FOREST/WOODLAND BOREAL	25	7.7	2.1	20.9
DRYLAND	18	5.4	1.5	14.9
DRYLAND HYPERARID	1	0.3	0.1	0.8
DRYLAND ARID	5	1.4	0.4	4.1
DRYLAND SEMIARID	25	7.7	2.1	20.9
DRYLAND SUBHUMID	35	11.3	3.2	29.6
ISLAND	17	5.0	1.4	14.1
MOUNTAINS	12	3.5	0.9	9.9
MOUNTAINS 300-1000	13	3.8	1	10.7
MOUNTAINS 1000-2500	13	3.8	1	10.7
MOUNTAINS 2500-4500	6	1.7	0.5	4.9
MOUNTAINS 4500+	0.3	0.1	0	0.2
POLAR	0.4	0.1	0	0.3
GLOBAL	38	12.4	3.5	32.2

1439  
1440

1441 Assuming the average  $z_{MAR}$ , and utilising tree cover from the Global Forest Watch  
1442 (<https://www.globalforestwatch.org>), which estimates 0.6-1.1% of transformation per year  
1443 across Canada, United States and Australia, we extrapolated genetic diversity loss in the next  
1444 50 years for tree species to be 8-15% genetic diversity loss.

1445

1446 Assuming that the calculated  $z_{MAR}$  estimates (Table 1) are representative of plant  
1447 species, we conducted an experiment to create a distribution of % of genetic diversity loss in  
1448 threatened species. We used the number of species in each IUCN category (Table S17) for a  
1449 total of 54,127 plant species. For plant species, one of the evaluation criteria of percentage of  
1450 population loss likely translates faithfully to area reduction in the species. Thus, the  
1451 proportion of species per category gives a discrete probability distribution of the ranges of  
1452 percentage of area loss:  $P(0-29\%)=0.596$ ,  $P(30-49\%)=0.156$ ,  $P(50-79\%)=0.159$ ,  $P(80-$   
1453  $99\%)=0.086$ ,  $P(99\%-100\%)=0.003$ . Using a simulation-based sampling approach, we drew  
1454 350,000 random area reductions  $A_t/A_{t-1}$  from the previous distribution and a  $z_{MAR}$  from the  
1455 mean and variance of our estimates from Table 1 for plants. These were plugged into the  
1456 MAR equation (Section II.4) to calculate the percentage of genetic diversity loss of these  
1457 350,000 random draws. The resulting distribution had a median and interquartile range of  
1458 17.53 % [7.51- 31.82]..

1459

1460 Using the Land Use Harmonization 2 dataset, we also create per-species predictions  
1461 based on the % transformation of each of the sampled regions per species (Table S14). As  
1462 before, the land use transformations that merit be considered area losses are all transitions  
1463 from primary forest (primf), primary non-forest (primn), secondary forest (secdf) and  
1464 secondary non-forest (secdn) lands to any other category. Taking all the locations where each  
1465 species has been sampled, we extracted the predicted % of land use change per cell and  
1466 summed over all cells where individuals had been sampled (we call this LUH<sup>2</sup> change ‘50,  
1467 see column in Table S10). We also produced the alternative area loss estimate taking that at  
1468 least 10% predicted habitat transformation for a grid cell renders the entire area of that grid  
1469 cell as impacted or lost (we call this LUH<sup>2</sup> >10% change ‘50). These per-species area losses,  
1470 in combination with the matched  $z_{MAR}$ , provided a range of potential loss estimates to 2050  
1471 ranging 0-36% depending on the species (Table S10).

1472

1473

### 1474 V.3 Community ecology simulations and MAR

1475

1476 To test whether intermediate levels of MAR would be expected across species in entire  
1477 ecosystems, we conducted community assembly simulations of ~100-500 species following  
1478 the Neutral Theory of Biodiversity (1, 41) and coalescent simulations (23) using the software  
1479 MESS (65). These simulations are computationally demanding and could not run in a  
1480 complete 2D spatial grid. Instead, they were simulated in a mainland-island system, with  
1481 islands of increasing areas. The community forms by species colonising an empty island  
1482 according to Hubbell's Unified Neutral Theory of Biodiversity and Biogeography (UNTB),  
1483 where all species are equally likely to colonise and persist in the local community. Continued  
1484 colonisation and migration to the local community continues to bring in new species that may  
1485 or may not survive, while also continuously bringing in individuals of species already in the  
1486 local community. The community assembly process ends when the community has reached  
1487 an equilibrium denoted as the balance between local extinction and new species dispersing  
1488 into the area (Hubbell 2001). Once the forward-time process has ended, we simulate the  
1489 coalescent history of each species backward in time. For this, MESS considers the population  
1490 size, divergence time, and migration rates of the meta and local communities. These  
1491 coalescent simulations provide us with genetic data and ultimately diversity estimates for  
1492 each species in the community.

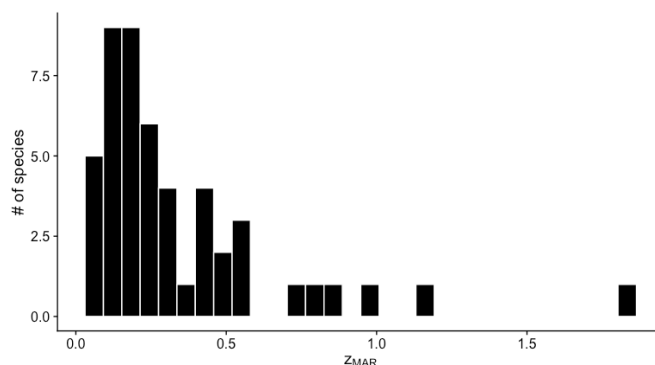
1493

1494 We simulated 100 MESS communities, and for each community the size of the local  
1495 community was varied from 1K to 100K. We varied the size of communities to emulate  
1496 variation in area occupied by a given community because we assume as the number of  
1497 individuals in a community increases from 1,000 to 100,000, so does the area occupied. All  
1498 other parameters were kept consistent across each of these community simulations, and most  
1499 remained at their default value. The parameters changed were the length of the sequences  
1500 simulated for the coalescent-based simulations, which was fixed at 10,000 bp, and the  
1501 migration rate, which was fixed at 0.01.

1502

1503 The simulation output was used to then compute a single  $z_{SAR}$  for the system as  
1504  $S=cA^{z_{SAR}}$ , and one  $z_{MAR}$  for each species in the same way,  $M=cA^{z_{MAR}}$ . This resulted in the  
1505 distribution of  $z_{MAR}$  from Fig. S24. This confirmed that we can recover typical  $z_{SAR}$  and  $z_{MAR}$   
1506 values from completely stochastic neutral yet spatially structured systems such as species in  
1507 communities and mutations in populations of a species.

1508



1509

1510 **Fig. S24 |  $z_{MAR}$  calculated from MESS eco-evolutionary simulations**

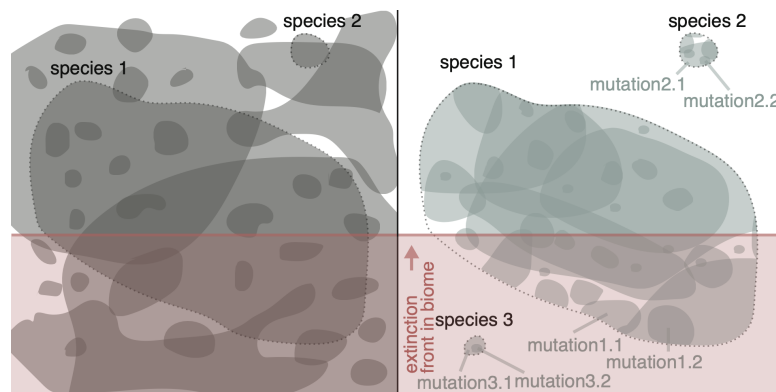
1511 Using the MESS framework of a mainland-island model with different island sizes,  $z_{MAR}$  per species is recovered. The  
1512 stochastic nature of the simulations results in each species having different abundances and migration histories that change

1513 the scaling value. Values were typically around 0.3. Rarely some species had values above 1, which appear could be noisy  
1514 estimates from recently colonising species in the simulations.  
1515  
1516  
1517

#### 1518 **V.4 The nested species extinction and genetic diversity loss processes** 1519

1520 Finally, we worried that our estimates of V.2 would be mistaken as overestimates. In fact, we  
1521 believe these may be underestimated. Recent policy proposals for the United Nations'  
1522 Sustainability Goals emphasize that the target of protecting 90% of species genetic diversity  
1523 for all species cannot leave the already-extinct species behind (66) (That is, one cannot  
1524 protect 90% of species and leave 10% to become extinct to meet this goal). This clearly  
1525 exemplifies a problem in conservation biology that what researchers can study is (most of the  
1526 time) what has escaped extinction, and therefore if we do not account for extinct species in  
1527 our overall estimates of genetic diversity loss we may naively think ecosystems have not  
1528 suffered genetic diversity loss (i.e. in the extreme scenario, an ecosystem that has lost all but  
1529 one abundant species may not really appear genetically eroded if such species is in good  
1530 shape).  
1531

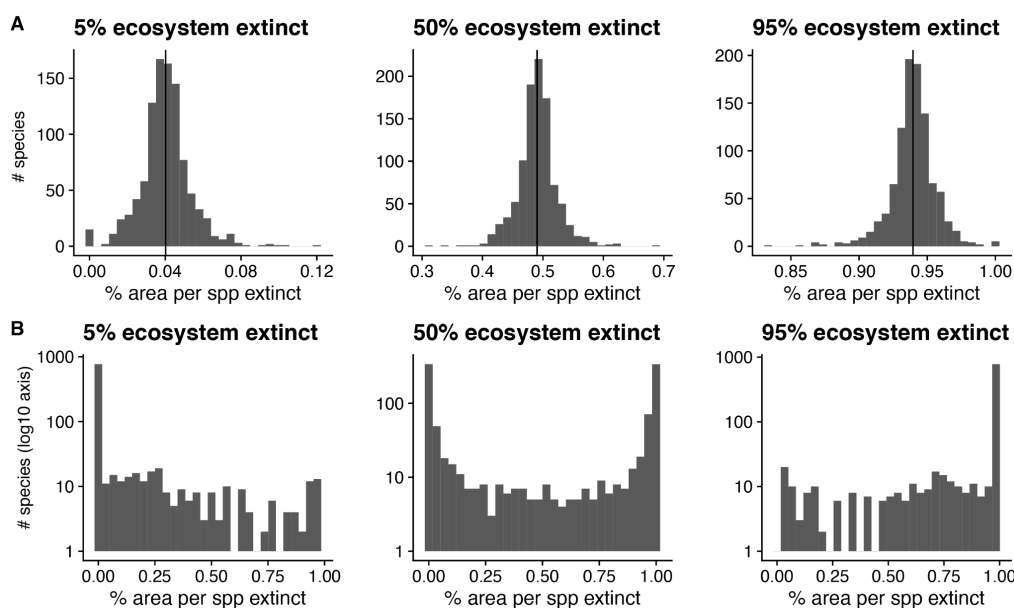
1532 We then created spatial simulations in R where 1,000 species are distributed in  
1533 100x100 grid cells following a UNTB abundance distribution and then proceeded with an  
1534 edge extinction of the ecosystem (see Fig. S25 for a cartoon).  
1535  
1536  
1537



1538  
1539 **Fig. S25 | Cartoon of nested extinction of species and genetic diversity loss.**  
1540 *An ecosystem with multiple species within it (left), distributed in space, with few species broadly distributed and many*  
1541 *narrowly distributed. Moving one level of biological organization lower, mutations within species (right) are also spatially*  
1542 *distributed with many narrowly distributed. As extinction happens (red line moving bottom to top), all species below the red*  
1543 *line go extinct, but only the mutations within species 1 below the line are lost, while mutations above the line remain.*  
1544 *Species 3 has already become extinct, and therefore also all the mutations within it.*  
1545  
1546

1547 Two extreme types of distributions of species can be imagined: species are randomly  
1548 placed in space, or species are found mostly in perfectly contiguous ranges (We ended up  
1549 using as an example a simulation with 85% of the individuals of a species found in a core  
1550 square continuous distribution and 15% found outside that core in fragmented observations,  
1551 as this scenario produced the canonical SAR of  $z \sim 0.3$ ). Spatial structure interestingly creates  
1552 two extreme distributions of area reductions across species (Fig. S26): random placement of  
1553 cell habitats essentially show that the average area reduction per ecosystem is followed by  
1554 most species, while autocorrelated placement of cell habitats create a U distribution in area  
1555 reductions, where at the beginning of the extinction process most species have not

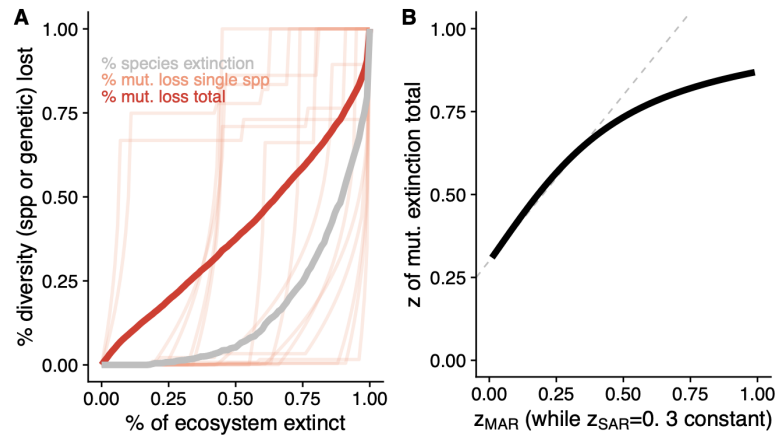
1556 experienced any impact (Fig. S26B left) but at the end of ecosystem reduction virtually all  
 1557 species are already extinct (Note we may be at the beginning of S26B process given the data  
 1558 from IUCN, Fig 3C).  
 1559  
 1560



1561

1562 **Fig. S26 | The distribution of per-species area lost and total ecosystem extinction with 1000 species**  
 1563 *Two ecosystems of 100x100 cells with 1000 species. Species are either randomly distributed in cells (A) or spatially*  
 1564 *autocorrelated with occupying mostly contiguous cells (B). As the extinction process wipes out part of the ecosystem*  
 1565 *(snapshots are provided at 5%, 50%, and 95%), the area loss per species (and hence genetic diversity lost) is tracked. In (A)*  
 1566 *the average area lost per species is roughly the total reduction of the ecosystem, whereas in (B) the distribution is U shaped*  
 1567 *(note the log-scaled y-axis). While in (B) the mean area lost in the distribution correctly captures the area loss of the*  
 1568 *ecosystem, per species losses are highly uneven.*  
 1569

1570 To study the consequence of the above differential area loss and the effect of some  
 1571 species going extinct on the total ecosystem genetic diversity, we conducted the next  
 1572 analysis: For extant species, we assumed they would lose genetic diversity following the  
 1573 MAR relationship (section II.4), with all species having  $z_{MAR} = 0.3$  for simplicity (i.e. all  
 1574 species lose genetic diversity at the same rate). For extinct species (100% of their area  
 1575 reduced), we considered genetic diversity loss was 100%. The compound total genetic  
 1576 diversity loss would then just be the sum of those  $X_{Tot} = \sum_{i=1}^{1000} X_i$  (Of course, in reality species  
 1577 may vary in their genome-wide diversity average, and we could for instance use Watterson's  
 1578  $\Theta_W$  (see section II.2) to scale the total loss of genetic diversity in the ecosystem accounting  
 1579 for different basal level of diversity per species:  $\sum_{i=1}^{1000} \Theta_{Wi} X_i$ ). Interestingly, if we calculate the  
 1580  $z$  of the slope of compound genetic diversity across species in an ecosystem it is much larger  
 1581 than MAR or SAR alone:  $z_{compounded} = 0.6$  (Fig. S27).  
 1582



1583

1584

1585

1586

1587

1588

1589

1590

1591

1592

1593

1594

1595

1596

1597

**Fig. S27 | Numeric simulation of nested species and genetic diversity loss.**

(A) Simulating the extinction of an ecosystem with 1,000 species that follow a log-normal species abundance curve. Extinction of the ecosystem creates a curve of species loss of  $z \sim 0.3$  (grey). Likewise, each species trajectory (light red, 15 species drawn randomly) follows a simulated genetic diversity loss of  $z_{MAR} \sim 0.3$  as they lose area. Because species' geographic distributions are by construction smaller than the whole ecosystem area, those distributed closer to the start of the extinction front lose area first, while those distributed farthest from the extinction front only lose area when the ecosystem is almost completely destroyed. Because genetic diversity loss is both due to complete extinction of species as well as area reduction of extant species, the compound genetic diversity loss curve (red) follows the faster loss dynamics. (B) Holding  $z_{SAR} = 0.3$  constant, and varying  $z_{MAR}$  in independent simulations shows that the compound genetic diversity across species is close to the sum of both  $z$  slopes (the SAR and the MAR), but it saturates at ca. 0.85 (grey dotted line shows  $z_{MAR} + z_{SAR}$ ).



## 1598 VI. Limitations and outlook

1599

1600 In this last section we list some potential limitations of an inherently simple scaling law, and  
1601 what approaches could be used to address those and improve genetic diversity loss  
1602 projections.

1603

### 1604 VI.1 Reasons for overestimations

1605

1606 Many researchers have posited that SAR likely overestimates species extinction (33, 67). For  
1607 instance:

- 1608 - Ignoring that a diversity-area relationship can be defined outwards, inwards, or  
1609 focusing on endemisms can have an impact (10, 33, 67). To address this, we  
1610 confirmed relative consistency between inward, outward, and random placement  
1611 MAR, and proposed that the EMAR may not be that appropriate to study genetic  
1612 diversity loss (or at least EMAR does not show predictability in our simulation).
- 1613 - Species may persist in altered habitats, like some animals are known to do (68). We  
1614 have focused some of the estimates in this study on plants, for which area loss should  
1615 equate to population loss and vice versa, but further extensions could be applied in the  
1616 future as described by Pereira and Daily (68).
- 1617 - SAR is not a mechanistic model (69). We have derived its ranges of possible values  
1618 and averages analytically and are beginning to understand how evolutionary forces  
1619 shape MAR. Realistic simulations can help understand in a process-based framework  
1620 how populations (and their MAR) react to partial population extinction (continuous  
1621 space simulations with progressive area reductions appear to fit well with the MAR  
1622 predictions calculated before the extinction process starts, section III.2.6).
- 1623 - There is a scale dependence in the SAR slope, with slight increase in the slope at large  
1624 scales (10). Since power laws are typically fit with large-scale datasets and used to  
1625 predict local scale extinctions, predictions could be overestimated at local scales.

1626

### 1627 VI.2 Reasons for underestimations

1628

1629 While the simplicity of power laws to make predictions of species extinction may lead to  
1630 overestimations, there are also important reasons to believe MAR would underestimate  
1631 genetic loss.

- 1632 - Perhaps even more so than in species list datasets and census, the discovery of low  
1633 frequency genetic variants is highly underpowered (70). These are highly prevalent,  
1634 but genomic pipelines, with the aim to be conservative, often filter out rare variants.  
1635 This would underestimate  $z_{MAR}$  and therefore the degree of genetic diversity loss with  
1636 area shrinkage. This is clear in the pre-selected-only marker dataset of *Pinus contorta*.
- 1637 - Related to the previous: Although sequencing methods have an error rate that  
1638 misreads true nucleotide sequences, this rate is typically extremely low (many  
1639 sequencing projects described here used Illumina HiSeq series, which has a 0.112%  
1640 error rate, or about 1 misread nucleotide in 1000). This could intuitively lead to  
1641 overestimates in mutations in space but in fact, the mis-reading of DNA ends up  
1642 causing an underestimation. This is because bioinformatic software that transforms  
1643 raw data into SNP variant tables errs towards the conservative direction, often not  
1644 calling mutations that have been observed very few times, and thus likely under-  
1645 representing rare mutations (71).
- 1646 - The use of scaled  $z_{MAR}$  proposed in section II.3.2. accounts for that the minimum  $z_{MAR}$   
1647 is rarely exactly 0, especially when sample sizes are limited. We use this correction

1648 scaling down  $z_{MAR}$  to be conservative. However,  $z_{MAR}$  could only in very exceptional  
1649 circumstances be 1, but we do not correct for this, again, to have a conservatively low  
1650  $z_{MAR}$ . Hence, our conservative approach would generally lead to underestimates of  
1651 genetic diversity loss.

- 1652 - When species shrink in area, the effective population size of the remaining population  
1653 decreases, increasing drift and moving towards a lower diversity equilibrium. This  
1654 reactive process is not captured by the phenomenological MAR relationship.
- 1655 - The nested extinction of species and genetic diversity loss (section V.3) would lead  
1656 us, by the right of “survival bias”, to underestimate how much genetic diversity has  
1657 been lost cumulative in an ecosystem.

1658

### 1659 **VI.3 Final notes**

1660

1661 Ultimately, to make accurate predictions of genetic diversity loss and increased extinction  
1662 risk of species, very detailed data and expert assessment per species will be required: census  
1663 sizes, genome size, migration in metapopulations, mating system, detailed maps of genetic  
1664 makeups, and finescale area transformations. This could enable mechanistic models projected  
1665 forward-in-time such as discussed in section II.3.6. The production of new genomic datasets  
1666 across entire ecosystems should further help create maps of genetic diversity at high  
1667 resolution to track losses (72–74).

1668

1669 Our philosophy in this work has been to err on the conservative side when projecting  
1670 genetic diversity loss (e.g. using area calculations that produce lower  $z_{MAR}$  values, scaling  
1671 them for low sample bias, using lower estimates of ecosystem transformation, etc.). However,  
1672 this conservative approach can also lead us into under-estimating loss. As described in V.4.,  
1673 the phenomenon of survival bias likely leads us to underestimate what has been lost given we  
1674 do not observe it. A phenomenon also highlighted as a possible explanation for the relatively  
1675 shy difference in genetic diversity between threatened and non-threatened species (75, 76)

1676

1677 Because to our knowledge, no other approaches exist to project genetic diversity, we  
1678 believe that MAR is a quantitative and scalable first-approximation of genetic diversity that  
1679 would just require accurate understanding of abundance or area reductions and minimal  
1680 information about population structure or mating/dispersal/range relationships. Given that  
1681 scaling relationships are already applied by conservation policy (77), and given that  
1682 assumptions and limitations are understood, we expect MAR to become a relevant tool to  
1683 project losses of a dimension of biodiversity so far mostly invisible or unaddressable in large  
1684 conservation projections.

1685

1686

1687

## 1688 SUPPLEMENTAL REFERENCES

1689  
1690

- 1691 1. S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography* (Monographs in  
1692 Population Biology, 2001).
- 1693 2. M. Tokeshi, Niche Apportionment or Random Assortment: Species Abundance Patterns  
1694 Revisited. *J. Anim. Ecol.* **59**, 1129–1146 (1990).
- 1695 3. M. Tokeshi, in *Advances in Ecological Research*, M. Begon, A. H. Fitter, Eds. (Academic Press,  
1696 1993; <https://www.sciencedirect.com/science/article/pii/S0065250408600422>), vol. 24, pp. 111–  
1697 186.
- 1698 4. MOTOMURA, I, A statistical treatment of ecological communities. *Zoological Magazine.* **44**,  
1699 379–383 (1932).
- 1700 5. R. H. MacArthur, On the relative abundance of bird species. *Proc. Natl. Acad. Sci. U. S. A.* **43**,  
1701 293–295 (1957).
- 1702 6. R. A. Fisher, A. S. Corbet, C. B. Williams, The Relation Between the Number of Species and the  
1703 Number of Individuals in a Random Sample of an Animal Population. *J. Anim. Ecol.* **12**, 42–58  
1704 (1943).
- 1705 7. F. W. Preston, The canonical distribution of commonness and rarity: Part I. *Ecology.* **43**, 185  
1706 (1962).
- 1707 8. F. W. Preston, The commonness, and rarity, of species. *Ecology.* **29**, 254–283 (1948).
- 1708 9. D. Alonso, A. J. McKane, Sampling Hubbell’s neutral theory of biodiversity. *Ecol. Lett.* **7**, 901–  
1709 910 (2004).
- 1710 10. D. Storch, P. Keil, W. Jetz, Universal species-area and endemics-area relationships at continental  
1711 scales. *Nature.* **488**, 78–81 (2012).
- 1712 11. S. L. Pimm, G. J. Russell, J. L. Gittleman, T. M. Brooks, The future of biodiversity. *Science.*  
1713 **269**, 347–350 (1995).
- 1714 12. C. D. Thomas, A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C. Collingham, B. F.  
1715 N. Erasmus, M. F. de Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A. S. van  
1716 Jaarsveld, G. F. Midgley, L. Miles, M. A. Ortega-Huerta, A. Townsend Peterson, O. L. Phillips,  
1717 S. E. Williams, Extinction risk from climate change. *Nature.* **427**, 145–148 (2004).
- 1718 13. J. F. C. Kingman, The coalescent. *Stochastic Process. Appl.* **13**, 235–248 (1982).
- 1719 14. R. C. Griffiths, S. Tavaré, The age of a mutation in a general coalescent tree. *Communications in*  
1720 *Statistics. Stochastic Models.* **14**, 273–295 (1998).
- 1721 15. R. A. Fisher, XVII.—The Distribution of Gene Ratios for Rare Mutations. *Proceedings of the*  
1722 *Royal Society of Edinburgh.* **50**, 204–219 (1931).
- 1723 16. M. W. Hahn, *Molecular Population Genetics* (Oxford University Press, 2018;  
1724 <https://play.google.com/store/books/details?id=3BDkswEACAAJ>).
- 1725 17. V. Buffalo, Quantifying the relationship between genetic diversity and population size suggests  
1726 natural selection cannot explain Lewontin’s paradox. *Elife.* **10** (2021), doi:10.7554/eLife.67509.

- 1727 18. D. R. Marshal, A. D. H. Brown, in *Crop genetic resources for today and tomorrow*, O. H.  
1728 Frankel, J. G. Hawkes, Eds. (Cambridge University Press, 1975;  
1729 <https://www.researchgate.net/publication/280057199>), vol. 2.
- 1730 19. S. Wright, Isolation by Distance. *Genetics*. **28**, 114–138 (1943).
- 1731 20. J. Novembre, M. Stephens, Interpreting principal component analyses of spatial population  
1732 genetic variation. *Nat. Genet.* **40**, 646–649 (2008).
- 1733 21. H. Fan 樊海英, Q. Zhang 张清臣, J. Rao 饶娟娟, J. Cao 曹静文, X. Lu 卢欣, Genetic Diversity-  
1734 Area Relationships across Bird Species. *Am. Nat.* **194**, 736–740 (2019).
- 1735 22. P. R. A. Campos, V. M. de Oliveira, A. Rosas, Epistasis and environmental heterogeneity in the  
1736 speciation process. *Ecol. Modell.* **221**, 2546–2554 (2010).
- 1737 23. J. Kelleher, A. M. Etheridge, G. McVean, Efficient Coalescent Simulation and Genealogical  
1738 Analysis for Large Sample Sizes. *PLoS Comput. Biol.* **12**, e1004842 (2016).
- 1739 24. B. C. Haller, P. W. Messer, SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher  
1740 Model. *Mol. Biol. Evol.* **36**, 632–637 (2019).
- 1741 25. T. R. Booker, S. Yeaman, M. C. Whitlock, The WZA: A window-based method for  
1742 characterizing genotype–environment association. *bioRxiv* (2021), p. 2021.06.25.449972.
- 1743 26. J. Kelleher, K. R. Thornton, J. Ashander, P. L. Ralph, Efficient pedigree recording for fast  
1744 population genetics simulation. *PLoS Comput. Biol.* (2018), p. e1006581.
- 1745 27. B. C. Haller, J. Galloway, J. Kelleher, P. W. Messer, P. L. Ralph, Tree-sequence recording in  
1746 SLiM opens new horizons for forward-time simulation of whole genomes. *Mol. Ecol. Resour.*  
1747 **19**, 552–566 (2019).
- 1748 28. R. R. Hudson, M. Slatkin, W. P. Maddison, Estimation of levels of gene flow from DNA  
1749 sequence data. *Genetics*. **132**, 583–589 (1992).
- 1750 29. 1001 Genomes Consortium, 1,135 Genomes Reveal the Global Pattern of Polymorphism in  
1751 *Arabidopsis thaliana*. *Cell*. **166**, 481–491 (2016).
- 1752 30. B. J. McGill, B. J. Enquist, E. Weiher, M. Westoby, Rebuilding community ecology from  
1753 functional traits. *Trends Ecol. Evol.* **21**, 178–185 (2006).
- 1754 31. P. I. Prado, M. Miranda, A. Chalom, *sads: R package for fitting species abundance distributions*  
1755 (Github, 2018; <https://github.com/piLaboratory/sads>).
- 1756 32. T. J. Matthews, K. A. Triantis, R. J. Whittaker, F. Guilhaumon, *sars: an R package for fitting,*  
1757 *evaluating and comparing species–area relationship models.* *Ecography*. **42**, 1446–1455 (2019).
- 1758 33. F. He, S. P. Hubbell, Species–area relationships always overestimate extinction rates from habitat  
1759 loss. *Nature*. **473**, 368–371 (2011).
- 1760 34. R. J. H. & J. van Etten, raster: Geographic analysis and modeling with raster data (2012),  
1761 (available at <http://CRAN.R-project.org/package=raster>).
- 1762 35. M. Exposito-Alonso, 500 Genomes Field Experiment Team, H. A. Burbano, O. Bossdorf, R.  
1763 Nielsen, D. Weigel, Natural selection in the *Arabidopsis thaliana* genome in present and future  
1764 climates. *Nature*. **573**, 126–129 (2019).

- 1765 36. Y. B. Simons, K. Bullaughey, R. R. Hudson, G. Sella, A population genetic interpretation of  
1766 GWAS findings for human quantitative traits. *PLoS Biol.* **16**, e2002985 (2018).
- 1767 37. M. Exposito-Alonso, C. Becker, V. J. Schuenemann, E. Reiter, C. Setzer, R. Slovak, B. Brachi, J.  
1768 Hagmann, D. G. Grimm, J. Chen, W. Busch, J. Bergelson, R. W. Ness, J. Krause, H. A. Burbano,  
1769 D. Weigel, The rate and potential relevance of new mutations in a colonizing plant lineage. *PLoS*  
1770 *Genet.* **14**, e1007155 (2018).
- 1771 38. H. Seebens, T. M. Blackburn, E. E. Dyer, P. Genovesi, P. E. Hulme, J. M. Jeschke, S. Pagad, P.  
1772 Pyšek, M. Winter, M. Arianoutsou, S. Bacher, B. Blasius, G. Brundu, C. Capinha, L. Celesti-  
1773 Grapow, W. Dawson, S. Dullinger, N. Fuentes, H. Jäger, J. Kartesz, M. Kenis, H. Kreft, I. Kühn,  
1774 B. Lenzner, A. Liebhold, A. Mosena, D. Moser, M. Nishino, D. Pearman, J. Pergl, W. Rabitsch,  
1775 J. Rojas-Sandoval, A. Roques, S. Rorke, S. Rossinelli, H. E. Roy, R. Scalera, S. Schindler, K.  
1776 Štajerová, B. Tokarska-Guzik, M. van Kleunen, K. Walker, P. Weigelt, T. Yamanaka, F. Essl,  
1777 No saturation in the accumulation of alien species worldwide. *Nat. Commun.* **8**, 14435 (2017).
- 1778 39. H. Seebens, F. Essl, W. Dawson, N. Fuentes, D. Moser, J. Pergl, P. Pyšek, M. van Kleunen, E.  
1779 Weber, M. Winter, B. Blasius, Global trade will accelerate plant invasions in emerging  
1780 economies under climate change. *Glob. Chang. Biol.* **21**, 4128–4140 (2015).
- 1781 40. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P.  
1782 Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: a tool set for whole-genome  
1783 association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- 1784 41. C. Mérot, R. A. Oomen, A. Tigano, M. Wellenreuther, A Roadmap for Understanding the  
1785 Evolutionary Significance of Structural Genomic Variation. *Trends Ecol. Evol.* **35**, 561–572  
1786 (2020).
- 1787 42. K. Lucek, Y. Willi, Drivers of linkage disequilibrium across a species' geographic range. *PLoS*  
1788 *Genet.* **17**, e1009477 (2021).
- 1789 43. J. M. Kreiner, D. A. Giacomini, F. Bemm, B. Waithaka, J. Regalado, C. Lanz, J. Hildebrandt, P.  
1790 H. Sikkema, P. J. Tranel, D. Weigel, J. R. Stinchcombe, S. I. Wright, Multiple modes of  
1791 convergent adaptation in the spread of glyphosate-resistant *Amaranthus tuberculatus*. *Proc. Natl.*  
1792 *Acad. Sci. U. S. A.* **116**, 21076–21084 (2019).
- 1793 44. M. A. Supple, J. G. Bragg, L. M. Broadhurst, A. B. Nicotra, M. Byrne, R. L. Andrew, A.  
1794 Widdup, N. C. Aitken, J. O. Borevitz, Landscape genomic prediction for restoration of a  
1795 *Eucalyptus* foundation species under climate change. *Elife.* **7** (2018), doi:10.7554/eLife.31835.
- 1796 45. M. Vallejo-Marín, J. Friedman, A. D. Twyford, O. Lepais, S. M. Ickert-Bond, M. A. Streisfeld,  
1797 L. Yant, M. van Kleunen, M. C. Rotter, J. R. Puzey, Population genomic and historical analysis  
1798 suggests a global invasion by bridgehead processes in *Mimulus guttatus*. *Commun Biol.* **4**, 327  
1799 (2021).
- 1800 46. J. T. Lovell, A. H. MacQueen, S. Mamidi, J. Bonnette, J. Jenkins, J. D. Napier, A. Sreedasyam,  
1801 A. Healey, A. Session, S. Shu, K. Barry, S. Bonos, L. Boston, C. Daum, S. Deshpande, A.  
1802 Ewing, P. P. Grabowski, T. Haque, M. Harrison, J. Jiang, D. Kudrna, A. Lipzen, T. H.  
1803 Pendergast 4th, C. Plott, P. Qi, C. A. Sasaki, E. V. Shakhov, D. Sims, M. Sharma, R. Sharma, A.  
1804 Stewart, V. R. Singan, Y. Tang, S. Thibivillier, J. Webber, X. Weng, M. Williams, G. A. Wu, Y.  
1805 Yoshinaga, M. Zane, L. Zhang, J. Zhang, K. D. Behrman, A. R. Boe, P. A. Fay, F. B. Fritschi, J.  
1806 D. Jastrow, J. Lloyd-Reilley, J. M. Martínez-Reyna, R. Matamala, R. B. Mitchell, F. M.  
1807 Rouquette Jr, P. Ronald, M. Saha, C. M. Tobias, M. Udvardi, R. A. Wing, Y. Wu, L. E. Bartley,  
1808 M. Casler, K. M. Devos, D. B. Lowry, D. S. Rokhsar, J. Grimwood, T. E. Juenger, J. Schmutz,  
1809 Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature* (2021),  
1810 doi:10.1038/s41586-020-03127-1.

- 1811 47. I. R. MacLachlan, T. K. McDonald, B. M. Lind, L. H. Rieseberg, S. Yeaman, S. N. Aitken,  
1812 Genome-wide shifts in climate-related variation underpin responses to selective breeding in a  
1813 widespread conifer. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021), doi:10.1073/pnas.2016900118.
- 1814 48. G. Tuskan, W. Muchero, J.-G. Chen, D. Jacobson, T. Tschaplinski, D. Rokhsar, W. Schackwitz,  
1815 J. Schmutz, S. DiFazio, Populus Trichocarpa Genome-Wide Association Study (GWAS)  
1816 Population SNP Dataset Released, (available at <https://doi.ccs.ornl.gov/ui/doi/55>).
- 1817 49. Anopheles gambiae 1000 Genomes Consortium, Data analysis group, Partner working group,  
1818 Sample collections—Angola:, Burkina Faso:, Cameroon:, Gabon:, Guinea:, Guinea-Bissau:,  
1819 Kenya:, Uganda:, Crosses:, Sequencing and data production, Web application development,  
1820 Project coordination, Genetic diversity of the African malaria vector Anopheles gambiae.  
1821 *Nature.* **552**, 96–100 (2017).
- 1822 50. Z. L. Fuller, V. J. L. Mocellin, L. A. Morris, N. Cantin, J. Shepherd, L. Sarre, J. Peng, Y. Liao, J.  
1823 Pickrell, P. Andolfatto, M. Matz, L. K. Bay, M. Przeworski, Population genetics of the coral  
1824 Acropora millepora: Toward genomic prediction of bleaching. *Science.* **369** (2020),  
1825 doi:10.1126/science.aba4674.
- 1826 51. K. Ruegg, E. C. Anderson, M. Somveille, R. A. Bay, M. Whitfield, E. H. Paxton, T. B. Smith,  
1827 Linking climate niches across seasons to assess population vulnerability in a migratory bird.  
1828 *Glob. Chang. Biol.* **27**, 3519–3531 (2021).
- 1829 52. R. A. Bay, R. J. Harrigan, V. Le Underwood, H. Lisle Gibbs, T. B. Smith, K. Ruegg, Genomic  
1830 signals of selection predict climate-driven population declines in a migratory bird. *Science.* **359**,  
1831 83–86 (2018).
- 1832 53. E. P. Kingsley, K. M. Kozak, S. P. Pfeifer, D.-S. Yang, H. E. Hoekstra, The ultimate and  
1833 proximate mechanisms driving the evolution of long tails in forest deer mice. *Evolution.* **71**, 261–  
1834 273 (2017).
- 1835 54. L. Smeds, J. Aspi, J. Berglund, I. Kojola, K. Tirronen, H. Ellegren, Whole-genome analyses  
1836 provide no evidence for dog introgression in Fennoscandian wolf populations. *Evol. Appl.* **14**,  
1837 721–734 (2021).
- 1838 55. R. M. Schweizer, J. Robinson, R. Harrigan, P. Silva, M. Galverni, M. Musiani, R. E. Green, J.  
1839 Novembre, R. K. Wayne, Targeted capture and resequencing of 1040 genes reveal  
1840 environmentally driven functional variation in grey wolves. *Mol. Ecol.* **25**, 357–379 (2016).
- 1841 56. 1000 Genomes Project Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M.  
1842 Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis, A global  
1843 reference for human genetic variation. *Nature.* **526**, 68–74 (2015).
- 1844 57. J. D. Palacio-Mejía, P. P. Grabowski, E. M. Ortiz, G. A. Silva-Arias, T. Haque, D. L. Des  
1845 Marais, J. Bonnette, D. B. Lowry, T. E. Juenger, Geographic patterns of genomic diversity and  
1846 structure in the C4 grass *Panicum hallii* across its natural distribution. *AoB Plants.* **13**, lab002  
1847 (2021).
- 1848 58. A. M. Royer, M. A. Streisfeld, C. I. Smith, Population genomics of divergence within an obligate  
1849 pollination mutualism: Selection maintains differences between Joshua tree species. *Am. J. Bot.*  
1850 **103**, 1730–1741 (2016).
- 1851 59. M. Kapun, J. C. B. Nunez, M. Bogaerts-Márquez, J. Murga-Moreno, M. Paris, J. Outten, M.  
1852 Coronado-Zamora, C. Tern, O. Rota-Stabelli, M. P. García Guerreiro, S. Casillas, D. J. Orengo,  
1853 E. Puerma, M. Kankare, L. Ometto, V. Loeschcke, B. S. Onder, J. K. Abbott, S. W. Schaeffer, S.  
1854 Rajpurohit, E. L. Behrman, M. F. Schou, T. J. S. Merritt, B. P. Lazzaro, A. Glaser-Schmitt, E.

- 1855 Argyridou, F. Staubach, Y. Wang, E. Tauber, S. V. Serga, D. K. Fabian, K. A. Dyer, C. W.  
1856 Wheat, J. Parsch, S. Grath, M. S. Veselinovic, M. Stamenkovic-Radak, M. Jelic, A. J. Buendía-  
1857 Ruíz, M. Josefa Gómez-Julián, M. Luisa Espinosa-Jimenez, F. D. Gallardo-Jiménez, A.  
1858 Patenkovic, K. Eric, M. Tanaskovic, A. Ullastres, L. Guio, M. Merenciano, S. Guirao-Rico, V.  
1859 Horváth, D. J. Obbard, E. Pasyukova, V. E. Alatortsev, C. P. Vieira, J. Vieira, J. Roberto Torres,  
1860 I. Kozeretska, O. M. Maistrenko, C. Montchamp-Moreau, D. V. Mukha, H. E. Machado, A.  
1861 Barbadilla, D. Petrov, P. Schmidt, J. Gonzalez, T. Flatt, A. O. Bergland, *Drosophila* Evolution  
1862 over Space and Time (DEST) - A New Population Genomics Resource. *bioRxiv* (2021), p.  
1863 2021.02.01.428994.
- 1864 60. L. N. Di Santo, S. Hoban, T. L. Parchman, J. W. Wright, J. A. Hamilton, Reduced representation  
1865 sequencing to understand the evolutionary history of Torrey pine (*Pinus torreyana* Parry) with  
1866 implications for rare species conservation. *bioRxiv* (2021), p. 2021.07.02.450939.
- 1867 61. J. von Seth, N. Dussex, D. Díez-Del-Molino, T. van der Valk, V. E. Kutschera, M. Kierczak, C.  
1868 C. Steiner, S. Liu, M. T. P. Gilbert, M.-H. S. Sinding, S. Prost, K. Guschanski, S. K. S. S.  
1869 Nathan, S. Brace, Y. L. Chan, C. W. Wheat, P. Skoglund, O. A. Ryder, B. Goossens, A.  
1870 Götherström, L. Dalén, Genomic insights into the conservation status of the world's last  
1871 remaining Sumatran rhinoceros populations. *Nat. Commun.* **12**, 2393 (2021).
- 1872 62. Millennium Ecosystem Assessment, *Millennium ecosystem assessment* (Millennium Ecosystem  
1873 Assessment, 2005; <http://chapter.ser.org/europe/files/2012/08/Harris.pdf>).
- 1874 63. G. C. Hurtt, L. Chini, R. Sahajpal, S. Frolking, B. L. Boudirsky, K. Calvin, J. C. Doelman, J. Fisk,  
1875 S. Fujimori, K. Klein Goldewijk, T. Hasegawa, P. Havlik, A. Heinemann, F. Humpenöder, J.  
1876 Jungclaus, J. O. Kaplan, J. Kennedy, T. Krisztin, D. Lawrence, P. Lawrence, L. Ma, O. Mertz, J.  
1877 Pongratz, A. Popp, B. Poulter, K. Riahi, E. Shevliakova, E. Stehfest, P. Thornton, F. N. Tubiello,  
1878 D. P. van Vuuren, X. Zhang, Harmonization of global land use change and management for the  
1879 period 850–2100 (LUH2) for CMIP6. *Geosci. Model Dev.* **13**, 5425–5464 (2020).
- 1880 64. S. Theodoridis, C. Rahbek, D. Nogues-Bravo, Exposure of mammal genetic diversity to mid-21st  
1881 century global change. *Ecography (Cop.)* (2021), doi:10.1111/ecog.05588.
- 1882 65. I. Overcast, M. Ruffley, J. Rosindell, L. Harmon, P. A. V. Borges, B. C. Emerson, R. S. Etienne,  
1883 R. Gillespie, H. Krehenwinkel, D. Luke Mahler, F. Massol, C. E. Parent, J. Patiño, B. Peter, B.  
1884 Week, C. Wagner, M. J. Hickerson, A. Rominger, A unified model of species abundance, genetic  
1885 diversity, and functional diversity reveals the mechanisms structuring ecological communities.  
1886 *bioRxiv* (2020), p. 2020.01.30.927236.
- 1887 66. S. Díaz, N. Zafra-Calvo, A. Purvis, P. H. Verburg, D. Obura, P. Leadley, R. Chaplin-Kramer, L.  
1888 De Meester, E. Dulloo, B. Martín-López, M. R. Shaw, P. Visconti, W. Broadgate, M. W.  
1889 Bruford, N. D. Burgess, J. Cavender-Bares, F. DeClerck, J. M. Fernández-Palacios, L. A.  
1890 Garibaldi, S. L. L. Hill, F. Isbell, C. K. Khoury, C. B. Krug, J. Liu, M. Maron, P. J. K.  
1891 McGowan, H. M. Pereira, V. Reyes-García, J. Rocha, C. Rondinini, L. Shannon, Y.-J. Shin, P.  
1892 V. R. Snelgrove, E. M. Spehn, B. Strassburg, S. M. Subramanian, J. J. Tewksbury, J. E. M.  
1893 Watson, A. E. Zanne, Set ambitious goals for biodiversity and sustainability. *Science.* **370**, 411–  
1894 413 (2020).
- 1895 67. C. Rahbek, R. K. Colwell, Biodiversity: Species loss revisited. *Nature.* **473** (2011), pp. 288–289.
- 1896 68. H. M. Pereira, G. C. Daily, Modeling biodiversity dynamics in countryside landscapes. *Ecology.*  
1897 **87**, 1877–1885 (2006).
- 1898 69. J. Harte, A. B. Smith, D. Storch, Biodiversity scales from plots to biomes with a universal  
1899 species-area curve. *Ecol. Lett.* **12**, 789–797 (2009).

- 1900 70. D. R. Lockwood, C. M. Richards, G. M. Volk, Probabilistic models for collecting genetic  
1901 diversity: Comparisons, caveats, and limitations. *Crop Sci.* **47**, 861–866 (2007).
- 1902 71. L. Czech, M. Exposito-Alonso, grenepipe: A flexible, scalable, and reproducible pipeline to  
1903 automate variant and frequency calling from sequence reads. *arXiv* (2021), (available at  
1904 <http://arxiv.org/abs/2103.15167>).
- 1905 72. A. Miraldo, S. Li, M. K. Borregaard, A. Flórez-Rodríguez, S. Gopalakrishnan, M. Rizvanovic, Z.  
1906 Wang, C. Rahbek, K. A. Marske, D. Nogués-Bravo, An Anthropocene map of genetic diversity.  
1907 *Science.* **353**, 1532–1535 (2016).
- 1908 73. D. H. Parks, T. Mankowski, S. Zangooei, M. S. Porter, D. G. Armanini, D. J. Baird, M. G. I.  
1909 Langille, R. G. Beiko, GenGIS 2: Geospatial Analysis of Traditional and Genetic Biodiversity,  
1910 with New Gradient Algorithms and an Extensible Plugin Framework. *PLoS One.* **8**, e69885  
1911 (2013).
- 1912 74. C. Li, D. Tian, B. Tang, X. Liu, X. Teng, W. Zhao, Z. Zhang, S. Song, Genome Variation Map: a  
1913 worldwide collection of genome variations across multiple species. *Nucleic Acids Res.* **49**,  
1914 D1186–D1191 (2021).
- 1915 75. M. Kardos, E. Armstrong, S. W. Fitzpatrick, S. Hauser, P. Hedrick, J. Miller, D. Tallmon, W.  
1916 Chris Funk, The crucial role of genome-wide genetic variation in conservation. *bioRxiv* (2021),  
1917 p. 2021.07.05.451163.
- 1918 76. D. Spielman, B. W. Brook, R. Frankham, Most species are not driven to extinction before  
1919 genetic factors impact them. *Proceedings of the National Academy of Sciences.* **101** (2004), pp.  
1920 15261–15264.
- 1921 77. Ipbes, *Global assessment report of the Intergovernmental Science-Policy Platform on*  
1922 *Biodiversity and Ecosystem Services* (IPBES Secretariat, Bonn, Germany, 2019;  
1923 <https://www.ipbes.net/news/ipbes-global-assessment-summary-policy-makers-pdf>).
- 1924