

# Power and limitations of the mutations-area relationship to assess within-species genetic diversity targets for post-2020 Sustainable Development Goals

Moises Exposito-Alonso<sup>1,2,3\*</sup>, Jeffrey P. Spence<sup>4</sup>, Megan Ruffley<sup>1</sup>, Lucas Czech<sup>1</sup>, Meixi Lin<sup>1</sup>, Oliver Selmoni<sup>1,5</sup>, Lauren Gillespie<sup>1,6</sup>, Kristy S. Mualim<sup>1,2</sup>, Shannon Hateley<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 Genetics, Stanford University, Stanford, CA 94305, USA.

<sup>5</sup>Department of Embryology, Carnegie Institution for Science, Baltimore, MD 21218, USA.

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

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

## Summary

**To evaluate the United Nation’s preliminary post-2020 sustainable goals on protecting high levels of genetic diversity per species, Exposito-Alonso et al. (2022) proposed a new framework to predict a species’ loss of genetic diversity given its loss of habitat area. This method, called the mutations-area relationship (MAR), is analogous to the species-area relationship (SAR), often used to assess and design species diversity targets. To advise conservation practitioners, here we discuss the power of MAR, its limitations, and potential improvements.**

**Keywords:** extinction, genetic diversity, climate change, habitat loss, macrogenetics, biodiversity loss, mutations-area relationship.

## Arising from:

Exposito-Alonso, M., Booker, T. A., Czech, L., Gillespie, L., Hateley, S., Kyriazis, C. C., Lang, P., Leventhal, L., Nogues-Bravo, D., Pagowski, V., Ruffley, M., Spence, J. P., Toro Arana, S., Weiss, C. and Zess, Erin. 2022. “Genetic Diversity Loss in the Anthropocene.” *Science* 377 (6613): 1431–35. <https://doi.org/10.1126/science.abn5642>.

## Introduction

Populations of the same species possess varying levels of genetic diversity arising from a complex relationship between a variety of factors, including mutation rate, recombination rate, population bottlenecks, and others (Hamrick & Godt, 1996; Buffalo, 2021). This genetic diversity codes for phenotypic diversity with important consequences for within-species adaptation to different environments (Clausen *et al.*, 1941; Leimu & Fischer, 2008; Hereford, 2009; Bontrager *et al.*, 2021) including future global change threats (Hoffmann & Sgrò, 2011).

The field of conservation genetics has focused on monitoring genetic diversity in populations as a means to understand and prevent isolation between populations and drift and inbreeding in small populations (Allendorf *et al.*, 2022), especially for those species categorized as endangered in the International Union for Conservation of Nature (IUCN) Red List (IUCN.org). The new field of macrogenetics aims to summarize an ever expanding set of DNA datasets for many species (Leigh *et al.*, 2021), to design essential biodiversity variables, and to keep track of biodiversity preservation targets (Hoban *et al.*, 2020). However, these methods are based on experimental assessment of individuals or populations of a species, requiring field collections, DNA sampling, and sequencing, which are challenging to build at global scales and hard to conduct for many groups of species and in remote ecosystems.

One way to overcome these challenges is by building models that are built on theoretical understanding of how genetic diversity of a species is distributed in space. These models can be calibrated in well-documented species and used to extrapolate to those where gaps of data exist. One such predictive model is the mutations-area relationship (MAR) (Exposito-Alonso *et al.*, 2022) that describes the relationship between genetic diversity within a species (measured as the number of unique alleles or variable positions in the genome) and its habitat size. This is analogous to the species-area relationship (SAR), which has been used to extrapolate how many species may be lost in an ecosystem by area loss due to direct land-use transformations or climate change-driven alterations (Matthews *et al.*, 2021). Using the same principle, when MAR is calibrated for a species or a group of species, one may derive how much genetic diversity will be lost for a given amount of habitat loss in one or a group of species.

Below, we provide an overview of how to apply MAR and discuss its limitations and discuss future improvements. This information will aid in the correct interpretation of analyses and prevent over- or under-predictions of genetic diversity loss.

## MAR equation and application to conservation

For detailed methods, we refer readers to the original publication (Exposito-Alonso *et al.*, 2022). In brief, MAR is a power law function that follows:  $M = c A^z$ , where  $M$  is the number of DNA mutations (also called nucleotide polymorphisms, genetic variants, or alleles) among a group of individuals collected in a given geographic area  $A$ . The constant  $c$  depends on

several parameters including the base level of genetic diversity of the species analyzed, the DNA sequencing method, and other potential biases that are dataset- or species-dependent. The key parameter for inference is  $z$  (or  $z_{MAR}$  to distinguish it from the SAR), the relative increase in genetic diversity per increase of habitat area, which is robust to the biases captured by  $c$  and scale invariant (meaning  $z$  does not change with the spatial units of area  $A$ , or regions of the genome used to calculate  $M$ , as long as these are representative). The scaling parameter  $z_{MAR}$  may vary across species depending on certain species-level properties such as migration rates or geographic range sizes, or their mating system. Based on the small number of species that we have analyzed so far, the average  $z_{MAR}$  is approximately 0.3 and ranges from 0.1 to 0.8 across species (Exposito-Alonso *et al.*, 2022). In the absence of genomic data in a species, and without any other way of inference, a practitioner may then assume the average  $z_{MAR}$  of 0.3 to conduct predictions across large numbers of species and ecosystems, and could consider values of  $z_{MAR}$  between 0.1 and 0.8 to account for broad uncertainties.

When we know that the geographic range of a species has suffered a range reduction due to any habitat loss—for instance by land-use transformation, clear cuttings, fires, or climate change-driven mortality—we can use the  $z_{MAR}$  parameter to predict the fraction of genetic diversity lost due to such instantaneous geographic range reductions by rearranging MAR:

$$M_{loss} = 1 - (1 - A_{loss})^{z_{MAR}}$$

where  $M_{loss}$  is the fraction (or to obtain a percentage, multiply by 100) of genetic variation lost and  $A_{loss}$  is the fraction of habitat lost.

The MAR can be applied using proposed biodiversity indicators within the Global Biodiversity Framework (CBD, 2021) (for an extended discussion of indicators and examples see [(Hoban, 2022)]). Proposed indicators include the number of populations with a substantial number of individuals (indicator #1) and the number of populations lost (indicator #2) can be used to quantify the effective total geographic range of a species and the area loss. The last indicator (indicator #3) dealing with direct genetic monitoring by DNA sequencing populations over time could help validate and expand MAR models that also include long-term genetic drift in addition to short-term genetic losses from population losses (see below).

### **Reasons for uncertain genetic diversity loss estimations**

1. Scaling relationships such as MAR and its counterpart the species-area relationship (SAR) are approximate projections that should capture the average dynamics of loss of richness due to loss of habitat. Spatial heterogeneity in the distribution of genetic diversity creates noise in the estimates, depending on whether habitat loss occurs first in geographic regions with high or low genetic diversity.

2. Scaling relationships describe a phenomenon emerging from multiple eco-evolutionary forces interacting in space, but are not a mechanistic model of said forces (Harte *et al.*, 2009). We have derived the ranges of possible  $z_{MAR}$  values given such forces (migration, genetic drift, natural selection, mutation) (Exposito-Alonso *et al.*, 2022), but our understanding of what  $z$  values are expected in any species, given its life history traits, and geographic and migratory ranges is still limited.

3. The power law MAR method specifically refers to the metric of genetic diversity based on number of mutations (i.e. number of segregating sites or allelic richness); note this is equivalent to species richness in SAR at the genetic level. Other metrics of genetic diversity such as population heterozygosity ( $\pi$ ) behave more unpredictably and need further modeling. Therefore, using MAR for predicting genetic diversity loss of different types of metrics will create uncertainty.

### Reasons for overestimating genetic diversity loss

1. SAR has been proposed to overestimate species extinction (He & Hubbell, 2011; Rahbek & Colwell, 2011), for instance due to the way it is statistically constructed through spatial re-sampling: using outward vs inward vs random subsampling or focusing on endemisms (i.e. counting segregating sites that are private to the area subsample) (He & Hubbell, 2011; Rahbek & Colwell, 2011; Storch *et al.*, 2012). We tested these potential biases in MAR (see Supplemental Materials [(Exposito-Alonso *et al.*, 2022)]), and we confirmed relative consistency between MAR calculated with inward, outward, and random spatial sampling, and found that the endemic MAR (EMAR) may not be appropriate to study genetic diversity loss since EMAR does not reliably predict genetic diversity loss in our simulations.

2. Species may persist in altered habitats (Pereira & Daily, 2006). This is known to bias SAR predictions to overestimate species loss when area is lost, due to the habitat change being overestimated. To account for this, Pereira and Daily (2006) proposed the “countryside SAR”. If the tolerance of a species to an altered habitat was known, for example as a ratio,  $H$ , of abundance of individuals before and after an alteration, the MAR equation could be adjusted to:  $M_{loss} = 1 - (1 - A_{loss} \times H)^{z_{MAR}}$ . However, this partial tolerance to habitat loss is often unknown (but see an approximation based on diversity surveys in altered lands: [(Gallego-Zamorano *et al.*, 2022)]). This proposed countryside MAR has not yet been tested but could yield accurate estimates for species which persist despite major habitat alteration.

However, note that if the habitat transformation involves a population loss and subsequent expansion after transformation—e.g., re-colonization from a nearby population—the countryside MAR may not be appropriate because genetic diversity has been lost regardless and its recovery will still be slow through the gradual accumulation of new mutations.

3. In SAR, there is a scale dependence in the slope  $z$ , with slight increase in slope at large spatial scales (Storch *et al.*, 2012). Since power laws are typically fit with continental-scale datasets and used to predict local scale extinctions, predictions could be overestimated at

local scales. This bias could also exist in MAR, although its consequences are still not well understood. Exposito-Alonso *et al.* (2022) showed that there is a scale dependency of slope  $z_{MAR}$  with small number of genomes sampled, where  $z_{MAR}$  and therefore genetic diversity losses are overestimated for species with very few genetic samples (i.e. below 100 sampling sites). A correcting factor has been proposed but further research is needed.

Note: Supplemental Material in (Exposito-Alonso *et al.*, 2022) addresses these three points.

### **Reasons for underestimating genetic diversity loss**

1. The discovery of low frequency mutations is underpowered (Lockwood *et al.*, 2007). This bias in genomic data is extremely prevalent, as sequencing pipelines aim to be conservative and often filter out rare mutations as possible sequencing errors. Underestimating sites of variation in the genome would underestimate  $z_{MAR}$  and therefore the degree of genetic diversity loss associated with area shrinkage. Therefore, we strongly advise that low frequency mutations are not filtered out from sequencing datasets when estimating the MAR.
2. The correction factor accounting for scale-dependency for low sample sizes (section #3 in the previous section of potential overestimations) effectively scales down  $z_{MAR}$  to avoid upwardly biased estimates. However, we do not have a correction factor for downwardly biased estimates which may arise, for instance, from incomplete sampling across the geographic distribution of a species missing important genetic variation. This conservative strategy would generally lead to underestimates of genetic diversity loss, and more research is needed to understand such biases.
3. When habitat area is lost, not only are mutations within a species lost, but also entire species. When a species goes extinct, it loses 100% of its genetic diversity. Therefore, the loss of genetic diversity in an ecosystem will be driven by partial geographic range losses of some species and entire extinctions of others. If full species extinctions are not accounted for when predicting genetic diversity loss, total losses would be underestimated. We referred to this as “nested extinction” (Exposito-Alonso *et al.*, 2022).
4. Finally, one of the most dramatic causes of MAR underestimation is geographic area reductions leading to high genetic drift. Species with fragmented and smaller ranges have higher drift and thus lose more genetic diversity due to stochastic demographic processes. This means that after the immediate genetic diversity loss from area reduction, predicted by MAR, even when area loss is stopped, genetic diversity will lead to further genetic diversity loss over generations. However, predictive MAR models including this reactive long-term process are still unexplored.

### **Conclusion and outlook**

In summary, MAR is a fast and scalable tool to predict species- or ecosystem-wide genetic diversity loss in macrogenetic and global conservation studies. These predictions may be

most helpful to set data- and theory-driven global targets as part of the Sustainability Goals from United Nations' Convention for Biological Diversity (CBD, 2021), or to rapidly assess potential habitat loss impacts for any given species.

Ultimately, to make accurate predictions of genetic diversity loss and increased extinction risk of species, very detailed data and expert assessment per species will be required: census sizes, genome sizes, migration in metapopulations, mating systems, detailed maps of genetic makeups, and area transformations. Recently there have been attempts to employ such a data-driven approach for a fraction of Red List species in Sweden (Thurfjell *et al.*, 2022) and nine additional other countries (Hoban, 2022). To enable the production of new genomic datasets across entire ecosystems is critical—especially in continents and ecosystems that are underrepresented such as the tropics and the Global South (Marks *et al.*, 2021)—and will help continue efforts to create maps of genetic diversity (Parks *et al.*, 2013; Miraldo *et al.*, 2016; Li *et al.*, 2021).

We envision that a combination of genome-wide DNA sequencing (i.e. reduced-representation-, whole-genome-, or pool-sequencing) and theory to conduct high-confidence extrapolations could be most promising, where for data-rich species mechanistic models projected forward-in-time (e.g. using SLiM simulations [(Haller & Messer, 2019)]) could be used to calibrate MAR functions, and MAR-based extrapolation could be calibrated to fill in data and prediction gaps for other species and ecosystems.

Given that SAR is already applied in global conservation policies (IPBES, 2019), and that no other statistical approaches exist to project genetic diversity losses within species, we expect the MAR approximation to be widely used. This perspective article intends to make conservation practitioners aware of the power and potential limitations of this method, which we will expand as evidence accumulates.

### **Acknowledgments**

We are grateful to members of the Moi Lab, Tad Fukami, Henrique Pereira, for comments or discussions on this research. This research is supported by the Office of the Director of the National Institutes of Health's Early Investigator Award with award # 1DP5OD029506-01; by the U.S. Department of Energy, Office of Biological and Environmental Research, grant # DE-SC0021286; by a USDA NIFA, grant # 2022-67019-36366; and by the U.S. National Science Foundation's DBI grant # 2213983 (Water and Life Interface Institute (WALII)); and by the Carnegie Institution for Science.

### **Authors contributions**

All authors discussed and wrote the manuscript.

### **Competing Interest Statement**

The authors have declared no competing interest.

## References

- Allendorf FW, Chris Funk W, Aitken SN, Byrne M, Luikart G. 2022. *Conservation and the Genomics of Populations*. Oxford University Press.
- Bontrager M, Usui T, Lee-Yaw JA, Anstett DN, Branch HA, Hargreaves AL, Muir CD, Angert AL. 2021. Adaptation across geographic ranges is consistent with strong selection in marginal climates and legacies of range expansion. *Evolution; international journal of organic evolution* **75**: 1316–1333.
- Buffalo V. 2021. Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain Lewontin’s paradox. *eLife* **10**.
- CBD. 2021. 1st Draft of The Post-2020 Global Biodiversity Framework.
- Clausen J, Keck DD, Hiesey WM. 1941. Regional Differentiation in Plant Species. *The American naturalist* **75**: 231–250.
- Exposito-Alonso M, Booker TR, Czech L, Gillespie L, Hateley S, Kyriazis CC, Lang PLM, Leventhal L, Nogues-Bravo D, Pagowski V, *et al.* 2022. Genetic diversity loss in the Anthropocene. *Science* **377**: 1431–1435.
- Gallego-Zamorano J, Huijbregts MAJ, Schipper AM. 2022. Changes in plant species richness due to land use and nitrogen deposition across the globe. *Diversity & distributions* **28**: 745–755.
- Haller BC, Messer PW. 2019. SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Molecular biology and evolution* **36**: 632–637.
- Hamrick JL, Godt MJW. 1996. Effects of life history traits on genetic diversity in plant species. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **351**: 1291–1298.
- Harte J, Smith AB, Storch D. 2009. Biodiversity scales from plots to biomes with a universal species-area curve. *Ecology letters* **12**: 789–797.
- He F, Hubbell SP. 2011. Species-area relationships always overestimate extinction rates from habitat loss. *Nature* **473**: 368–371.
- Hereford J. 2009. A quantitative survey of local adaptation and fitness trade-offs. *The American naturalist* **173**: 579–588.
- Hoban S. 2022. Monitoring status and trends in genetic diversity for the Convention on Biological Diversity: an ongoing assessment of genetic indicators in nine countries. *Authorea Preprints*.
- Hoban S, Bruford M, D’Urban Jackson J, Lopes-Fernandes M, Heuertz M, Hohenlohe PA, Paz-Vinas I, Sjögren-Gulve P, Segelbacher G, Vernesi C, *et al.* 2020. Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. *Biological conservation* **248**: 108654.
- Hoffmann AA, Sgrò CM. 2011. Climate change and evolutionary adaptation. *Nature* **470**: 479–485.
- IPBES. 2019. *Global Assessment Report on Biodiversity and Ecosystem Services* (E. S. Brondizio, J. Settele, S. Díaz, and H. T. Ngo, Ed.). IPBES Secretariat, Bonn, Germany.
- Leigh DM, van Rees CB, Millette KL, Breed MF, Schmidt C, Bertola LD, Hand BK, Hunter

- ME, Jensen EL, Kershaw F, et al. 2021.** Opportunities and challenges of macrogenetic studies. *Nature reviews. Genetics* **22**: 791–807.
- Leimu R, Fischer M. 2008.** A meta-analysis of local adaptation in plants. *PloS one* **3**: e4010.
- Li C, Tian D, Tang B, Liu X, Teng X, Zhao W, Zhang Z, Song S. 2021.** Genome Variation Map: a worldwide collection of genome variations across multiple species. *Nucleic acids research* **49**: D1186–D1191.
- Lockwood DR, Richards CM, Volk GM. 2007.** Probabilistic models for collecting genetic diversity: Comparisons, caveats, and limitations. *Crop science* **47**: 861–866.
- Marks RA, Hotaling S, Frandsen PB, VanBuren R. 2021.** Representation and participation across 20 years of plant genome sequencing. *Nature plants* **7**: 1571–1578.
- Matthews TJ, Triantis KA, Whittaker RJ. 2021.** *The Species-Area Relationship: Theory and Application* (TJ Matthews, KA Triantis, and RJ Whittaker, Eds.). Cambridge University Press.
- Miraldo A, Li S, Borregaard MK, Flórez-Rodríguez A, Gopalakrishnan S, Rizvanovic M, Wang Z, Rahbek C, Marske KA, Nogués-Bravo D. 2016.** An Anthropocene map of genetic diversity. *Science* **353**: 1532–1535.
- Parks DH, Mankowski T, Zangooei S, Porter MS, Armanini DG, Baird DJ, Langille MGI, Beiko RG. 2013.** GenGIS 2: Geospatial Analysis of Traditional and Genetic Biodiversity, with New Gradient Algorithms and an Extensible Plugin Framework. *PloS one* **8**: e69885.
- Pereira HM, Daily GC. 2006.** Modeling biodiversity dynamics in countryside landscapes. *Ecology* **87**: 1877–1885.
- Rahbek C, Colwell RK. 2011.** Biodiversity: Species loss revisited. *Nature* **473**: 288–289.
- Storch D, Keil P, Jetz W. 2012.** Universal species-area and endemics-area relationships at continental scales. *Nature* **488**: 78–81.
- Thurfjell H, Laikre L, Ekblom R, Hoban S, Sjögren-Gulve P. 2022.** Practical application of indicators for genetic diversity in CBD post-2020 Global Biodiversity Framework implementation. *bioRxiv*: 2022.02.18.481087.