

# Supplemental Materials | Response to Shaw et al. (2025)

Moi Exposito-Alonso

Department of Integrative Biology, University of California Berkeley, California, USA  
Howard Hughes Medical Institute, University of California Berkeley, California, USA

Correspondance: moisesexpositoalonso@gmail.com

24 July, 2025

## List of Texts

Text S1: Subset dataset to robust and comparable genetic diversity metrics ( $\pi$ and S) . . . . .	3
Summary of numbers from the original dataset . . . . .	3
Summary of numbers from filtered dataset . . . . .	4
Re-calculate effect sizes . . . . .	5
Calculate a percentage diversity change . . . . .	6
Summary statistics of changes . . . . .	7
Text S2: Test the temporal shift of genetic diversity changes . . . . .	9
Calculate average change with different methods . . . . .	9
Variations of meta-regression . . . . .	9
Test per-study statistical differences: are there more or less positive trends? . . . . .	15
Test relationship with Red List . . . . .	18
Test differences with increasing time. . . . .	18
Test differences with increasing data size. . . . .	18
Test differences with decreasing noise. . . . .	18
Test differences with domesticated vs not species. . . . .	18
Text S3: Test within-species consistency. . . . .	23
Text S4: Interpretation of changes under an evolutionary population genetic model of species contraction. .	24
BIOLOGICAL SAMPLING – <i>null scenario</i> . . . . .	24
SUBSAMPLE Single population, immediate – <i>scenario 1a</i> . . . . .	28
TIME Single population, drift over time – <i>scenario 1b</i> . . . . .	30
SPACE. Spatial meta-populations, immediate – <i>scenario 2a</i> . . . . .	32
SPACE & TIME. Spatial meta-populations via WFmoments, short drift – <i>scenario 2b</i> . . . . .	34
Text S5: Interpret increase in genetic diversity . . . . .	36
SINGLE POPULATION GROWTH OVER TIME – <i>gain scenario 1b</i> . . . . .	36
SPATIAL METAPOPULATIONS MIXING – <i>gain scenario 2a</i> . . . . .	37
Text S6: Simple population contraction averages from LPI . . . . .	37
Text S7: Limitations, caveats, and thoughts forward . . . . .	39
Thoughts on data limitations . . . . .	39
Thoughts on inverse model limitations . . . . .	39

## List of Figures

1     Funnel plot of effect size vs error shows a classic relationship of smaller effect size in measurements with less error. . . . .	5
--	---

2	Distribution of effect sizes (g) and raw genetic diversity changes (%). . . . .	8
3	P-value distribution in QQ plot. . . . .	16
4	Relationship between raw values of diversity change $\pi$ and Hedge's g significance. . . . .	17
5	Relationship between raw values of diversity change $\pi$ and Hedge's g values. . . . .	17
6	Significant decline trends and Red List categories. . . . .	18
7	Relationship between genetic changes and length of time interval. . . . .	19
8	Sample size (loci and individuals) and raw changes in $\pi\%$ genetic diversity. . . . .	20
9	Sample numbers across past and present studies. . . . .	20
10	Sample and marker numbers across studies . . . . .	21
11	Average diversity change of different data sizes. . . . .	21
12	Relationship between raw values, significance and noise. . . . .	22
13	Cases of species with multiple observations. . . . .	23
14	Positive and negative cases of species with multiple observations. . . . .	23
15	Ranges of expected changes of diversity $\pi\%$ with biological sampling . . . . .	25
16	Expected maximum changes of diversity $\pi\%$ with biological sampling . . . . .	26
17	Expected noise in percent diversity change $\pi\%$ with different sample sizes . . . . .	27
18	Expected immediate population bottleneck given observed $\pi\%$ with noise . . . . .	28
19	Expected immediate population bottleneck given observed M%. . . . .	29
20	Expected population bottleneck and drift given observed $\pi\%$ . . . . .	30
21	Expected immediate population bottleneck given observed M%. . . . .	31
22	Expected immediate geographic area loss given observed $\pi\%$ . . . . .	32
23	Expected immediate geographic area loss given observed M%. . . . .	33
24	Expected immediate geographic area loss given observed $\pi\%$ . . . . .	34
25	Expected population growth given increases in $\pi\%$ . . . . .	36
26	The shape of genetic diversity increases is slower than decreases. . . . .	36
27	Expected $F_{ST}$ to cause a inflation of $\pi\%$ . . . . .	37
28	Expected noisy inflation of $\pi\%$ from area loss and fragmentation. . . . .	38

## List of Tables

1	Number of observations by diversity type . . . . .	3
2	Shaw effect size meta-regression model comparison. . . . .	10
3	Shaw meta-regression summary. . . . .	11
4	Shaw meta-regression summary excluding $\pi$ and M. . . . .	11
5	Percentage diversity $\pi$ meta-regression model comparison. . . . .	11
6	Percentage diversity $\pi$ (delta SE) meta-regression model comparison. . . . .	12
7	Log ratio diversity $\pi$ meta-regression model comparison. . . . .	13
8	Percentage allelic richness M meta-regression model comparison. . . . .	14
9	Log ratio diversity M meta-regression model comparison. . . . .	15

## Text S1: Subset dataset to robust and comparable genetic diversity metrics ( $\pi$ and S).

### Summary of numbers from the original dataset

Shaw et al (2025) include a number of metrics of genetic composition in their analyses. The definition (from Authors) is below:

AR (allelic richness)  
A (mean alleles)  
TA (total alleles)  
pA (total private alleles)  
NPL (number of polymorphic loci)  
Other1 (other metric associated with variant counts)  
He (expected heterozygosity)  
pi (nucleotide diversity)  
h (haplotype diversity)  
H (Shannon diversity index)  
PIC (polymorphic information content)  
NEA (number of effective alleles)  
Fis (inbreeding coefficient)  
R (mean relatedness/kinship among individuals)  
apFst (among-population fixation index)  
Freq (frequency of an allele of interest)  
NPD (mean individual nucleotide p-distance)  
BS (bandsharing score)  
Other2 (other metric associated with Variant frequencies)  
Ho (observed heterozygosity)  
SH (standardised multilocus heterozygosity)  
Ai (mean alleles per individual)  
F (inbreeding coefficient at the individual level)  
NH (number of polymorphic loci per individual)  
Other3 (other metric associated with Individual-level diversity)  
Ne (effective population size estimated from molecular data)  
Nd (effective population size estimated from demographic data)  
Nb (effective number of breeders),  
Nf (female effective population size),  
Nc (effective population size estimated from a population census, and calculated based on an assumption about the Ne:Nc ratio),  
Other4 (other metric associated with integrated statistics).

These were further grouped in four categories (definitions from Authors):

Group1 = metrics based on variant counts (AR, A, TA, pA, NPL, Other1)  
Group2 = metrics based on evenness of variant frequencies (He, pi, h, H, PIC, NEA, Fis, R, apFst, Freq, NPD, BS, Other2)  
Group3 = metrics based on population means of individual-level diversity (Ho, SH, Ai, F, NH, Other3)  
Group4 = integrated statistics (Ne, Nd, Nb, Nf, Nc, Other4).

In this reanalyses, the dataset is filtered to only standard metrics of diversity:

Table S 1: Number of observations by diversity type

Diversity type	# observations
He	1177
Ho	872

Diversity type	# observations
AR	539
A	460
Ne	347
pi	153
h	139
Other2	87
Nb	71
Fis	53
F	19
apFst	17
Freq	16
H	14
Other1	12
TA	8
R	7
NPL	6
Nd	5
Other3	5
PIC	4
NEA	3
Ai	2
Nf	2
pA	2
Nc	1
NPD	1
Other4	1

- We then focus on the one hand on average difference metrics,  $\pi$  diversity, He and  $\pi$  diversity. There are a total of 1177 and 153 studies with He and Pi measurements. We decided to combine them so there is sufficient sample size for downstream analyses (for biallelic loci, expected heterozygosity and  $\pi$  are the same value). We also focus on the number of alleles, 999, for which we also have population genetic expectations. Together, these are 57.8921203% of original dataset.
- We also focus on studies with metrics in 2 time points, rather than linear trajectories or coalescent, which makes comparisons most straightforward. The number of 2 timepoint measurements is: 3803, linear: 64, and coalescent: 156. Coalescent is 3.8777032% of the original dataset.
- The number of observations by Kingdom and Phylum mostly show observations include insects, vertebrates, and flowering plants. We do not do any filtering.
- Number of observations of domesticated species:  
Domesticated 767 observations (% 19.0653741). Domesticated/pathogens/pests 3256, 767 observations (% 80.9346259, 19.0653741). We removed domesticated species.
- Number of nuclear vs other genomes:  
Nuclear genome markers 3658 observations (% 90.9271688). We focus on nuclear, as chloroplast and mitochondria undergo different evolutionary dynamics (maternal inheritance, no recombination during meiosis). There are 3658 nuclear, 282 mitochondrial, and 13 mixed genomes. Non-nuclear is 9.0728312% of original dataset.

### Summary of numbers from filtered dataset

Total number 1701 of which 54% are  $\pi$ . The final subset is 42% of the original dataset.

Summaries of allelic richness  $\pi$

- 916 studies
- 323 species
- Summary of number of generations in the final dataset is: Mean=6.5 generations, Median=2 generations, IQR= 0.59
- 6 generations.

### Summaries of allelic richness $M$

- 785 studies
- 283 species
- Summary of number of generations in the final dataset is: Mean=6.8 generations, Median=2 generations, IQR= 0.62
- 6.9 generations.

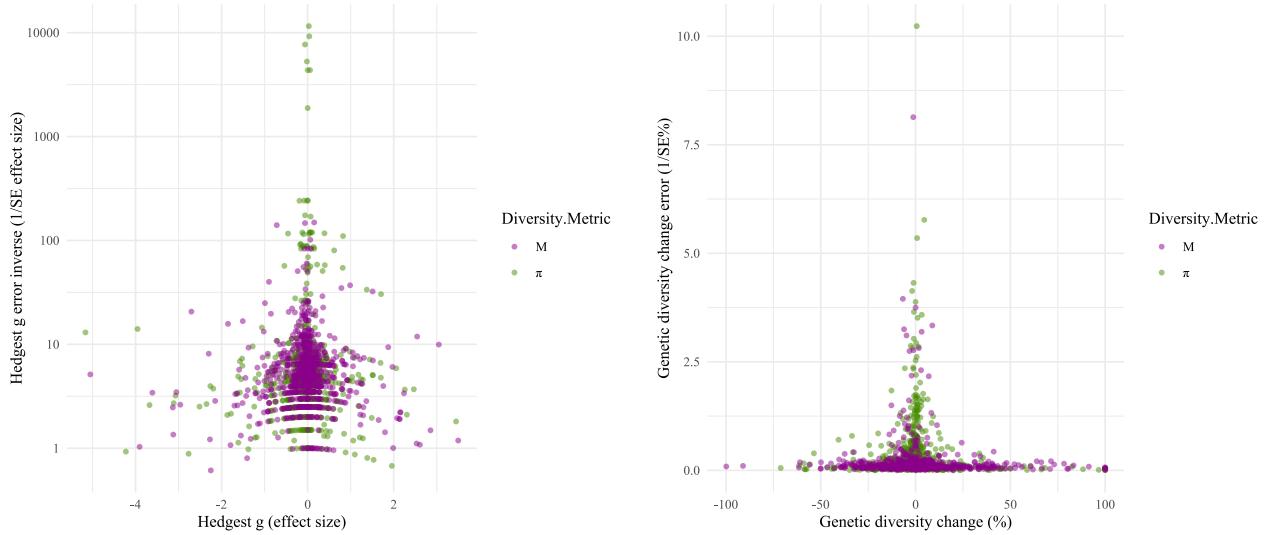


Figure S 1: Funnel plot of effect size vs error shows a classic relationship of smaller effect size in measurements with less error.

Figure S 1 shows the typical funnel shape whereby most precise estimates (high 1/SE) show smaller effects.

### Re-calculate effect sizes

To reproduce Shaw calculation of effect size:

$$g = J \cdot \frac{\pi_1 - \pi_0}{s_{g\text{-pooled}}}$$

To calculate the corrected version we have the degrees of freedom used for effect size correction is:  $df = n_{\text{early}} + n_{\text{recent}} - 2$ . We used Hedges' correction factor  $J$  for small-sample bias:  $J = 1 - \frac{3}{4 \cdot df - 1}$ .

Its standard error was estimated as:

$$SE_g = \sqrt{\frac{n_{\text{early}} + n_{\text{recent}}}{n_{\text{early}} \cdot n_{\text{recent}}} + \frac{g^2}{2 \cdot (n_{\text{early}} + n_{\text{recent}})}}$$

## Calculate a percentage diversity change

**Percentage change** The fraction (or percentage if multiplied  $\times 100$ ) of diversity change is defined as:

$$X_\pi = \frac{\pi_{present} - \pi_{past}}{\pi_{past}}$$

Each estimate came with sample size and some form of error extracted during the review process for both past (early) and present (recent) timepoints. The errors reported came in several forms: standard deviation (SD), standard error (SE), or the width of a 95% confidence interval (CI). All error types were converted to standard errors using the following rules (diversity  $\pi$  and  $M$  had mostly errors reported as SD or SE, only a few 95%CI and only 6 observations of  $M$  with IQR, discarded):

$$SE = \begin{cases} \frac{SD}{\sqrt{n}} & \text{if error type is SD} \\ SE & \text{if already reported as SE} \\ \frac{CI \text{ width}/2}{1.96} & \text{if reported as 95\% CI} \end{cases}$$

where  $n$  is the sample size. From this standard deviation (SD) was recomputed for completeness, as sample sizes were known for all observations as  $SD = SE \times \sqrt{n}$ . This was done separately for early and recent time points. Then pooled standard deviation was computed as:

$$SD_{pooled} = \sqrt{\frac{(n_{\text{early}} - 1) \cdot SD_{\text{early}}^2 + (n_{\text{recent}} - 1) \cdot SD_{\text{recent}}^2}{n_{\text{early}} + n_{\text{recent}} - 2}}$$

Just to have something informative of both studies, the effective sample size was computed as the harmonic mean of early and recent sample sizes:  $n_{\text{eff}} = \frac{2 \cdot n_{\text{early}} \cdot n_{\text{recent}}}{n_{\text{early}} + n_{\text{recent}}}$ .

From this, we aim to compute a error of the measured change ( $X_\pi = 100 \times \frac{\pi_{present} - \pi_{past}}{\pi_{past}} = \frac{\Delta_\pi}{\pi_{past}}$ ) The standard error of the change is computed assuming errors in the two time points are independent and estimated the standard error of the difference:  $SE_\Delta = \sqrt{SE_{\text{early}}^2 + SE_{\text{recent}}^2}$ , and scaled to be percentage change error:  $SE\%_\Delta = \left(\frac{SE_\Delta}{\pi_0}\right) \times 100$ .

This standard error is not exact because of the division of  $\pi_0$ , which itself contains noise. We can attempt to create a second standard error uses the delta method. While both are approximations, the delta method attempts to provide a more sophisticated approximation. Consider  $f(X, Y) = (\frac{X}{Y} - 1) \times 100$ , where  $X = \pi$  and  $Y = \pi_0$ . We apply the **delta method** to approximate the variance of  $f(X, Y)$ . The first-order Taylor expansion is:

$$\text{Var}[f(X, Y)] \approx \left(\frac{\partial f}{\partial X}\right)^2 \cdot \text{Var}[X] + \left(\frac{\partial f}{\partial Y}\right)^2 \cdot \text{Var}[Y] + 2 \cdot \frac{\partial f}{\partial X} \cdot \frac{\partial f}{\partial Y} \cdot \text{Cov}[X, Y]$$

Assuming independence between  $X$  and  $Y$  (i.e.,  $\text{Cov}[X, Y] = 0$ ), the partial derivatives are:

$$\frac{\partial f}{\partial X} = \frac{100}{Y}, \quad \frac{\partial f}{\partial Y} = -\frac{100X}{Y^2}$$

Substituting into the variance expression:

$$\text{Var}[\% \Delta \pi] \approx \left(\frac{100}{Y}\right)^2 \cdot \text{Var}[X] + \left(\frac{100X}{Y^2}\right)^2 \cdot \text{Var}[Y]$$

Thus, the **standard error** of the percent change is:

$$\text{SE}[\% \Delta \pi] \approx \sqrt{\left(\frac{100}{Y}\right)^2 \cdot \text{SE}_X^2 + \left(\frac{100X}{Y^2}\right)^2 \cdot \text{SE}_Y^2}$$

**Log ratio change** Although likely an arithmetic percentage change is the most understandable metric, for completeness, I also calculate the ratio of change:

$$LR_\pi = \log\left(\frac{\pi_{present}}{\pi_{past}}\right)$$

When averaging the log ratio we get a geometric (rather than arithmetic) mean, which may be helpful.

We can also calculate the standard error of the log ratio using the delta method. Let:  $X = \pi$  (recent diversity) and  $Y = \pi_0$  (early diversity), and  $f(X, Y) = \log\left(\frac{X}{Y}\right) = \log(X) - \log(Y)$ . We aim to compute the standard error of the log-ratio,  $\text{SE}[\log\left(\frac{X}{Y}\right)]$ , using the **delta method**, assuming that  $X$  and  $Y$  are independent.

The variance of  $f(X, Y)$  is approximated as:

$$\text{Var}[f(X, Y)] \approx \left(\frac{\partial f}{\partial X}\right)^2 \cdot \text{Var}[X] + \left(\frac{\partial f}{\partial Y}\right)^2 \cdot \text{Var}[Y]$$

The partial derivatives are:

$$\frac{\partial f}{\partial X} = \frac{1}{X}, \quad \frac{\partial f}{\partial Y} = -\frac{1}{Y}$$

Substituting into the variance approximation:

$$\text{Var}\left[\log\left(\frac{X}{Y}\right)\right] \approx \left(\frac{1}{X}\right)^2 \cdot \text{SE}_X^2 + \left(\frac{1}{Y}\right)^2 \cdot \text{SE}_Y^2$$

Therefore, the standard error of the log-ratio is:

$$\text{SE}\left[\log\left(\frac{X}{Y}\right)\right] = \sqrt{\left(\frac{\text{SE}_X}{X}\right)^2 + \left(\frac{\text{SE}_Y}{Y}\right)^2}$$

### Summary statistics of changes

Summaries of  $\pi$  change:

Median [IQR] Hedges g = -0.026 [-0.17 - 0.1]  
 Median [IQR] Error of Hedges g = 0.22 [0.15 - 0.33]  
 Median [IQR] Percentage change = -0.78% [-5 - 2.8%]  
 Median [IQR] SE of Percentage change = 7.4% [3.5 - 16%]  
 Mean Percentage change = 0.99%  
 Mean SE of Percentage change = 14%

Summaries of  $M$  change:

Median [IQR] Hedges g = -0.0071 [-0.23 - 0.16]  
 Median [IQR] Error of Hedges g = 0.2 [0.13 - 0.3]  
 Median [IQR] Percentage change = -0.44% [-7.9 - 6.8%]  
 Median [IQR] SE of Percentage change = 10% [6.5 - 16%]

Mean Percentage change= 1.5%

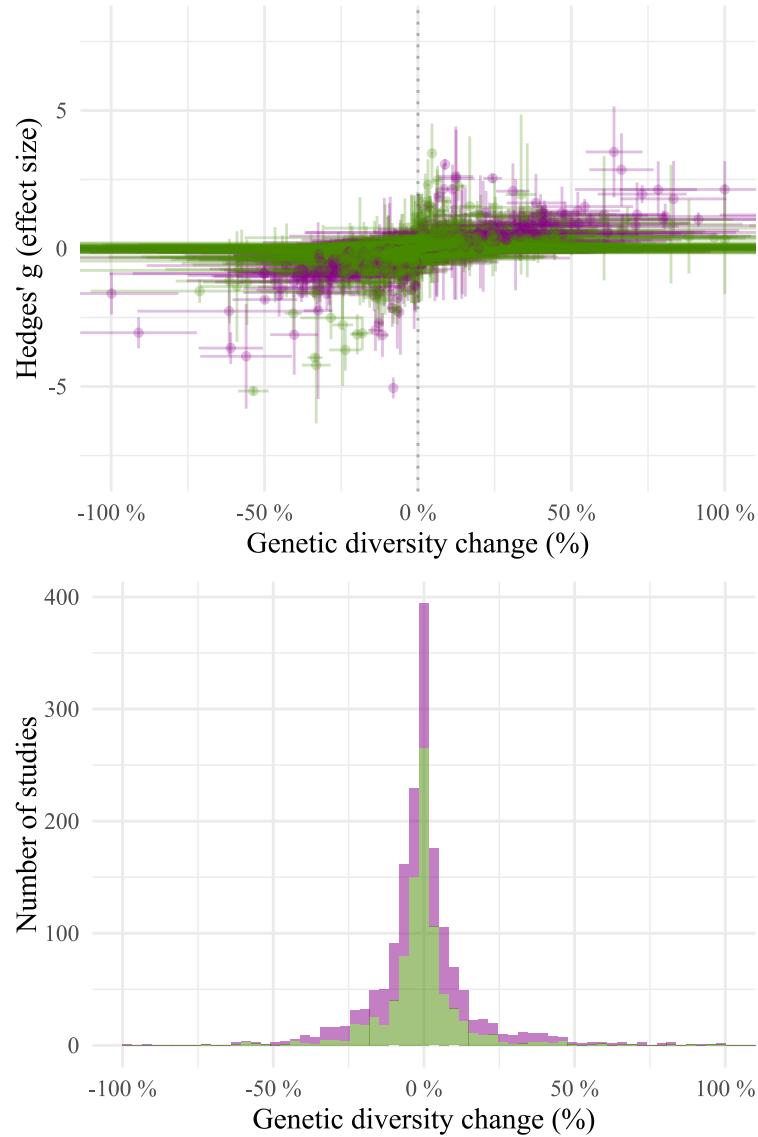


Figure S 2: Distribution of effect sizes (g) and raw genetic diversity changes (%).

Hedges' g and genetic diversity change in percentage are correlated since raw data is used to calculate g (Figure S2), although not perfectly (Pearson's  $r=0.53$ ). The extremest values in  $\pi$ , which comprises the majority of the data however, typically do not coincide. We can see that of the 5% top  $\pi$  values that decline based on Hedge's g (46), only a small fraction are also among the top raw % values of decline (43%). Same for  $M$  (50%). Figure S5 shows the raw average and estimated effect sizes in the comparison of Hedges' g and raw values of change (colors indicate significance, see section below).

## Text S2: Test the temporal shift of genetic diversity changes

### Calculate average change with different methods

For diversity  $\pi$

Arithmetic mean Hedge's  $g = -0.057 \pm 0.038$  (95% CI).

Weighted SE mean Hedge's  $g = 0.01 \pm 0.00011$  (95% CI) ( $z = 194, P = 0$ ).

MCMCglmm mean (intercept) of Hedge's  $g = -0.05 \pm 0.03$  (95% HPI).

Arithmetic mean  $\% \pi = 0.99\% \pm 1.9\%$  (95% CI).

Weighted SE mean  $\% \pi = 0.16\% \pm 0.089\%$  (95% CI) ( $z = 3.6, P = 3.3515656 \times 10^{-4}$ ).

MCMCglmm mean (intercept) of  $\% \pi = -1.20 \pm 0.73$  (95% HPI) .

Geometric mean  $\% \pi = -1.2\% \pm 1.2\%$  (95% CI).

Weighted SE geometric mean  $\% \pi = 0.25\% \pm 0.089\%$  (95% CI).

For allelic richness  $M$

Average Hedge's  $g = -0.044 \pm 0.049$  (95% CI).

Weighted SE mean Hedges'  $g = -0.087 \pm 0.005$  (95% CI) ( $z = -34, P = 4.3656128 \times 10^{-258}$ ).

MCMCglmm mean (intercept) of Hedges'  $g = -0.04 \pm 0.04$  (95% HPI).

Arithmetic mean  $\% M = 1.5\% \pm 1.9\%$  (95% CI).

Weighted SE mean  $\% M = -1.1 \pm 0.13$  (95% CI) ( $z = -16, P = 1.0272574 \times 10^{-55}$ ).

MCMCglmm mean (intercept) of  $\% M = 0.46 \pm 1.57$  (95% HPI).

Geometric mean  $\% \pi = -1.8\% \pm 2.3\%$  (95% CI).

Weighted SE geometric mean  $\% \pi = -0.86\% \pm 0.13\%$  (95% CI).

### Variations of meta-regression

To study the robustness of the original MCMCglmm meta-regression from Shaw, we conduct several variations of their function. From their supplement, they run MCMCglmm(  $g \sim Z.year.Midpoint + Z.Gen.interval$ , random= ~PaperID) using weak priors. Following Shaw, I used a weak inverse-gamma model prior ( $V=1, nu=0.002$ ) Aka. "pweak".

I generated variations over this model to understand robustness and explain why the results may differ from an arithmetic or error-weighted mean (models were named numerically with increasing fixed or random terms, "m1"..."m6"):

Changing fixed effects. I first changed the mean-centered generation interval fixed effect (shaw1 model). The mean-centering is necessary so that the intercept corresponds with the average fixed effect variable. However, the generation interval was not normally distributed. In fact, >75% of the dataset showed lower number of generations (quantile 75% = 6 generations) than the mean (6.5 generations). This is due to a long tail effect. A approach less driven by long tails is including a fixed effect of generation interval median-centered. When conducting that, the effect typically becomes much weaker or non-significant (this corrected model was named shaw2).

Changing random effects. Shaw includes Paper ID, and I tested the inclusion of Kingdom, species, and Marker type, to account for other possible heterogeneity in the data.

Changing priors. Although a weak prior seems sensible, in Bayesian statistics is important to test model robustness based on priors, especially when signals appear weak. I tested not including any prior (using default MCMCglmm,  $nu=0, V=1$ , aka. "none"). I tested a more regularizing prior "preg" ( $V=1e-4, nu=10$ ), and a prior that would limit unexplained variance to focus on the observation measurement error term ( $1/SE^2$ ), aka "pfix" ( $V=1e-4, nu=100$ ).

The last one is closer to fixed effect meta-analysis. These priors represent a range from low to strong shrinking, from weak information but unstable estimates to stronger information in model from standard error measurements.

The same set of models was fitted to the entire dataset (using effect sizes) as well as standard average diversity  $\pi$  and allelic richness  $M$  both effect sizes and percentage change. Models in reporting tables are ranked by DIC from MCMCglmm.

Several top models in both  $\pi$  and  $M$  percentage changes included non-significant, positive and negatively significant trends.

### All Shaw data effect size

Table S 2: Shaw effect size meta-regression model comparison.

label	DIC	summary	fixed effects	random effects
m5_none	8874.6	-0.03 ± 0.28 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m5_pweak	8873.8	-0.03 ± 0.27 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
mshaw2_pfix	10080.5	-0.05 ± 0.03 *	M.Year.Midpoint + M.Gen.interval	PaperID
m5_pfix	10087.2	-0.05 ± 0.03 *	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m3_preg	10110.6	-0.06 ± 0.03 *	M.Gen.interval	none
m3_pfix	10087.0	-0.06 ± 0.03 *	M.Gen.interval	none
m4_pfix	10087.5	-0.06 ± 0.03 *	M.Gen.interval	PaperID
m3_none	10112.9	-0.06 ± 0.03 *	M.Gen.interval	none
m3_pweak	10113.3	-0.06 ± 0.03 *	M.Gen.interval	none
m2_pfix	10086.5	-0.06 ± 0.03 *	none	PaperID
m6_pfix	10086.2	-0.06 ± 0.03 *	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
m1_pweak	10112.1	-0.06 ± 0.03 *	none	none
mshaw1_pfix	10080.6	-0.06 ± 0.03 *	Z.Year.Midpoint + Z.Gen.interval	PaperID
m1_pfix	10087.1	-0.06 ± 0.03 *	none	none
m1_preg	10109.5	-0.06 ± 0.03 *	none	none
m1_none	10112.6	-0.06 ± 0.03 *	none	none
mshaw2_preg	8877.4	-0.11 ± 0.05 *	M.Year.Midpoint + M.Gen.interval	PaperID
mshaw2_pweak	8878.5	-0.11 ± 0.05 *	M.Year.Midpoint + M.Gen.interval	PaperID
mshaw2_none	8878.0	-0.11 ± 0.05 *	M.Year.Midpoint + M.Gen.interval	PaperID
mshaw1_preg	8876.9	-0.11 ± 0.05 *	Z.Year.Midpoint + Z.Gen.interval	PaperID
mshaw1_pweak	8878.4	-0.11 ± 0.05 *	Z.Year.Midpoint + Z.Gen.interval	PaperID
mshaw1_none	8878.4	-0.11 ± 0.05 *	Z.Year.Midpoint + Z.Gen.interval	PaperID
m5_preg	8872.7	-0.11 ± 0.05 *	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m4_preg	8873.2	-0.12 ± 0.05 *	M.Gen.interval	PaperID
m4_none	8874.2	-0.12 ± 0.05 *	M.Gen.interval	PaperID
m4_pweak	8874.3	-0.12 ± 0.05 *	M.Gen.interval	PaperID
m2_preg	8871.8	-0.12 ± 0.05 *	none	PaperID
m2_none	8873.9	-0.12 ± 0.05 *	none	PaperID
m2_pweak	8873.5	-0.12 ± 0.05 *	none	PaperID
m6_none	8782.2	-0.25 ± 0.46 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
m6_pweak	8780.4	-0.25 ± 0.47 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
m6_preg	8804.7	-0.30 ± 0.18 *	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type

The exact model from Shaw et al. (“mshaw\_pweak”) recapitulates the found  $g=-0.11$  reported in the paper. We note that the fixed effect generation interval is non-significant, indicating that with increasing generation interval there is a positive trend in the effect size.

Table S 3: Shaw meta-regression summary.

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC
(Intercept)	-0.1121301	-0.1649134	-0.0621774	1166.092	0.001
Z.Gen.interval	0.0344938	-0.0388307	0.1004779	1000.000	0.346
Z.Year.Midpoint	0.0948754	0.0216149	0.1693552	1000.000	0.018

Exclusion of the key diversity variables makes the Shaw model non-significant.

Table S 4: Shaw meta-regression summary excluding  $\pi$  and M.

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC
(Intercept)	-0.0578699	-0.1316435	0.0082411	1000	0.106
Z.Gen.interval	0.0357412	-0.0396524	0.1116065	1000	0.364
Z.Year.Midpoint	0.1116704	0.0394472	0.2072596	1000	0.012

#### Average genetic diversity $\pi$ percent change

Table S 5: Percentage diversity  $\pi$  meta-regression model comparison.

label	DIC	summary	fixed effects	random effects
mshaw1_pfix	-4852.1	$0.26 \pm 0.32$ ns	Z.Year.Midpoint + Z.Gen.interval	PaperID
mshaw2_preg	-4856.6	$0.24 \pm 1.24$ ns	M.Year.Midpoint + M.Gen.interval	PaperID
mshaw2_pfix	-4852.0	$0.21 \pm 0.10$ *	M.Year.Midpoint + M.Gen.interval	PaperID
m2_pfix	-4853.1	$0.16 \pm 0.09$ *	none	PaperID
m1_pfix	-4852.7	$0.16 \pm 0.09$ *	none	none
m3_pfix	-4853.6	$0.13 \pm 0.09$ *	M.Gen.interval	none
m4_pfix	-4851.1	$0.13 \pm 0.09$ *	M.Gen.interval	PaperID
m5_pfix	-4854.0	$0.13 \pm 0.09$ *	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m6_pfix	-4852.2	$0.13 \pm 0.09$ *	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
mshaw2_none	6371.4	$-0.78 \pm 1.08$ ns	M.Year.Midpoint + M.Gen.interval	PaperID
mshaw2_pweak	6371.6	$-0.78 \pm 1.11$ ns	M.Year.Midpoint + M.Gen.interval	PaperID
m6_preg	6778.0	$-1.05 \pm 0.72$ *	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
m3_pweak	6794.5	$-1.05 \pm 0.74$ *	M.Gen.interval	none
m3_none	6795.0	$-1.05 \pm 0.74$ *	M.Gen.interval	none
m4_preg	6778.3	$-1.05 \pm 0.74$ *	M.Gen.interval	PaperID
m3_preg	6778.6	$-1.06 \pm 0.72$ *	M.Gen.interval	none
m5_preg	6010.0	$-1.11 \pm 0.91$ *	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m1_none	6797.9	$-1.19 \pm 0.72$ *	none	none
m1_pweak	6798.8	$-1.19 \pm 0.72$ *	none	none
m2_preg	6781.8	$-1.19 \pm 0.71$ *	none	PaperID
m1_preg	6782.3	$-1.20 \pm 0.71$ *	none	none
m4_none	6461.3	$-1.27 \pm 1.06$ *	M.Gen.interval	PaperID

Table S 5: Percentage diversity  $\pi$  meta-regression model comparison. (*continued*)

label	DIC	summary	fixed effects	random effects
m4_pweak	6462.4	-1.28 ± 1.08 *	M.Gen.interval	PaperID
m2_pweak	6460.5	-1.40 ± 1.03 *	none	PaperID
m2_none	6460.4	-1.40 ± 1.02 *	none	PaperID
m5_pweak	6460.8	-1.93 ± 4.75 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m5_none	6462.3	-1.95 ± 4.53 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
mshaw1_preg	6765.2	-3.12 ± 1.52 *	Z.Year.Midpoint + Z.Gen.interval	PaperID
mshaw1_none	6371.1	-4.23 ± 1.95 *	Z.Year.Midpoint + Z.Gen.interval	PaperID
mshaw1_pweak	6371.4	-4.23 ± 1.96 *	Z.Year.Midpoint + Z.Gen.interval	PaperID
m6_none	6422.7	-4.47 ± 8.30 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
m6_pweak	6423.2	-4.54 ± 8.45 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type

#### Average genetic diversity $\pi$ percent change (delta method for SE)

Table S 6: Percentage diversity  $\pi$  (delta SE) meta-regression model comparison.

label	DIC	summary	fixed effects	random effects
mshaw2_pfix	-4848.3	0.05 ± 0.10 ns	M.Year.Midpoint + M.Gen.interval	PaperID
m2_pfix	-4835.7	-0.06 ± 0.09 ns	none	PaperID
m1_pfix	-4868.3	-0.06 ± 0.09 ns	none	none
m3_pfix	-4832.5	-0.06 ± 0.09 ns	M.Gen.interval	none
m6_pfix	-4838.4	-0.06 ± 0.10 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
m5_pfix	-4847.1	-0.06 ± 0.09 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m4_pfix	-4850.0	-0.06 ± 0.09 ns	M.Gen.interval	PaperID
mshaw1_pfix	-4836.6	-0.44 ± 0.33 *	Z.Year.Midpoint + Z.Gen.interval	PaperID
mshaw2_preg	-4807.6	-0.52 ± 1.12 ns	M.Year.Midpoint + M.Gen.interval	PaperID
mshaw2_pweak	6376.6	-1.47 ± 0.96 *	M.Year.Midpoint + M.Gen.interval	PaperID
mshaw2_none	6371.0	-1.48 ± 0.97 *	M.Year.Midpoint + M.Gen.interval	PaperID
m4_preg	6777.3	-1.60 ± 0.70 *	M.Gen.interval	PaperID
m3_none	6791.4	-1.60 ± 0.72 *	M.Gen.interval	none
m3_pweak	6790.6	-1.62 ± 0.69 *	M.Gen.interval	none
m3_preg	6777.8	-1.64 ± 0.73 *	M.Gen.interval	none
m1_pweak	6795.3	-1.80 ± 0.71 *	none	none
m1_none	6795.8	-1.82 ± 0.71 *	none	none
m1_preg	6785.0	-1.82 ± 0.69 *	none	none
m4_pweak	6456.3	-1.94 ± 0.93 *	M.Gen.interval	PaperID
m4_none	6451.8	-1.94 ± 1.03 *	M.Gen.interval	PaperID
m2_pweak	6453.0	-2.07 ± 0.91 *	none	PaperID
m2_preg	6443.3	-2.11 ± 0.90 *	none	PaperID
m2_none	6457.3	-2.13 ± 0.90 *	none	PaperID
m5_pweak	6453.2	-2.70 ± 5.61 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m5_preg	-4812.2	-3.01 ± 1.06 *	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m5_none	6450.5	-3.03 ± 5.67 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
mshaw1_none	6376.6	-4.93 ± 1.75 *	Z.Year.Midpoint + Z.Gen.interval	PaperID

Table S 6: Percentage diversity  $\pi$  (delta SE) meta-regression model comparison. (*continued*)

label	DIC	summary	fixed effects	random effects
mshaw1_pweak	6369.6	-4.97 ± 1.75 *	Z.Year.Midpoint + Z.Gen.interval	PaperID
m6_pweak	6413.6	-5.65 ± 8.00 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
m6_none	6409.2	-5.85 ± 8.15 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
m6_preg	-4893.2	-6.24 ± 6.16 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
mshaw1_preg	-4862.0	-10.04 ± 1.61 *	Z.Year.Midpoint + Z.Gen.interval	PaperID

#### Average genetic diversity $\pi$ log ratio (geometric)

Table S 7: Log ratio diversity  $\pi$  meta-regression model comparison.

label	DIC	summary	fixed effects	random effects
mshaw2_pfix	-1631.3	-0.01 ± 0.01 *	M.Year.Midpoint + M.Gen.interval	PaperID
mshaw2_pweak	-1762.9	-0.01 ± 0.01 ns	M.Year.Midpoint + M.Gen.interval	PaperID
mshaw2_none	-1767.6	-0.01 ± 0.01 ns	M.Year.Midpoint + M.Gen.interval	PaperID
mshaw2_preg	-1758.6	-0.01 ± 0.01 ns	M.Year.Midpoint + M.Gen.interval	PaperID
m3_none	-1489.4	-0.01 ± 0.01 *	M.Gen.interval	none
m4_pfix	-1615.2	-0.01 ± 0.01 *	M.Gen.interval	PaperID
m3_pfix	-1611.3	-0.01 ± 0.01 *	M.Gen.interval	none
m3_pweak	-1489.9	-0.01 ± 0.01 *	M.Gen.interval	none
m3_preg	-1505.4	-0.01 ± 0.01 *	M.Gen.interval	none
m4_preg	-1665.7	-0.01 ± 0.01 *	M.Gen.interval	PaperID
m2_pfix	-1610.3	-0.01 ± 0.01 *	none	PaperID
m1_pfix	-1607.3	-0.01 ± 0.01 *	none	none
m1_pweak	-1486.1	-0.01 ± 0.01 *	none	none
m1_none	-1485.9	-0.01 ± 0.01 *	none	none
m1_preg	-1498.2	-0.01 ± 0.01 *	none	none
m4_pweak	-1698.7	-0.01 ± 0.01 *	M.Gen.interval	PaperID
m4_none	-1701.9	-0.01 ± 0.01 *	M.Gen.interval	PaperID
m5_pfix	-1624.2	-0.01 ± 0.01 *	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m2_pweak	-1700.9	-0.01 ± 0.01 *	none	PaperID
m2_preg	-1694.9	-0.01 ± 0.01 *	none	PaperID
m2_none	-1697.1	-0.01 ± 0.01 *	none	PaperID
m6_pfix	-1624.5	-0.01 ± 0.01 *	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
m5_preg	-1641.9	-0.02 ± 0.02 *	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m5_pweak	-1709.2	-0.02 ± 0.12 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m5_none	-1712.4	-0.03 ± 0.11 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m6_preg	-1636.7	-0.03 ± 0.03 *	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
mshaw1_pfix	-1629.8	-0.04 ± 0.01 *	Z.Year.Midpoint + Z.Gen.interval	PaperID
mshaw1_preg	-1761.5	-0.04 ± 0.02 *	Z.Year.Midpoint + Z.Gen.interval	PaperID
mshaw1_none	-1762.0	-0.04 ± 0.02 *	Z.Year.Midpoint + Z.Gen.interval	PaperID
mshaw1_pweak	-1769.9	-0.04 ± 0.02 *	Z.Year.Midpoint + Z.Gen.interval	PaperID

Table S 7: Log ratio diversity  $\pi$  meta-regression model comparison. (continued)

label	DIC	summary	fixed effects	random effects
m6_pweak	-1778.1	-0.07 ± 0.16 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
m6_none	-1775.9	-0.07 ± 0.15 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type

### Allelic richness M percent change

Table S 8: Percentage allelic richness M meta-regression model comparison.

label	DIC	summary	fixed effects	random effects
mshaw2_preg	-4072.7	2.09 ± 1.99 *	M.Year.Midpoint + M.Gen.interval	PaperID
m3_pweak	6761.7	0.64 ± 1.59 ns	M.Gen.interval	none
m3_none	6762.0	0.63 ± 1.57 ns	M.Gen.interval	none
m4_preg	6755.5	0.63 ± 1.58 ns	M.Gen.interval	PaperID
m3_preg	6755.6	0.60 ± 1.55 ns	M.Gen.interval	none
m6_preg	-10836.1	0.47 ± 5.33 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
m1_pweak	6760.3	0.45 ± 1.52 ns	none	none
m1_none	6760.7	0.45 ± 1.50 ns	none	none
m1_preg	6754.5	0.45 ± 1.49 ns	none	none
mshaw2_none	6591.6	0.14 ± 2.02 ns	M.Year.Midpoint + M.Gen.interval	PaperID
mshaw2_pweak	6589.7	0.13 ± 2.00 ns	M.Year.Midpoint + M.Gen.interval	PaperID
m4_none	6592.6	-0.18 ± 1.94 ns	M.Gen.interval	PaperID
m4_pweak	6592.8	-0.19 ± 1.95 ns	M.Gen.interval	PaperID
m2_none	6591.0	-0.25 ± 1.90 ns	none	PaperID
m2_pweak	6591.2	-0.26 ± 1.85 ns	none	PaperID
mshaw2_pfix	-4057.9	-0.56 ± 0.15 *	M.Year.Midpoint + M.Gen.interval	PaperID
m5_preg	-4074.9	-0.72 ± 1.89 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m2_preg	-8584.4	-0.77 ± 1.88 ns	none	PaperID
m5_pweak	6592.6	-0.77 ± 3.92 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m5_none	6591.9	-0.86 ± 4.27 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m5_pfix	-4057.0	-1.06 ± 0.14 *	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m3_pfix	-4052.8	-1.06 ± 0.14 *	M.Gen.interval	none
m6_pfix	-4056.2	-1.06 ± 0.14 *	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
m4_pfix	-4051.2	-1.06 ± 0.14 *	M.Gen.interval	PaperID
m1_pfix	-4055.5	-1.07 ± 0.13 *	none	none
m2_pfix	-4050.4	-1.07 ± 0.14 *	none	PaperID
m6_none	6592.0	-1.10 ± 5.25 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
m6_pweak	6592.5	-1.15 ± 5.19 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
mshaw1_pweak	6590.1	-1.95 ± 3.61 ns	Z.Year.Midpoint + Z.Gen.interval	PaperID
mshaw1_none	6590.0	-1.95 ± 3.61 ns	Z.Year.Midpoint + Z.Gen.interval	PaperID
mshaw1_pfix	-4059.6	-3.38 ± 0.35 *	Z.Year.Midpoint + Z.Gen.interval	PaperID
mshaw1_preg	-4107.1	-13.06 ± 3.07 *	Z.Year.Midpoint + Z.Gen.interval	PaperID

### Allelic richness M log ratio (geometric)

Table S 9: Log ratio diversity M meta-regression model comparison.

label	DIC	summary	fixed effects	random effects
mshaw2_pfix	-359.4	0.00 ± 0.01 ns	M.Year.Midpoint + M.Gen.interval	PaperID
m3_pfix	-355.9	0.00 ± 0.01 ns	M.Gen.interval	none
m4_pfix	-358.1	0.00 ± 0.01 ns	M.Gen.interval	PaperID
m3_preg	-310.4	-0.00 ± 0.01 ns	M.Gen.interval	none
m3_none	-305.7	-0.00 ± 0.02 ns	M.Gen.interval	none
m6_pfix	-358.3	-0.00 ± 0.02 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
m3_pweak	-304.2	-0.00 ± 0.02 ns	M.Gen.interval	none
m5_pfix	-358.7	-0.00 ± 0.01 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m1_pfix	-360.1	-0.00 ± 0.01 ns	none	none
m1_none	-306.3	-0.00 ± 0.02 ns	none	none
m2_pfix	-360.2	-0.00 ± 0.01 ns	none	PaperID
m1_preg	-312.7	-0.00 ± 0.01 ns	none	none
m1_pweak	-306.7	-0.00 ± 0.02 ns	none	none
mshaw2_preg	-561.5	-0.01 ± 0.02 ns	M.Year.Midpoint + M.Gen.interval	PaperID
mshaw2_none	-560.3	-0.01 ± 0.02 ns	M.Year.Midpoint + M.Gen.interval	PaperID
mshaw2_pweak	-561.3	-0.01 ± 0.02 ns	M.Year.Midpoint + M.Gen.interval	PaperID
m2_preg	-560.3	-0.01 ± 0.02 ns	none	PaperID
m4_preg	-563.2	-0.01 ± 0.02 ns	M.Gen.interval	PaperID
m2_pweak	-560.0	-0.01 ± 0.02 ns	none	PaperID
m4_pweak	-558.2	-0.01 ± 0.02 ns	M.Gen.interval	PaperID
m4_none	-555.9	-0.01 ± 0.02 ns	M.Gen.interval	PaperID
m2_none	-555.0	-0.01 ± 0.02 ns	none	PaperID
m5_preg	-559.4	-0.01 ± 0.02 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
m6_preg	-555.4	-0.01 ± 0.03 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
m5_none	-556.6	-0.02 ± 0.14 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
mshaw1_pfix	-358.3	-0.02 ± 0.03 ns	Z.Year.Midpoint + Z.Gen.interval	PaperID
m5_pweak	-557.7	-0.02 ± 0.15 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl
mshaw1_preg	-569.8	-0.03 ± 0.03 ns	Z.Year.Midpoint + Z.Gen.interval	PaperID
mshaw1_none	-560.0	-0.03 ± 0.04 ns	Z.Year.Midpoint + Z.Gen.interval	PaperID
mshaw1_pweak	-557.7	-0.03 ± 0.04 ns	Z.Year.Midpoint + Z.Gen.interval	PaperID
m6_none	-560.0	-0.04 ± 0.15 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type
m6_pweak	-561.0	-0.04 ± 0.16 ns	M.Gen.interval	PaperID + Kingdom + Binomen.rotl + Marker.Type

### Test per-study statistical differences: are there more or less positive trends?

Is the distribution of Hedge's g significantly different from zero for a given study? To test this, we follow a t-statistic distribution, which is the expected for Hedge's g. In addition we also conduct the test with a Gaussian distribution (which may be more akin to the meta-regression modeling, but this Normal distribution causes less significant results than t-distribution because is wider and thus more conservative). Figure S4 shows all studies ranked by  $-\log_{10}(p)$  of Hedge's g test per study, which has a substantial inflation. Figure 3 shows the inflation of P-values assuming Hedges'g follow a t-distribution, making our detection of significant effects anti-conservative.

Distribution of positive and negative  $\pi$  div trajectory between two time points. Using Hedge's g for significance using two sided test and family-wise Bonferroni  $P$ -value correction  $0.05/\text{number of tests}$ . The number of significant tests is as follows: Non-significant 823. Declining 50. Increasing 43.

- *For diversity  $\pi$*

Number significant trends (i.e. either positive or negative) = 93 of total 916 points.

Percentage of significant trends (i.e. either positive or negative) = 10%

- *For richness  $M$*

Number significant trends (i.e. either positive or negative) = 112 of total 785 points.

Percentage of significant trends (i.e. either positive or negative) = 14%

Are the number of significantly negative (decline of genetic diversity) and positive (increase in genetic diversity) more or less common? We conduct a Binomial test of enrichment of significant tests that are negative vs positive:

- *For diversity  $\pi$*

Prob. success (i.e. positive vs negative) = 0.4623656

$P$  value = 0.5340599

There are also 46  $P$ -values=0 (5.02% of dataset). The cases of  $P$ -value=0 are roughly equal in both directions, 24 decreasing, 22 increasing.

- *For richness  $M$*

Prob. success (i.e. positive vs negative) = 0.4196429

$P$  value = 0.1077823

There are also 31  $P$ -values=0 (3.95% of dataset). The cases of  $P$ -value=0 are roughly equal in both directions, 17 decreasing, 14 increasing.

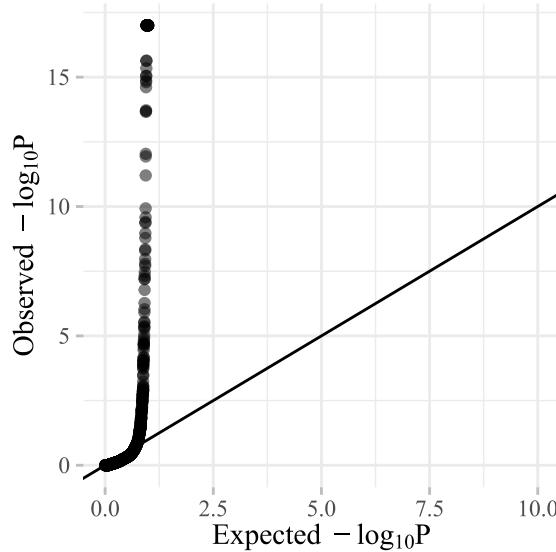


Figure S 3: P-value distribution in QQ plot.

We also test this with all the data points from Shaw et al (4023) to make sure this is not a subset effect:

- *For all dataset*

Prob. success (i.e. positive vs negative) = 0.5099237

$P$  value = 0.6391896

In addition, I compute the same binomial test without test statistics. Simply, count how many studies have Hedges' g (effect size) that is negative (2095) out of all the dataset (4023) and ask whether this is different from 50/50. There is no significant divergence from 50/50, with estimate 0.5207557 and  $P$ -value=0.995964.

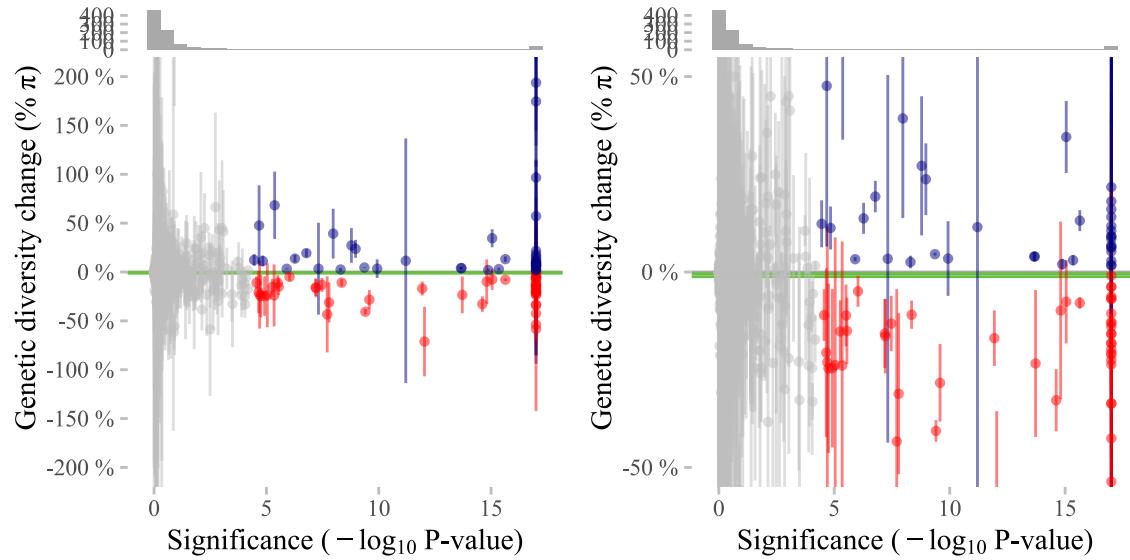


Figure S 4: Relationship between raw values of diversity change  $\pi$  and Hedge's g significance.

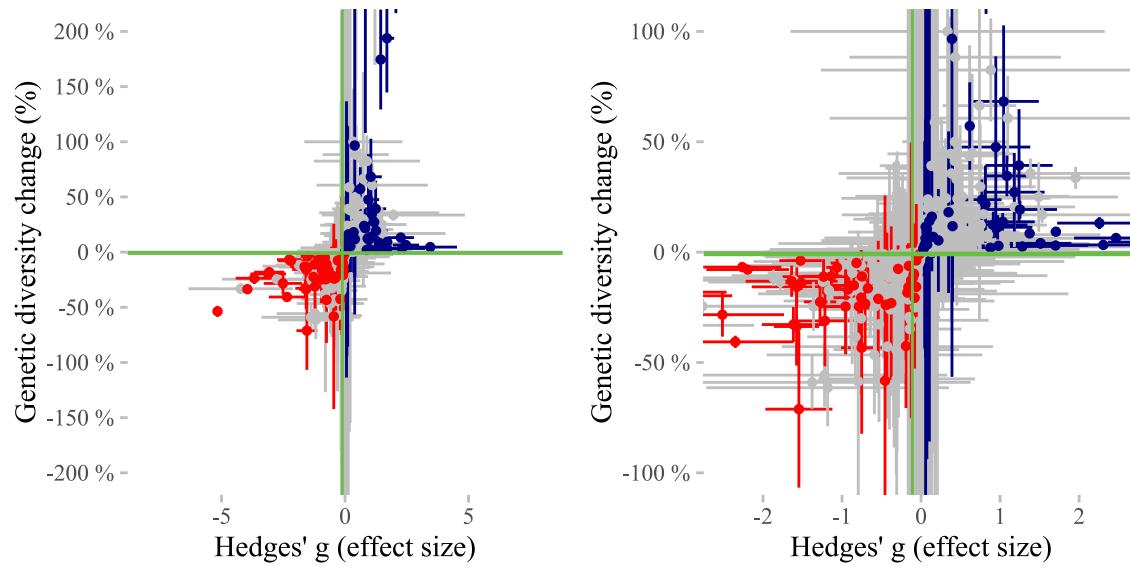


Figure S 5: Relationship between raw values of diversity change  $\pi$  and Hedge's g values.

## Test relationship with Red List

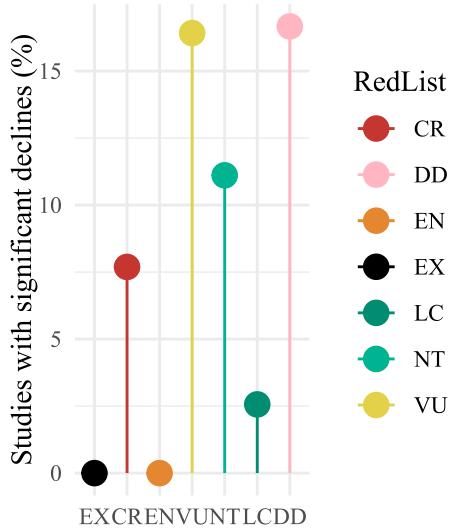


Figure S 6: Significant decline trends and Red List categories.

The number of significant decline cases in  $\pi$  (50), was significantly enriched with species in the Red List (all but least concern [LC] or data deficient [DD], 475, Fisher's test  $P=0.0012471$ , Odds=2.7783496) (Fig. 6). The number of significant genetic diversity increase cases too (Fisher's test  $P=0.1606518$ , Odds=1.6000509). This was mostly driven by vulnerable species (6).

### Test differences with increasing time.

The number of significant decline cases in  $\pi$  (50), were marginally significantly enriched with high number of generation interval ( $>30$  generations # studies= 37, Fisher's test  $P=0.0457039$ , Odds=2.8906597) (Fig. 7). This was non-significant if excluding positive trends (which may not be realistic) ( $P=0.0543964$ , Odds 2.7419411). Only focusing on studies with increasing genetic diversity trends showed also no enrichment (Fisher's test  $P=0.411707$ , Odds=0).

### Test differences with increasing data size.

There was a significant enrichment of trajectories of  $\pi\%$  decline in studies with number of loci above 10 (Odd=15.8647827, Fisher's test  $P=1.2975938 \times 10^{-12}$ ) as well as  $\pi\%$  increases (Odd=7.5969563, Fisher's test  $P=1.51457 \times 10^{-8}$ ).

There was not a significant enrichment of trajectories of  $\pi\%$  decline in studies with  $>100$  individuals sampled (Odd=0.1889661, Fisher's test  $P=0.0774391$ ) nor  $\pi\%$  increases (Odd=0.9965217, Fisher's test  $P=1$ ).

### Test differences with decreasing noise.

It has become clear that the very extreme values of genetic diversity increases and decreases are owed to high standard error. A good rule of thumb is that standard error  $< 5\%$  are accurate measurements, while  $>5\%$  are highly uncertain, as the measurements we are aiming to achieve are in the same or smaller order of magnitude. We then look at enrichment of significant values in the more certain studies. I find a marginal enrichment of significant decline diversity cases  $\pi$  and low error measurements (Odd=1.8501645, Fisher's test  $P=0.0443237$ ) but not with significantly increasing  $\pi$  studies (Odd=1.7378191, Fisher's test  $P=0.0966327$ ).

### Test differences with domesticated vs not species.

Finally, hat the domesticated/pathogens/pest speices not being removed, we would find an upward bias in genetic diversity change as there is an enrichment of domesticated species and significant positive trends in the original data (Odds= 1.8437971, Odds=  $3.9582345 \times 10^{-6}$ ).

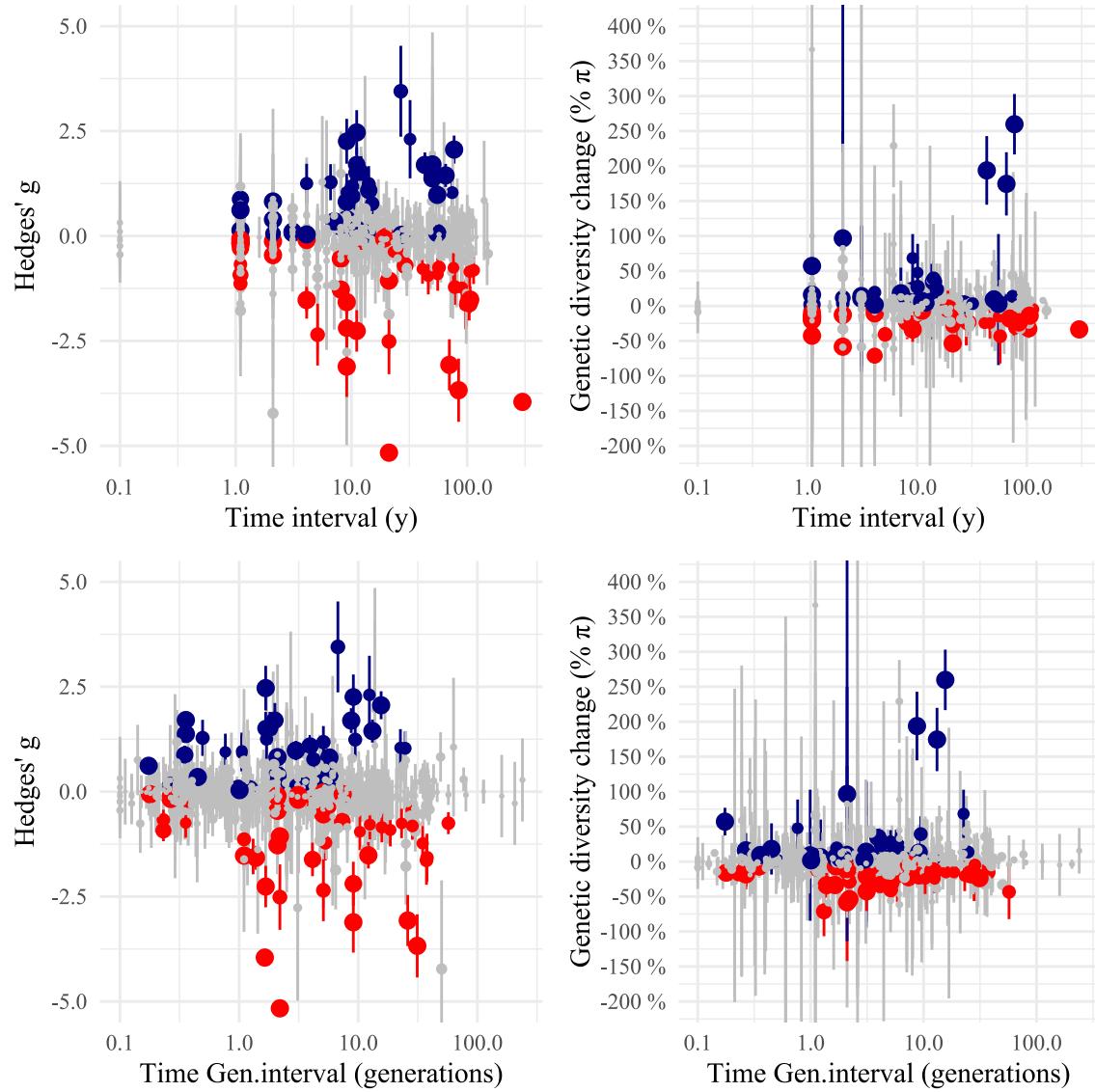


Figure S 7: Relationship between genetic changes and length of time interval.

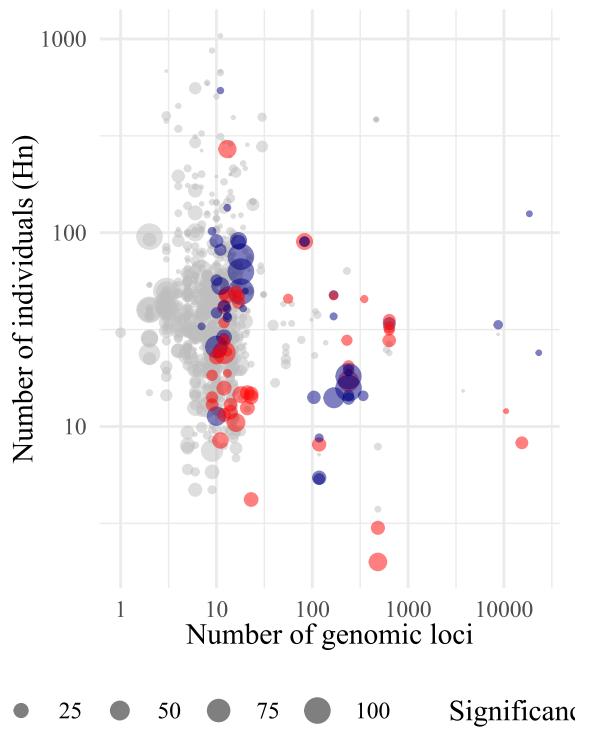


Figure S 8: Sample size (loci and individuals) and raw changes in  $\pi\%$  genetic diversity.

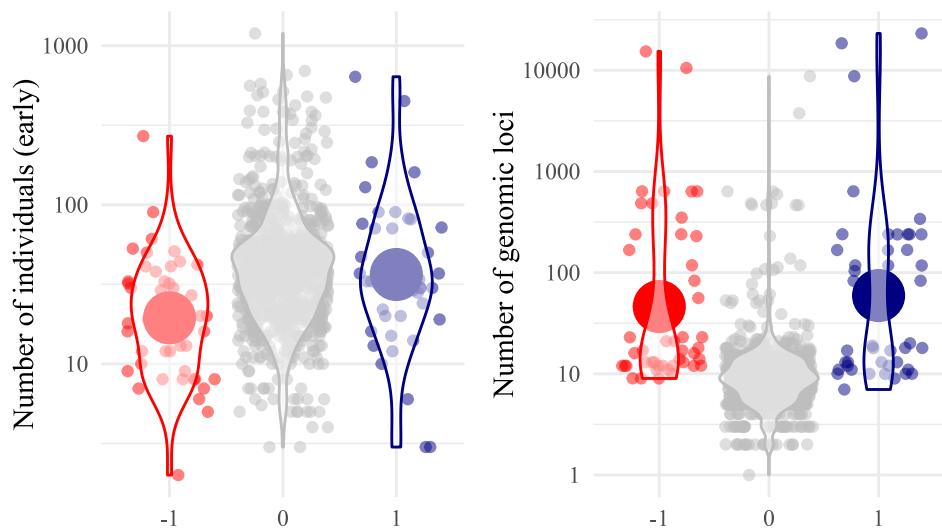


Figure S 9: Sample numbers across past and present studies.

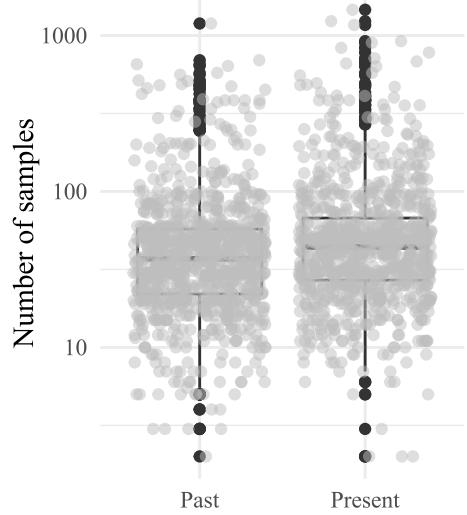


Figure S 10: Sample and marker numbers across studies

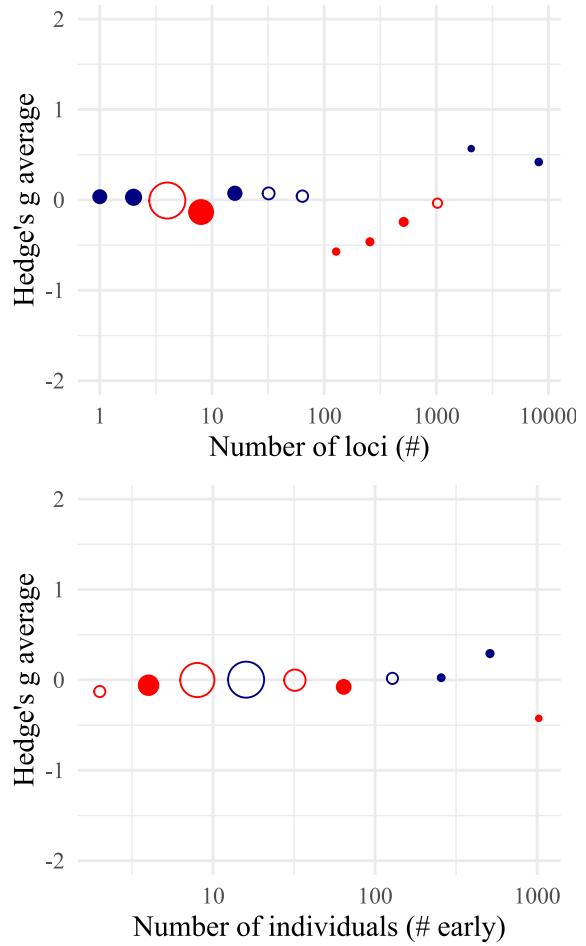


Figure S 11: Average diversity change of different data sizes.

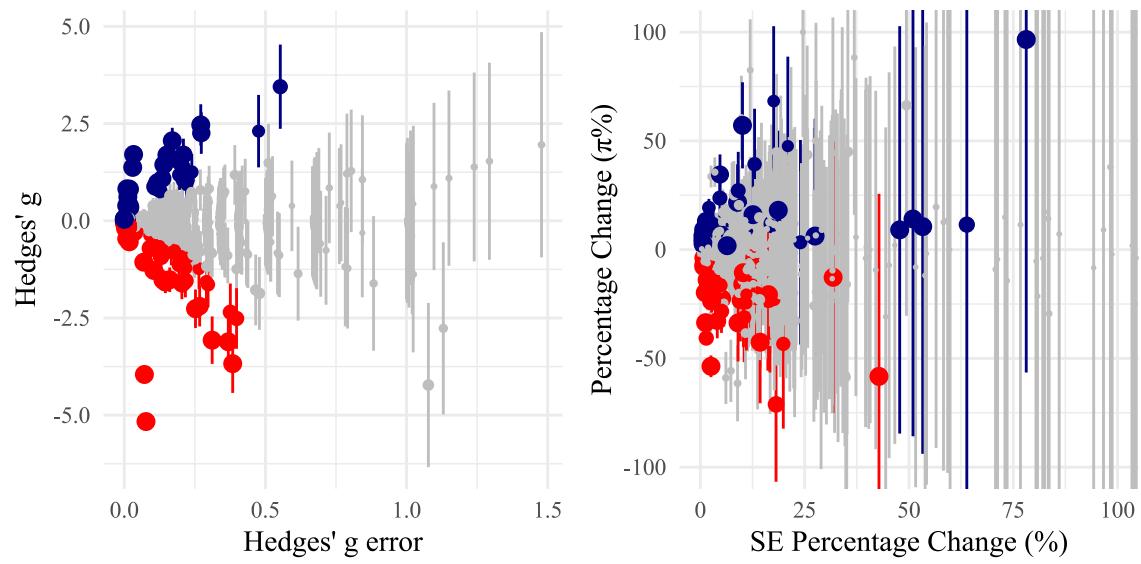


Figure S 12: Relationship between raw values, significance and noise.

### Text S3: Test within-species consistency.

There were repeated measures per species in the dataset. For instance, for  $\pi$  there were 916 studies, but within these there were a total of 323 unique species.

We hypothesize that if the observed decline are robust they should occur in the same direction within the same species. We then conduct a Binomial test of species with multiple significant observations, asking whether they tend to have more positive or more negative significant values.

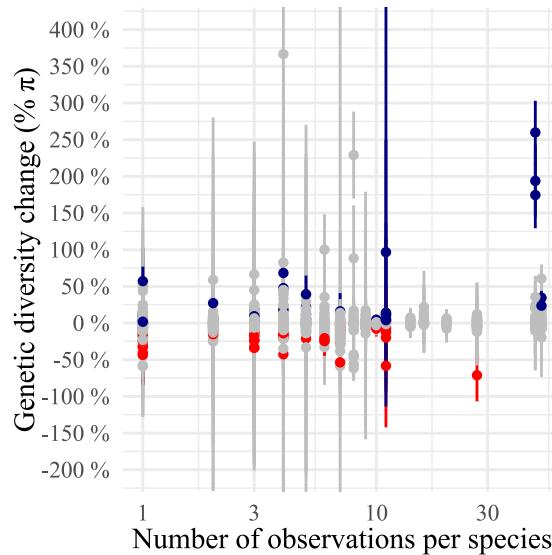


Figure S 13: Cases of species with multiple observations.

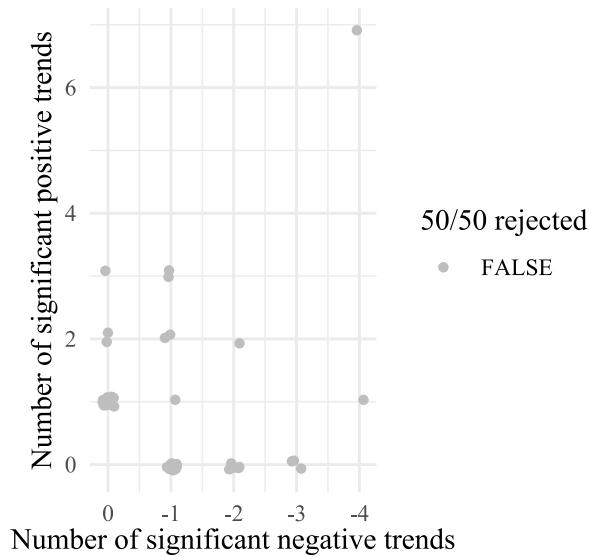


Figure S 14: Positive and negative cases of species with multiple observations.

Of 50 species with multiple populations sampled and significant trends, none had a ratio of negative trends (decline) vs positive trend (increase) that was significant under a Binomial test of expected 50/50% probability.

## **Text S4: Interpretation of changes under an evolutionary population genetic model of species contraction.**

Below we explore what % genetic diversity changes (non-significant or significant) could mean in terms of population contraction models of different scenarios (we assume positive changes are zero, as they are treated separately, see Text S5). Equations are derived in the **Mathematical Appendix**.

Some key notation:

- $\pi$  nucleotide diversity or expected heterozygosity (which are identical under a biallelic loci condition).
- $M$  allelic richness, or segregating sites, or number of variable positions, or number of mutations in a strand of DNA. Note we use  $M$  instead of more standard  $S$  or  $A$  to avoid confusions with species richness from ecology, and also to avoid confusions with area of habitat.
- $N_0$  the population size at some past time. We may also use  $N$  as this is the long-term equilibrium size.
- $N_t$  the population size at a given time  $t$ . Because we mostly deal with two time points (0 and 1), we often use  $N_1$  the population size instantaneous reduction from human impact or contraction.
- $A_0$  the habitat area of a species with multiple populations (useful to use area to avoid defining discrete entities).
- $A_1$  the habitat area of a species reduced from human impact or contraction.
- $X_\pi$  the fraction loss of genetic diversity  $\pi$ . Intuitive % of genetic diversity loss when  $X_\pi \times 100$ .
- Same for  $X_M$  the fraction loss of allelic richness  $M$ .

### **BIOLOGICAL SAMPLING – *null scenario***

**Average genetic diversity  $\pi$**  The high noise is likely due to the small sampling sizes:

Present sampling median  $n_0=48$

Present sampling median  $n_1=52$

Number of markers median  $L=9$

Taking biological sampling into account, we found 22 studies out of 306 with changes in measured genetic diversity (-18%) beyond the 5% tail of the null distribution from subsampling (-8.4%). This corresponds to an excess of -10%.

Bootstrapping 1000 times one simulation per paper gives a range of average  $\pi\%$  change= -1.1 - 6%.

The reason observations are very noisy is because the median sampling effort in Shaw is  $n_0=37$ ,  $n_1=44$ , and the median number of loci is  $L=10$ . Fig. 18 shows the results of standard deviation of percent diversity change (% $\pi$ ) in a population of constant size sampled two times, in a grid of values from 5 individuals to 10,000 individuals, and a number of markers 1 to 1M. The median values of Shaw are marked in green, and indicate that the median error rate is over 10%, just as we see in the data.

Based on Fig. S17, the expected error for 50% of the Shaw dataset is a minimum of 32.6933738%, and for 75% percentile of the dataset a minimum of 11.3687606%.

**Allelic richness  $M$**  TBD.

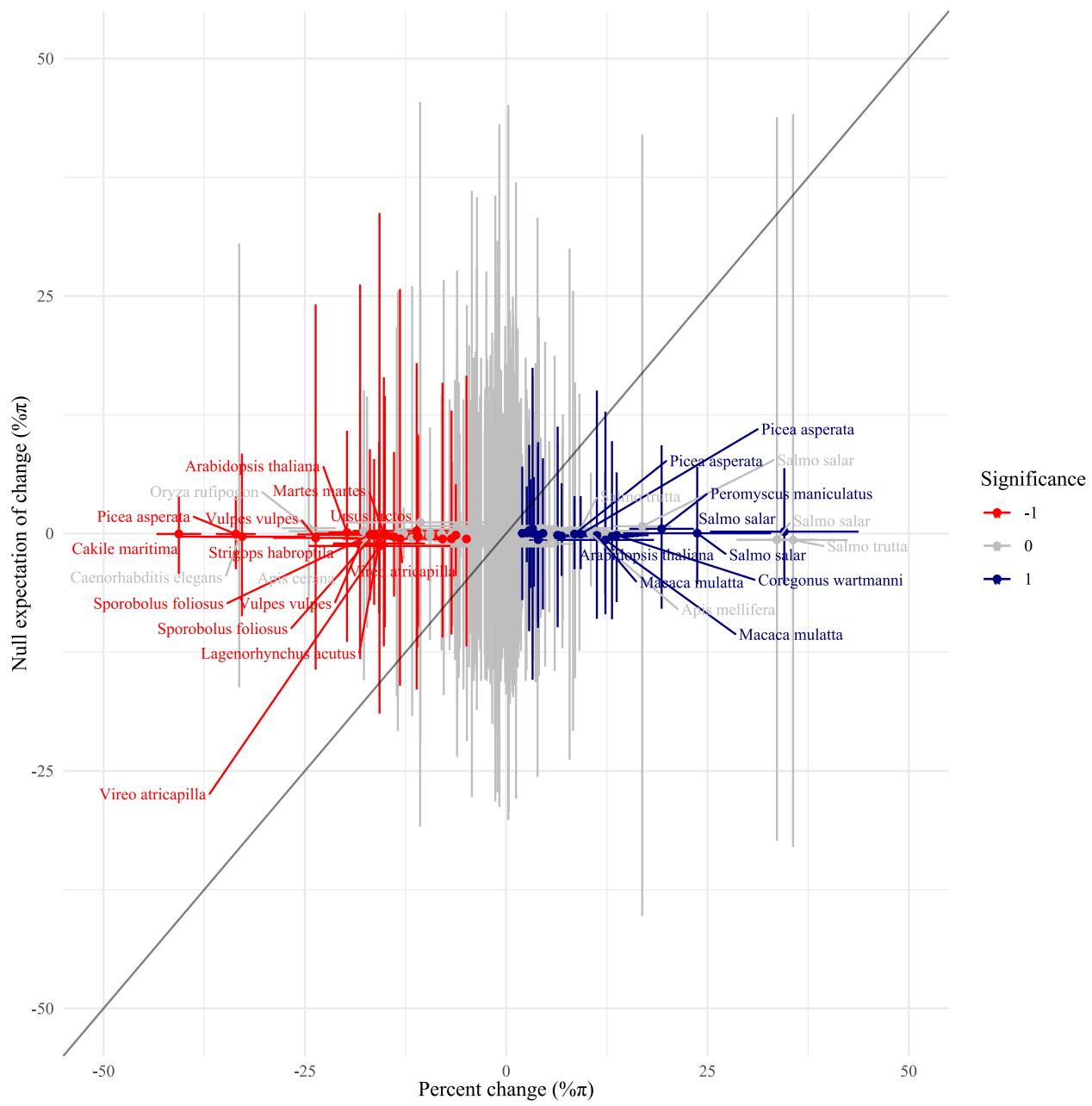


Figure S 15: Ranges of expected changes of diversity  $\pi\%$  with biological sampling

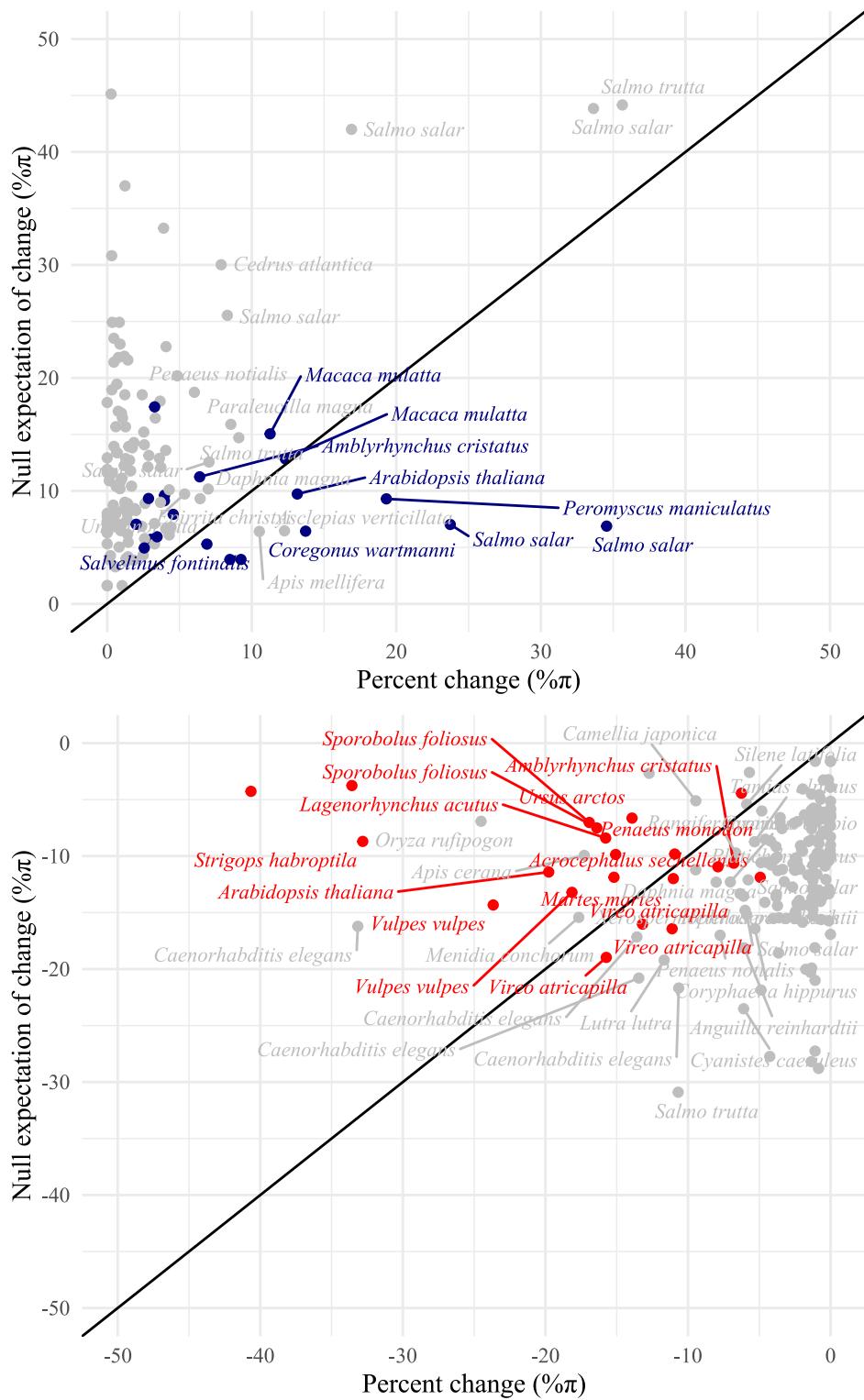


Figure S 16: Expected maximum changes of diversity  $\pi\%$  with biological sampling

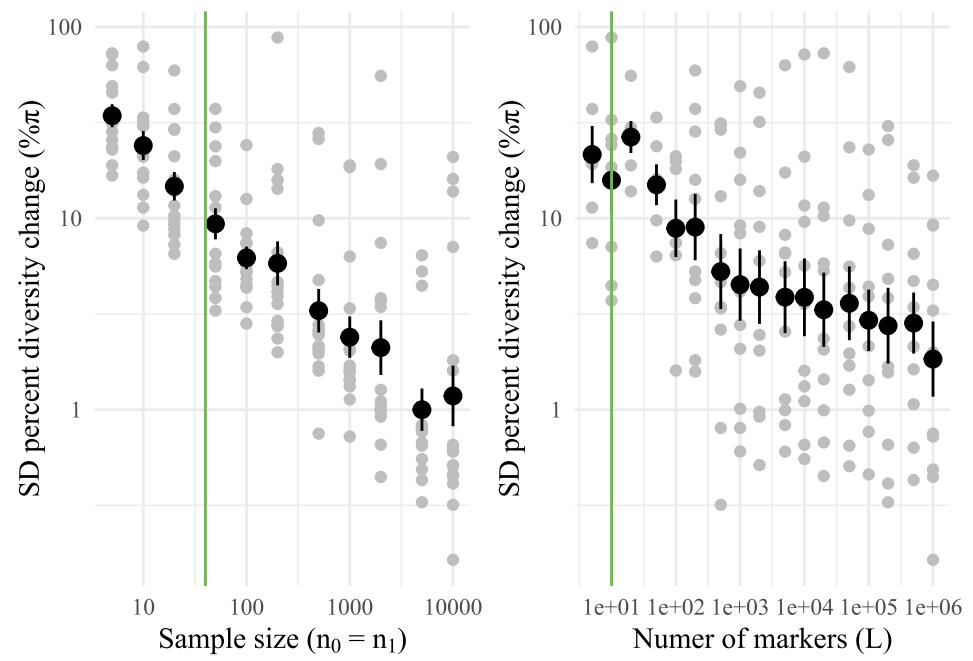


Figure S 17: Expected noise in percent diversity change  $\pi\%$  with different sample sizes

### SUBSAMPLE Single population, immediate – scenario 1a

**Average genetic diversity  $\pi$**  Based on population evolutionary genetic theory, in the **Mathematical Appendix**, we show that the expected relative  $\pi\%$  change (or the fraction change,  $X_\pi$ , which is more mathematically convenient) from a reduction of  $N_0$  to  $N_1$  within a generation is null. The only reason genetic diversity may change as a subsample of a population is due to sampling noise or if for some reason the remaining fraction of the population is more related to each other. This affects a fraction of the dataset: 295 observations, 32%. The results in Fig. 18, corroborate the classic observation that substantial losses ( $\sim 5\%$ ) of genetic diversity are not expected unless the bottleneck in a species is  $\ll 100$  individuals. NB. Fig. 18: A jitter of 0.1N units was added for visualization. Min value of bottleneck truncated to 10M. Positive trend % of genetic diversity collapsed to 0%.

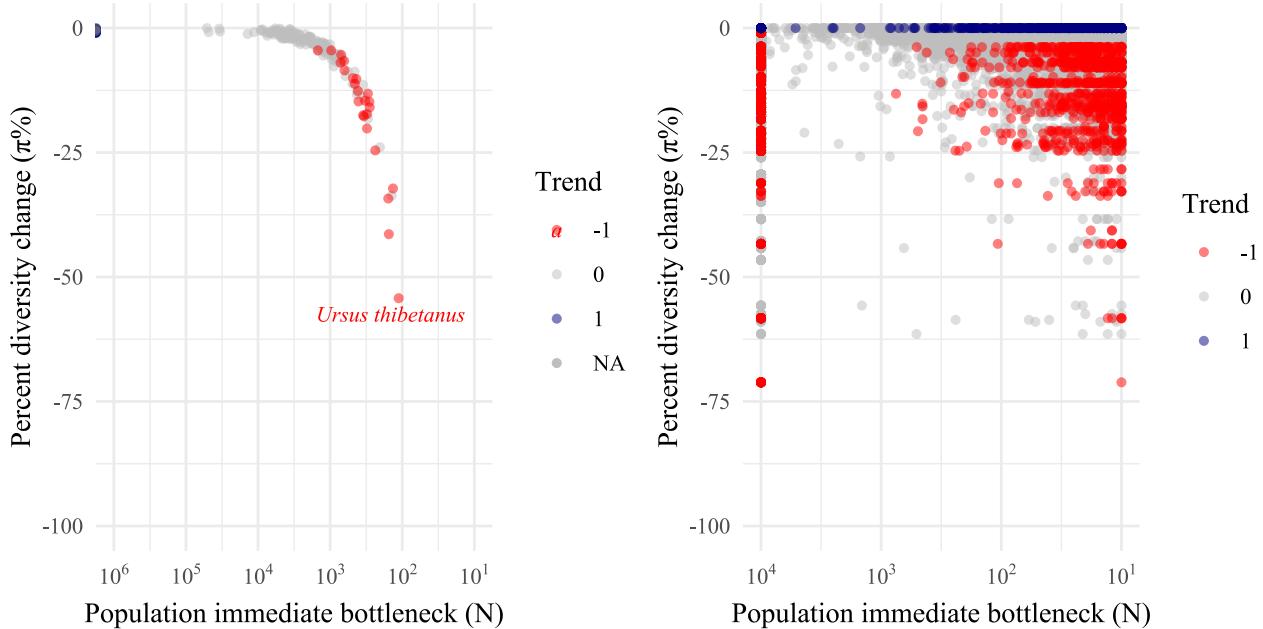


Figure S 18: Expected immediate population bottleneck given observed  $\pi\%$  with noise

*No noise* Median = 330 individuals.

Mean = 436 individuals.

IQR = 270 - 501 individuals.

Range = 93 - 1307 individuals.

*Noise* Median = 4 individuals.

Mean = 10 individuals.

IQR = 2 - 9 individuals.

Range = 1 - 752 individuals.

**Allelic richness  $M$**  The loss of alleles  $M$  follows under the coalescent a function of:  $X_M = -\frac{\log(1-X_N)}{\log(N_0)}$ . In the case of alleles  $M$  the loss not only depends on the fraction loss of individuals  $X_N$ , but also the starting population sizes  $N_0$  (this is unknown, so Fig. ?? shows trajectories with several  $N_0$ ). NB Fig. ?? A jitter of 1N units was added for visualization. Positive trend % of genetic diversity collapsed to 0%. The pattern observed of four trajectories correspond to inferences assuming  $N_0 = 100, 1000, 10000, 1000000$  starting individuals prior the bottleneck (respectively left to right).

Using the results assuming a starting size of  $N_0=10000$  individuals. For allelic richness  $M$  the loss is faster than for  $\pi$

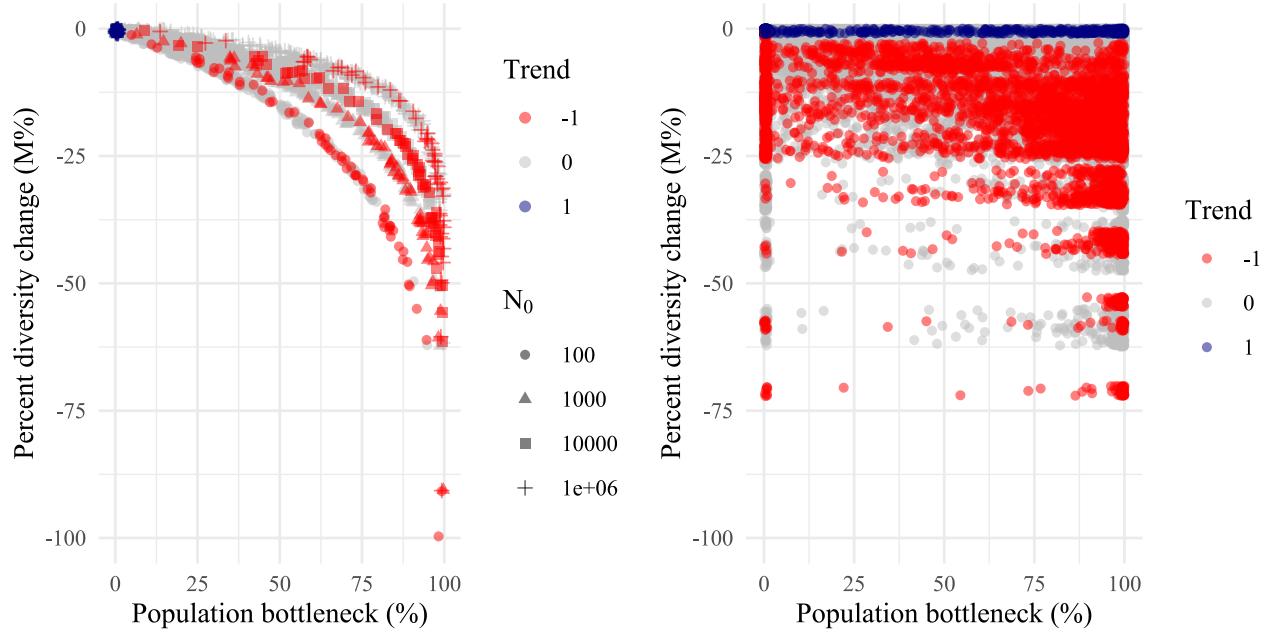


Figure S 19: Expected immediate population bottleneck given observed M%.

diversity, as expected.

*No noise* (assume the case of  $N_0=10000$ ) Median = NA% of original population

Mean = NaN% of original population

IQR = NA - NA% of original population

IQR = Inf - -Inf% of original population

*Noise* (assume the case of  $N_0=10000$ ) Median = 80% of original population

Mean = 62% of original population

IQR = 23 - 96% of original population

IQR = 0 - 100% of original population

### TIME Single population, drift over time – scenario 1b

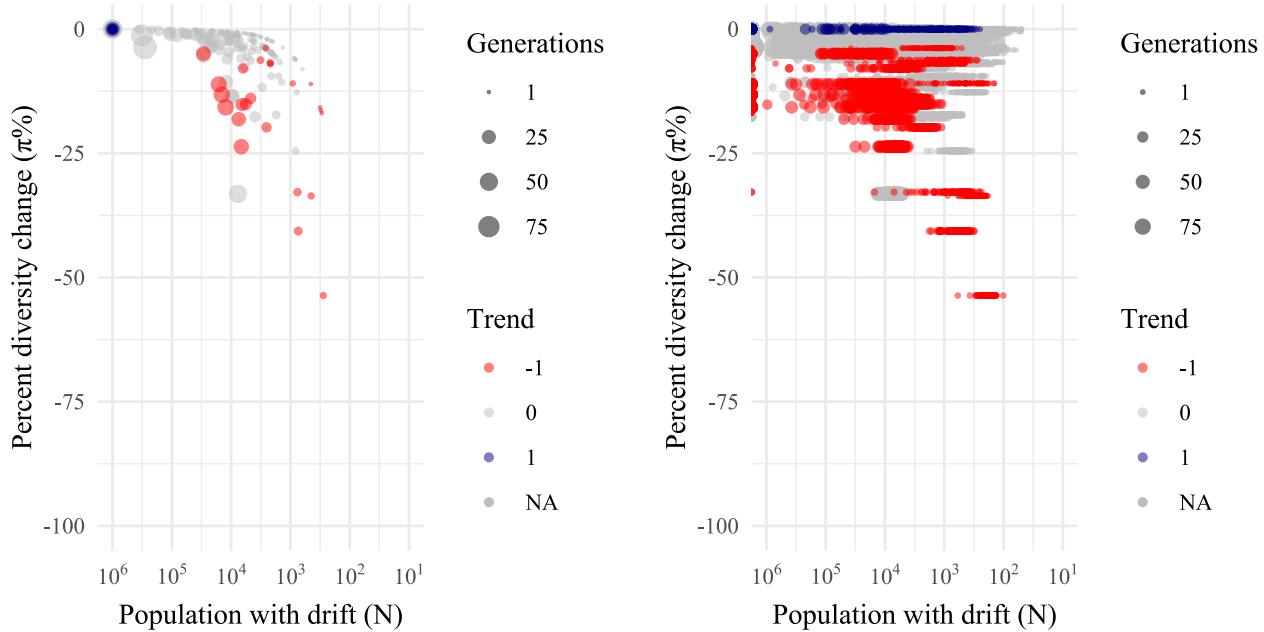


Figure S 20: Expected population bottleneck and drift given observed  $\pi\%$ .

#### Average genetic diversity $\pi$

NB. Positive trend % of genetic diversity collapsed to 0%.  
 Number of generations truncated to 100+ for visualization purposes.  
 Number of population size needed to explain loss truncated to 1M.

#### No noise

Median = 2570 individuals.  
 Mean = 5292 individuals.  
 IQR = 666 - 6623 individuals.  
 Ranges = 279 - 29401 individuals.

#### Noise

Median = 2624 individuals.  
 Mean = Inf individuals.  
 IQR = 426 - 9195 individuals.  
 Ranges = 0.5 - Inf individuals.

#### Allelic richness $M$

Population bottleneck truncated up to  $10^5$  for visualization

Positive trend % of genetic diversity collapsed to 0%.  
 Starting population also needs to be assumed,  $N_0 = 10^6$ , which is a conservative value (as a smaller starting opulation is assumed, the bottleenck predicted is even larger)

Expected instantaneous population bottleneck percentage for studies with significant decline of  $M$  :  
 Median = 72% of original population  
 IQR = 22 - 486

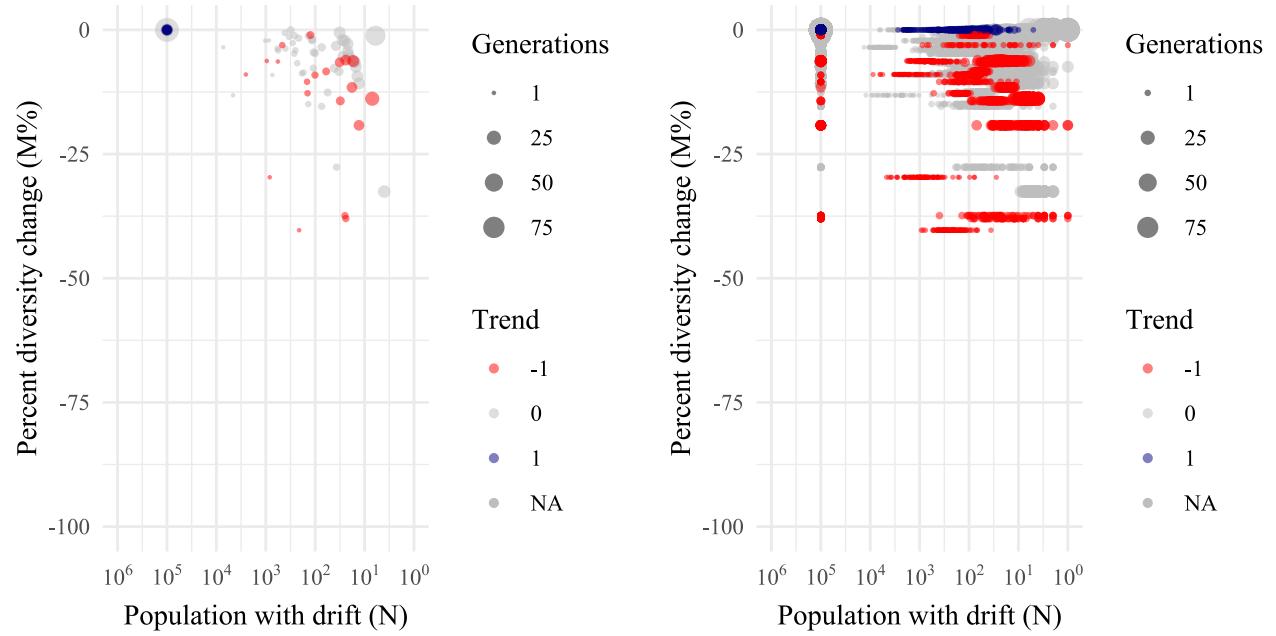


Figure S 21: Expected immediate population bottleneck given observed M%.

### SPACE. Spatial meta-populations, immediate – scenario 2a

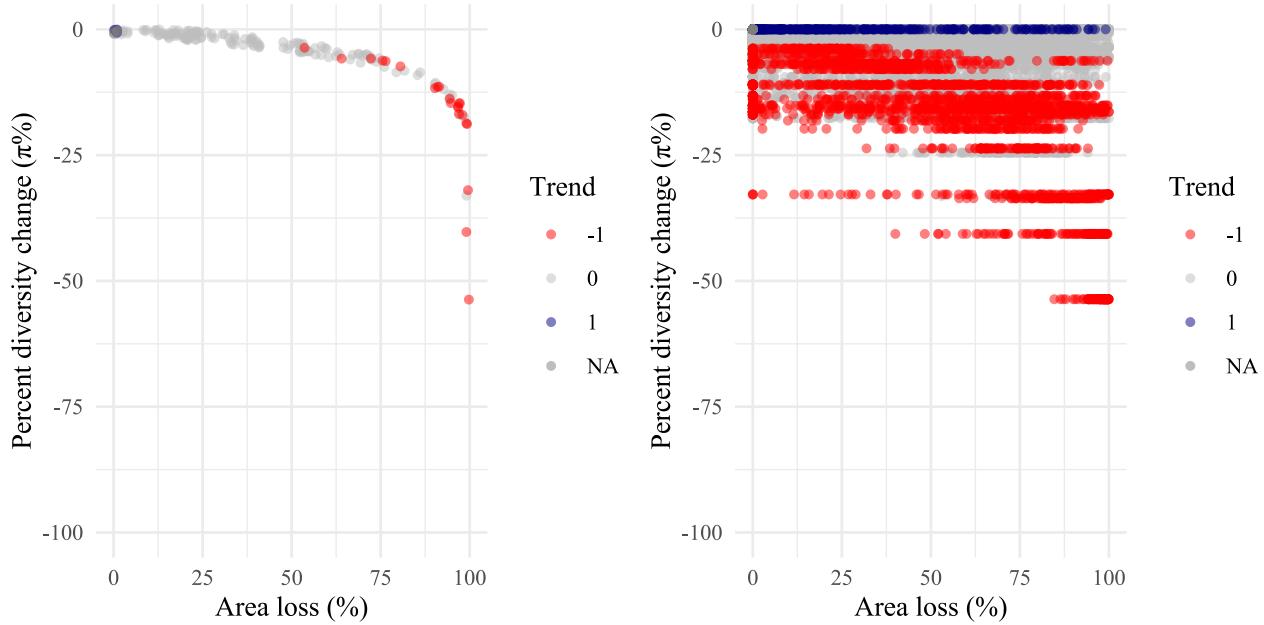


Figure S 22: Expected immediate geographic area loss given observed  $\pi\%$ .

**Average genetic diversity  $\pi$**  NB. A jitter of 1% area units was added for visualization. Positive trend % of genetic diversity collapsed to 0%.

#### No noise

Inverse modeling of instantaneous  $M$  richness loss with spatial contraction:

Median = 96% area

Mean = 90% area

IQR = 88 - 98% area

Range = 54 - 100% area

#### Noise

Inverse modeling of instantaneous  $M$  richness loss with spatial contraction:

Median = 55% area

Mean = 51% area

IQR = 25 - 79% area

Range = 0 - 100% area

Expected instantaneous area contraction percentage for studies with significant decline of  $\pi$  :

Median = 55% of original range

IQR = 25 - 79

**Allelic richness  $M$**  Positive trend % of genetic diversity collapsed to 0%.

The above reconstructions take at face value the average genetic diversity change, but as seen in Text S2 there is much noise in these estimates. A more true visualization would re-generate each observation 100 times following the observed error in the quantity.

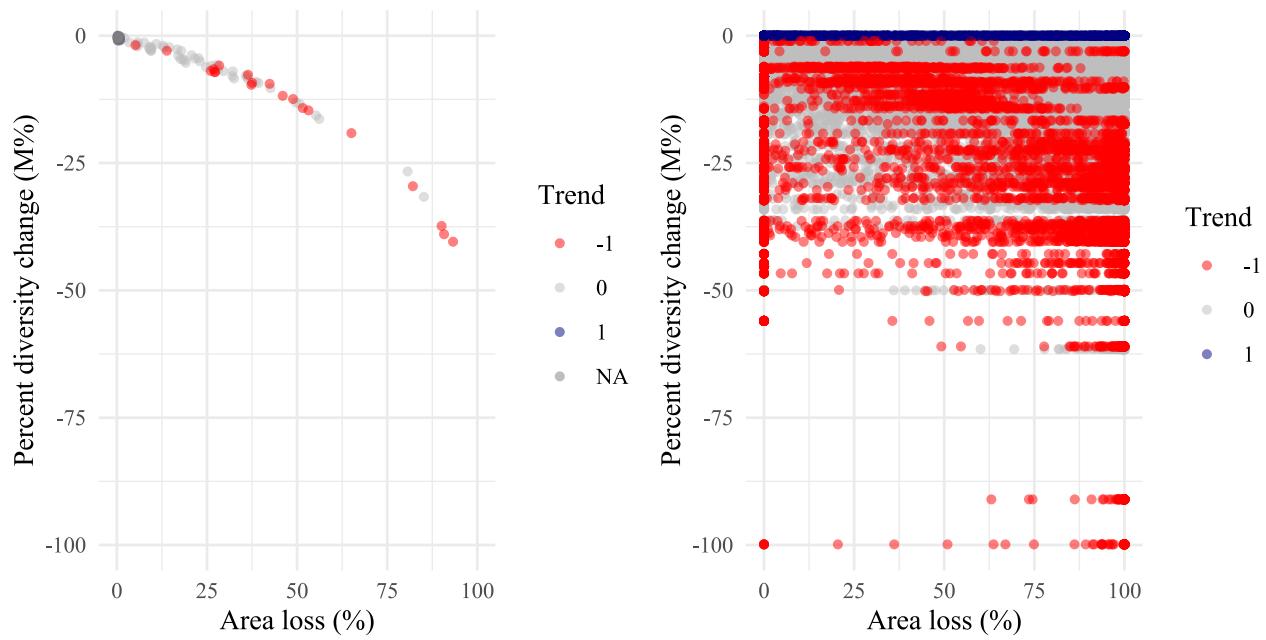


Figure S 23: Expected immediate geographic area loss given observed M%.

Expected instantaneous area contraction percentage for studies with significant decline of  $M$  :  
 Median = 61% of original range.  
 IQR = 0 - 97.

### SPACE & TIME. Spatial meta-populations via WFmoments, short drift – scenario 2b

Both spatial contraction and short drift. This is the most complicated scenario as we do not have a direct simple mathematical method. Instead, we have pre-computed scenarios of a solved numerical method we call WFmoments, that can predict the non-equilibrium genetic diversity of any meta-population system over time. We also only have solutions for average genetic diversity  $\pi$ . In this approach, then we created a number of loss area scenarios in a system of 100 demes in a 2D lattice, losing demes in an edge contraction manner. We then inverse model the area loss corresponding to the matched  $\pi$  change simulated and the observed  $\pi$  changes. We can use multiple functions, although they give qualitatively similar results. We decide for a Gaussian kernel, where the distance between the observed and simulated  $\pi$  changes exponentially decreases.

By definition, meta-population dynamics will be most relevant to species with multiple populations, and thus may not be as affected by single population dynamics (i.e. with a number of populations in the landscape, their effective population size will unlikely be below the ~500 where we see single population drift losses). Then, when in a meta-population landscape some populations are lost, drift will increase but most importantly the whole system will tend towards a future lower genetic diversity equilibrium. Population genetics tells us that for populations to reach equilibrium about  $4Ne$  generations. Assuming different starting  $N_0$  values and using the information of time interval, we can then compute

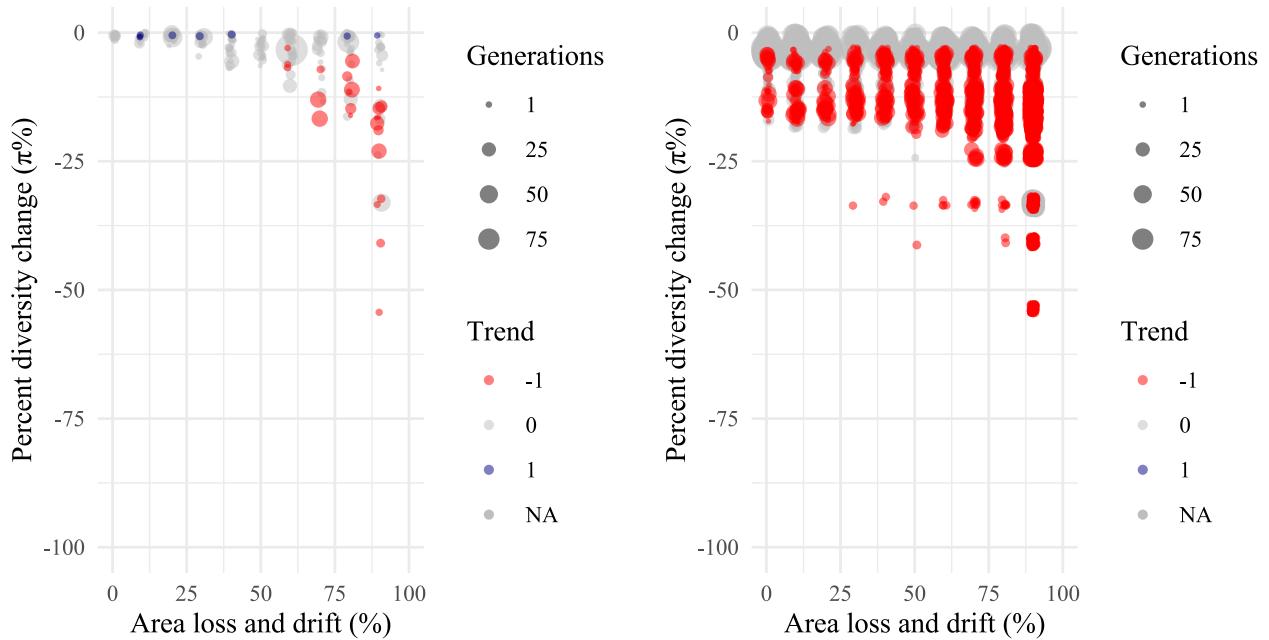


Figure S 24: Expected immediate geographic area loss given observed  $\pi\%$ .

#### Average genetic diversity $\pi$ Summary:

##### No noise

Median = NA% of original range lost.

Mean = NaN% of original range lost.

IQR = NA - NA% of original range lost.

Ranges = Inf - -Inf% of original range lost.

##### Noise

Median = 90% of original range lost.

Mean = 73% of original range lost.

IQR = 60 - 90% of original range lost.

Ranges = 0 - 90% of original range lost.

*Allelic richness*  $M$  We have no model for this.

## Text S5: Interpret increase in genetic diversity

There are three options to interpret population genetic diversity: (1) that genetic diversity is indeed increasing species-wide, (2) genetic diversity is inflated due to changes in meta-population dynamics, (3) measurement error.

### SINGLE POPULATION GROWTH OVER TIME – *gain scenario 1b*

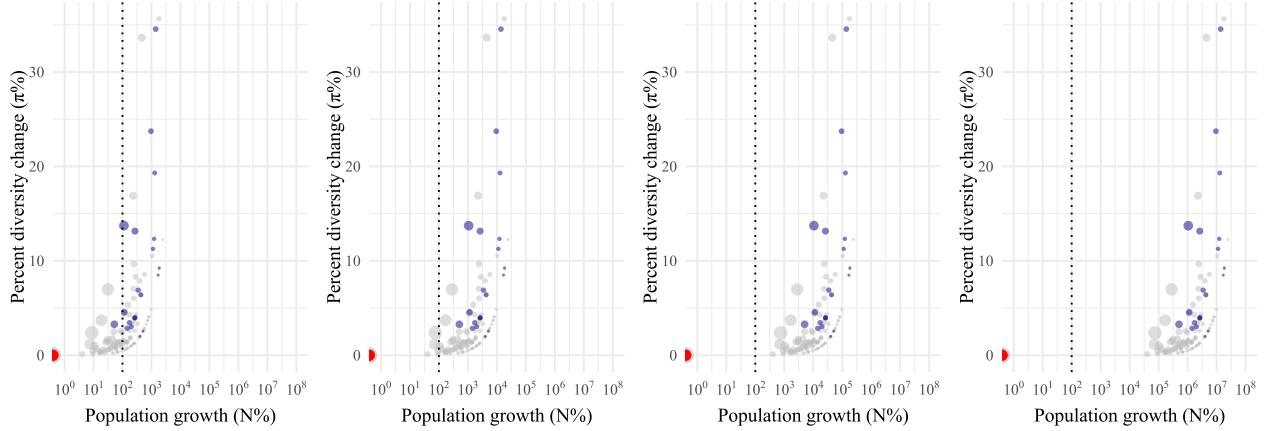


Figure S 25: Expected population growth given increases in  $\pi\%$ .

Fig. 27 shows plots assuming final population sizes:  $N1 = 100, 1000, 10000, 1000000$ . All data points of decline area assumed to be 0%. Dotted line marks the 100% increase, or doubling of population.

Fig. 26 explains why growth of genetic diversity is even less expected than declines. It assumes a population started at 10,000 individuals, and went down to 1,000 or increased to 15,000.

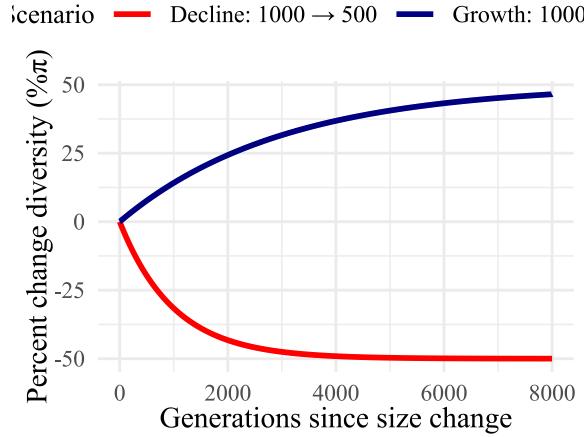


Figure S 26: The shape of genetic diversity increases is slower than decreases.

## SPATIAL METAPOPULATIONS MIXING – *gain scenario 2a*

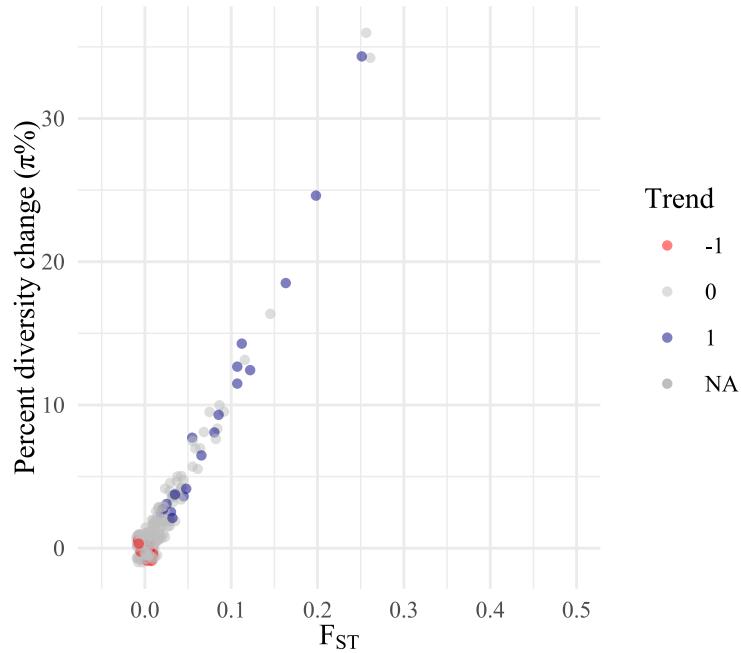


Figure S 27: Expected  $F_{ST}$  to cause a inflation of  $\pi\%$ .

The median  $F_{ST}$  required for a mixing of two popualtions to cause the percentage genetic change increase observed in positive cases will be: NA (27).

NB 28. Plot truncated to up to +100% in genetic diversity. Negative diversity trends truncated.

This also shows that it could be that even if genetic diversity increases, area loss and fragmentation could have happened.

Expected instantaneous area contraction percentage for studies with signican decline of  $\pi$  :  
Mean = 43% of original range.

Median = 40% of original range.

IQR = 28 - 70% of original range.

Range = 0 - 90% of original range.

## Text S6: Simple population contraction averages from LPI

As a though experiment – and to avoid controversial discussion on how or whether to measure abundance changes from LPI data – I simply take the earliest and latest census value for each LPI tracked population and compute a arithmetic average if a species has multiple observations. This yields -54% [IQR = -27 - 81%]. The median number of populations is: 2 populations [IQR = 1 - 5]. The median number of years in between earliest and latest observation is: 9.8 years [IQR = 4.4 - 17 years].

A thought experiment back-calculation of genetic diversity loss can be created using, for instance, scenario 2a,  $X_\pi = 16\%$  [IQR 1.6 - 8.1%].

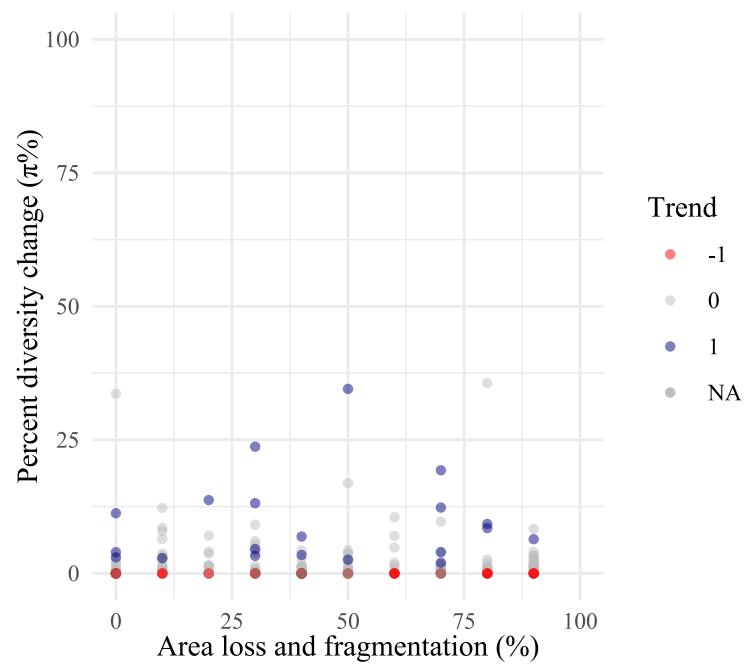


Figure S 28: Expected noisy inflation of  $\pi\%$  from area loss and fragmentation.

## **Text S7: Limitations, caveats, and thoughts forward**

### **Thoughts on data limitations**

- Published studies may have too short timelines to detect signals. About 10 years or 2 generations on average are in evolutionary trends effectively instantaneous and even if populations are undergoing bottlenecks and drift changes are expected to be small. Possibly applying historical/ancient DNA technologies would enable the temporal breadth backwards (see also next point). This will help avoiding the “baseline shifting problem” in biodiversity, where our reference point of what “original diversity” was shifts constantly leading to having a recent past (already eroded nature) as our pristine nature reference.
- A problem in genomic monitoring that we are facing now and will continue happening in the future is that genetic technologies are moving fast. Therefore, comparing observations published in the literature 10 years ago is today considered obsolete genetic technologies, so meta-analysis will not make use of the best current data. This is happening in the Shaw dataset, a typical molecular ecology study measuring genetic composition of a population may include 10 microsatellites for genotyping. Nowadays, even moderate budget projects could possibly use genomic technologies to generate much more accurate values of genetic diversity (NB the measurement error rate scales 1/number of bp sequenced). This can be done even using pool sequencing (i.e. sequencing multiple individuals’ DNA in one tube) since one aims to get a population-level diversity estimate. In addition, while many genomic approaches need of assembling a genome blueprint for a species, Kmer-based approaches (digesting sequencing reads directly in DNA sequence tables) has proven to be efficient at characterizing genetic diversity (Roberts & Josephs et al. 2025, Evolution Letters).
- A problem in genetic data often ignored or not reported is the scale of measurement of genetic diversity. While population genetics classically describes the individual, the population, and the species—which has been adopted to essential biodiversity variables (Hoban et al. 2022 Biological Reviews)—these three levels are not well captured in sampling protocols nor in policy texts (CBD). The difficulty is that genetic diversity within a population may change while it not changing at the species-level scale. Likewise genetic diversity within an individual (heterozygosity or inbreeding coefficient) may change but diversity at the population is maintained. Different demographic processes, evolutionary processes, may affect these without them being problematic, in fact they may go in different directions. It seems all levels are relevant (intuitively perhaps within-species may be the most relevant) and analyses should be conducted at all such three scales.

### **Thoughts on inverse model limitations**

Apart from data limitations, the inverse modeling proposed in the text also has limitations and possible improvements. No model is going to be perfect and one needs to balance usability (e.g. back of the envelope predictions vs complex process-based simulations). Here I mostly opted for back of the envelope mathematical predictions such as single population dynamics or genetic diversity-area relationships as these are very powerful to scale to many species (scenario 1a,1b,2a). For more accurate predictions and dynamics I also included some simulation/numeric based approaches (scenario 2b, Mualim et al 2024 bioRxiv) which can be extended to any population geographic conformation wanted, starting population size, migration rates, but these require further information per species.

- Shifting baseline problem underestimates impacts: Many species have already decreased in population sizes and thus contemporary measurements used in modeling (bottlenecks - to - genetics) or inverse modeling (genetic - to - bottlenecks) equations will underestimate past declines of genetic diversity before either genetic monitoring or population monitoring were established.
- Survival bias underestimates impacts: by definition genetic monitoring is only going to capture genetic diversity within a population, but cannot quantify genetic diversity loss of unsampled or extinct populations. It is likely that many populations within species are being altered with habitat losses (Exposito-Alonso 2022 Science) and thus genetic diversity loss should be larger than, e.g., what is inferred in Fig. 2.
- Wild populations often do not follow Wright-Fisher evolutionary assumptions, modeling may under- or overestimate impacts. The inferences based on evolutionary theory equations, equilibrium, sampling, necessarily use assumptions. Some assumptions involve no natural selection, random outcrossing of individuals (scenario 1a/1b), equal reproduction, etc. Underestimation may occur if, for instance, bottlenecks are impacting populations’ Ne

to a larger degree due to selfing or other. Models may overestimate impacts of bottlenecks, for instance, when in nature populations have overlapping generations that maintain genetic diversity.

- Unknown refugia/reservoirs of populations containing much of a species genetic diversity will protect it to impacts elsewhere thus models may over-estimate genetic diversity loss.
- Historical trends have legacy trajectories in genetic diversity. Species are not static, and are often been changing in population size for long periods of time. Whether these are upward or downward, will have effects in genetic diversity trajectories in the future (even with contemporary rapid changes in population).