Mathematical appendix | 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

Contents

SETUP

Overview of evolutionary population genetics parameters

Diversity parameters

- π = nucleotide diversity or expected heterozygosity (identical under biallelic loci)
- S = allelic richness, segregating sites, number of variable positions, or number of mutations in a DNA strand. **Note**: We use M instead of S to avoid confusion with species richness from ecology, and avoid A (alleles) to prevent confusion with area of habitat
- θ = diversity parameter of a population; under equilibrium $\theta = 4N_e\mu$

Population parameters

- N_0 = population size at some past time (long-term equilibrium size)
- N_1 = population size after instantaneous reduction from human impact or contraction (smaller population, not in equilibrium)
- N_t = population size at time t
- A_0 = original habitat area of a species with multiple populations
- A_1 = reduced habitat area after human impact or contraction

Parameters of change

- X_π = fraction loss of genetic diversity π (intuitive % loss when $X_\pi \times 100$)
- X_M = fraction loss of allelic richness M
- X_N = fraction loss of population size = $1 - (N_1/N_0)$

Stable state of genetic diversity in a population

Under equilibrium conditions (where $N_e = N_0$), the expected genetic diversity follows from the coalescent process. The total length of all coalescent branches in a population genealogy is L_{total}, and the number of mutations is $M = \mu L_{\text{total}}$.

The expected genealogy size under the coalescent is:

$$E[L_{\text{total}}] = 2N_e H_{n-1} \approx 2N_e \log(n-1) + c$$

where: - $2N_e$ = average time to coalescence between any two samples - H_{n-1} = harmonic number representing progressive coalescent of all samples - This can be approximated with $\log(n-1)$ for large n

Therefore:

$$E[M] = \mu E[L_{\text{total}}] \approx \mu \cdot 2N_e \log(n-1)$$

For nucleotide diversity π (equivalent to expected heterozygosity under biallelic variants), under equilibrium:

$$E[\pi] = 4N_e \mu$$

This represents the average number of genetic variants between two samples along a DNA stretch, derived from the average pairwise coalescent time:

$$E[L_{\text{pairwise}}] = 2 \times 2N_e = 4N_e$$

Key insight: M is more sensitive to low-frequency variants (any mutation in ≥ 1 individual counts), while π is more sensitive to intermediate-frequency variants. For a biallelic variant with frequency p, the probability that two random samples differ is $2p(1-p)$, maximized at $p = 0.5$.

Summary of loss and gain scenarios

We examine genetic diversity changes for both π and M under different spatio-temporal scenarios:

Spatial scenarios: - **Scenario 1**: Single population declining (applicable to threatened species) - **Scenario 2**: Multiple populations with fraction of populations lost

Temporal scenarios: - **Scenario a**: Sudden immediate decline (direct mortality effect) - **Scenario b**: Loss due to demographic stochasticity over time (genetic drift)

Combined scenarios: 1a, 1b, 2a, 2b

DIVERSITY DECLINES

Loss scenario 1a: Immediate population contraction (no drift effects)

When a population contracts immediately from N_0 to N_1 within a single generation, we can model this as statistical subsampling.

Loss of π   For nucleotide diversity, the sampling effect creates a downward bias. Under the coalescent, when sampling n individuals from a finite population of size N, there's a probability $\frac{1}{2N}$ that samples coalesce to the same ancestor (yielding zero diversity). Additionally, finite sampling creates bias captured by $(1 - \frac{1}{n})$.

The relationship becomes:

$$\pi_1 = \pi_0 \cdot \left(1 - \frac{1}{2N_1}\right) \cdot \left(1 - \frac{1}{n}\right)$$

Key result: Loss is proportional to $\frac{1}{2N_1}$, so $X_\pi \approx \frac{1}{2N_1}$. This means: - Very small genetic diversity loss unless $N_1 < 40$ - Even at $N_1 = 40$: $X_\pi = 0.0125$ (1.25% loss) - π is largely unaffected by bottlenecks within one generation unless near-complete population eradication occurs

Loss of alleles M Since $M_0 \propto \log(N_0)$ under equilibrium, the fractional loss is:

$$X_M = 1 - \frac{M_1}{M_0} = 1 - \frac{\log(N_1)}{\log(N_0)}$$

Substituting $N_1 = N_0(1 - X_N)$:

$$X_M = 1 - \frac{\log(N_0(1 - X_N))}{\log(N_0)} = 1 - \frac{\log(N_0) + \log(1 - X_N)}{\log(N_0)}$$

$$X_M = -\frac{\log(1 - X_N)}{\log(N_0)}$$

Key insights: - Genetic diversity loss X_M is log-proportional to population size loss $\log(1 - X_N)$ - Species with larger initial populations lose smaller fractions of genetic diversity - **Example**: $N_0 = 10,000$, 50% reduction ($X_N = 0.5$) $\rightarrow X_M = 7.5$ allelic diversity loss

---

Loss scenario 1b: Effect of drift over time

Loss of π over time Changes in nucleotide diversity over time follow the classic formula:

$$\pi_t = \pi_0 \times \left(1 - \frac{1}{2N}\right)^t$$

The fractional loss after t generations:

$$X_\pi = 1 - \left(1 - \frac{1}{2N}\right)^t$$

Rearranging to solve for effective population size:

$$N = \frac{1}{2\left[1 - (1 - X_\pi)^{1/t}\right]}$$

Example: For $N = 100$ individuals: - $X_\pi = 0.5\%$ per generation - ~5% loss over 10 generations - This insight supports the conservation genetics recommendation of maintaining $N > 500$

Loss of alleles M over time The probability that an allele with frequency p_i is lost in one generation is $(1 - p_i)^{2N}$. The allelic richness after one generation:

$$M_1 = M_0 - \sum_i (1 - p_i)^{2N}$$

Simplified approximation: Discretizing frequency distribution into: - Rare alleles (<1%): high loss probability (>50%) - Uncommon alleles (1-10%): moderate loss probability
- Common alleles (>10%): very low loss probability (<0.1%)
Using the standard frequency distribution $f(p) \propto \theta/p$:

$$X_M \approx f(p < 0.01) \cdot (1 - \bar{p}_{\text{rare}})^{2N} + f(p_{0.01-0.1}) \cdot (1 - \bar{p}_{\text{uncommon}})^{2N} + f(p > 0.1) \cdot (1 - \bar{p}_{\text{common}})^{2N}$$

Example: 90% population decline ($N_0 = 1000 \rightarrow N_1 = 100$): - $X_M \approx 15\%$ allelic richness loss per generation
Over t generations, survival probability for allele with frequency p:

$$P_{\text{survival}}(p) = 1 - (1 - p)^{2N_1 t}$$

Loss scenario 2a: Spatial geographic contraction (immediate)

Loss of π with population structure   We use the fixation index F_{ST} (Nei's definition):

$$F_{ST} = \frac{\pi_{\text{total}} - \pi_{\text{within}}}{\pi_{\text{total}}}$$

where $\pi_{\text{total}} = \pi_{\text{within}} + \pi_{\text{between}}$.
For a species with k populations, total genetic diversity is:

$$\pi_{\text{total}} = \frac{1}{k}\pi_{\text{within}} + \frac{k-1}{k}\pi_{\text{between}}$$

We can express components in terms of F_{ST}:

$$\pi_{\text{within}} = (1 - F_{ST}) \times \pi_{\text{total}}$$

$$\pi_{\text{between}} = \frac{F_{ST}}{k-1} \times k \times \pi_{\text{total}}$$

After losing x populations, the new total diversity:

$$\pi'_{\text{total}} = \frac{1}{k-x}\pi_{\text{within}} + \frac{k-x-1}{k-x}\pi_{\text{between}}$$

The fractional change:

$$X_\pi = \frac{\pi'_{\text{total}}}{\pi_{\text{total}}} - 1$$

Substituting and simplifying:

$$X_\pi = \frac{1}{k-x}(1 - F_{ST}) + \frac{k-x-1}{k-x} \times \frac{k - (1 - F_{ST})}{k-1} - 1$$

Key insights: - Minimum populations ($k = 2$): $X_\pi = F_{ST}$ - As k increases, loss per population decreases - **Example**: $F_{ST} = 0.2$, 50% area contraction $\rightarrow X_\pi \approx 0.2\%$ to 20% (depending on $k = 2$ to 100)
Continuous space approximation: Using the Genetic Diversity-Area Relationship (GDAR):

$$\pi = cA^{z_{\text{GDAR}}}$$

where empirically $z_{\text{GDAR}} \approx 0.05$. This gives:

$$X_\pi = 1 - \left(\frac{A_1}{A_0}\right)^{z_{\text{GDAR}}} = 1 - (1 - X_A)^{z_{\text{GDAR}}}$$

Example: 50% area reduction $\rightarrow X_\pi \approx 3.4\%$

Loss of alleles M with spatial contraction   Using the Mutations-Area Relationship (MAR):

$$M = cA^{z_{\text{MAR}}}$$

where empirically $z_{\text{MAR}} \approx 0.25$. This gives:

$$X_M = 1 - (1 - X_A)^{z_{\text{MAR}}}$$

Example: 50% area reduction $\rightarrow X_M \approx 19\%$ allelic richness loss

---

Loss scenario 2b: Spatial contraction with drift over time

This scenario combines spatial extinction with ongoing demographic stochasticity. The temporal dynamics require sophisticated numerical methods (e.g., Wright-Fisher moments) and are beyond simple analytical solutions. These models are particularly relevant for:

- Non-threatened species with moderate to large geographic ranges
- Global within-species diversity loss (vs. classic conservation studies of extremely reduced populations)
- Long-term predictions of genetic diversity under habitat fragmentation

---

DIVERSITY INCREASES

Gain scenario 1b: Population growth over time

Starting from the classic equation, rearranged for expansion. $\pi_0 = 4N_0\mu$ (diversity before expansion). $\pi_\infty = 4N_1\mu$ (new equilibrium after expansion). π_t = diversity t generations after expansion
The trajectory follows:

$$\pi_t = \pi_\infty + (\pi_0 - \pi_\infty)\left(1 - \frac{1}{2N_1}\right)^t$$

Derivation of fractional change:
Define the population growth ratio: $r = \frac{N_1}{N_0}$
Then $\frac{\pi_\infty}{\pi_0} = \frac{4N_1\mu}{4N_0\mu} = r$
The fractional change from baseline:

$$X_t = \frac{\pi_t - \pi_0}{\pi_0}$$

Substituting and simplifying:

$$X_t = r + (1-r)\left(1 - \frac{1}{2N_1}\right)^t - 1$$

$$X_t = (r-1) + (1-r)\left(1 - \frac{1}{2N_1}\right)^t$$

Expressing in terms of population change $X_n = \frac{N_1 - N_0}{N_0} = r - 1$:

$$X_t = X_n\left[1 - \left(1 - \frac{1}{2N_1}\right)^t\right]$$

Or in terms of original population size:

$$X_t = X_n\left[1 - \left(1 - \frac{1}{2N_0(1 + X_n)}\right)^t\right]$$

Interpretation: The genetic diversity approaches the new equilibrium value exponentially, with rate depending on the expanded population size N_1.

Gain scenario 2a: Spatial expansion (F_{ST} based)

When populations expand spatially and come into contact, genetic diversity increases. The proportional increase can be derived from population differentiation, starting from: $F_{ST} = 1 - \frac{\pi_{\text{within}}}{\pi_{\text{total}}}$.
The relationship between diversity gain and population structure can be re-arranged then as:

$$X_\pi = \frac{F_{ST}}{1 - F_{ST}}$$

Conversely: $F_{ST} = \frac{X_\pi}{1+X_\pi}$.

Interpretation: Higher population differentiation leads to greater potential diversity gains upon population mixing or expansion.

REFERENCES

Exposito-Alonso M, Booker TR, Czech L, et al. Genetic diversity loss in the Anthropocene. Science. 2022;377(6613):1431-1435. doi:10.1126/science.abn5642

Mualim KS, Spence JP, Weiß C, Selmoni O, Lin M, Exposito-Alonso M. Genetic diversity loss in the Anthropocene will continue long after habitat destruction ends. bioRxiv. Published online October 22, 2024:2024.10.21.619096. doi:10.1101/2024.10.21.619096