The Community Genetic Distribution (CGD): A unifying measure for monitoring biodiversity change

Isaac Overcast[1,2,3,4], Irene Calderon-Sanou[IRS], Simon Creer[SC], Virginia Dominguez-Garcia[VDG], Oskar Hagen[OH], Michael J Hickerson[MH], Theresa Jörger-Hickfang[TJH], Henrik Krehenwinkel[HK], Jay T. Lennon[JL], Laura Méndez[LM], Maria Méndez[MM], Renske E. Onstein[RO], Henrique Pereira[HP], Clara Qin[CQ], Douglas W. Yu[DY], Damaris Zurell[DZ], Rosemary Gillespie[RG]

[1] Institute of Biodiversity Science & Sustainability, California Academy of Sciences, San Francisco, CA, USA
[2] Department of Vertebrate Zoology, American Museum of Natural History, New York, New York, USA
[3] Virginia Museum of Natural History, Martinsville, VA, USA
[4] Illinois Natural History Survey, University of Illinois Urbana-Champaign, Champaign, IL, USA
[IRS] UMR EcoFoG (AgroParistech, CIRAD, CNRS, INRAE, Université des Antilles, Université de la Guyane), Kourou, France
[SC] Molecular Ecology and Evolution Group at Bangor, School of Environmental and Natural Sciences, Bangor University, UK
[VDG] Estación Biológica de Doñana (EBD-CSIC), Seville, Spain
[OH] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig
[OH]Center for Critical Computational Studies (C³S), Goethe-Universität Frankfurt, Theodor-W.-Adorno-Platz 1, 60323 Frankfurt am Main, Germany
[MH] Biology Department, City College of New York, New York, USA
[MH] Biology Ph.D. Program, Graduate Center, City University of New York, New York, USA
[MH] Division of Invertebrate Zoology, American Museum of Natural History, New York, New York, USA
[TJH] German Centre for Integrative Biodiversity Research (Halle-Jena-Leipzig), Leipzig, Germany
[TJH] Institute of Biology, Martin Luther University, Halle-Wittenberg, Halle (Saale), Germany
[HK] Department of Biogeography, Trier University, Trier, Germany
[JTL] Department of Biology, Indiana University, Bloomington, Indiana, USA
[LM] Department of Community Ecology, Helmholtz Centre for Environmental Research (UFZ), Halle (Saale), Germany
[LM] German Centre for Integrative Biodiversity Research (iDiv) Leipzig-Halle-Jena, Leipzig, Germany
[LM] Philipps-University Marburg, Marburg, Germany
[MM] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
[MM] Institute of Biology, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany
[RO] Naturalis Biodiversity Center, Leiden, Netherlands
[RO] Leiden University, Institute for Biology Leiden (IBL), Leiden, Netherlands
[HP] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
[HP] Institut für Biologie, Martin-Luther-University Halle-Wittenberg, Halle, Germany

[HP] CIBIO (Research Centre in Biodiversity and Genetic Resources)–InBIO (Research Network in Biodiversity and Evolutionary Biology), Universidade do Porto, Vairão, Portugal
[CQ] Society for the Protection of Underground Networks, Dover, DE
[DY] Yunnan Key Laboratory of Biodiversity and Ecological Conservation of Gaoligong Mountain, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China
[DY] School of Biological Sciences, University of East Anglia, Norwich, UK
[DZ] Ecology and Macroecology, Institute for Biochemistry and Biology, University of Potsdam, Potsdam, Germany
[RG] Environmental Science, Policy, and Management, University of California Berkeley, CA

*Corresponding author (present address): Isaac Overcast, Department of Ecology, Evolution, and Environmental Biology, Columbia University, New York, NY 10027. Email: iao2122@columbia.edu

Running headline: The distribution of genetic diversity in ecological communities

Abstract

Monitoring the condition of ecological communities is essential to understanding, managing and conserving biodiversity. Much needed is a means to measure holistic properties that emerge from ecological communities, in other words attributes that characterize communities as a whole rather than individual species or sets of species. Here, we propose the Community Genetic Distribution (CGD) as a novel metric that captures these emergent properties of ecological communities by integrating genetic data from a comprehensive sample of the community. The CGD compliments existing taxon-based metrics, focusing on properties that emerge from the interacting set of species within a given local community; as such, it is unifying across taxa, economically scalable, and geographically transferable. Using high-throughput biodiversity genetic inventories, the CGD captures information about past ecological and evolutionary dynamics that have shaped the local community, whether through local co-evolution and speciation within an isolated community, dispersal and ecological filtering from a larger co-evolved regional pool, or recent colonization from a novel anthropogenic pool. Here, we present empirical case studies demonstrating how CGD, derived from metabarcoding datasets, varies in space and time in response to human disturbances, and effectively captures patterns associated with restoration and succession gradients. These insights enhance our understanding of the community-level genetic condition and help develop reliable indicator metrics to assess ecological and evolutionary degradation. The CGD complements existing essential biodiversity variables and provides the basis for developing novel indicators of community status and integrity, and can thus directly contribute to the targets of the Kunming-Montreal Global Biodiversity Framework.

Keywords: Biodiversity, community ecology, metabarcoding, environmental DNA, essential biodiversity variables, eco-evolutionary modeling

1 Introduction

Biodiversity is a complex fabric, with the presence of any one species variably connected, directly or indirectly, with all others in a local community. At the same time, a major scientific and policy challenge in ecology and conservation is to find metrics of biodiversity that represent this complexity, yet are simultaneously simple and understandable (Humphries et al., 1995; Purvis & Hector, 2000). The GEOBON Essential Biodiversity Variables (EBVs) represent a coordinated effort to standardize the collection, sharing, and use of biodiversity data (Pereira et al., 2013), contributing to the development of headline indicators for the Kunming-Montreal Global Biodiversity Framework (on Biological, n.d.). However, like the EBVs, most existing metrics focus on patterns at single levels of biological organization, for example at the species level (e.g. within species genetic diversity), or the community level (e.g. species abundance distributions). There is thus a growing need for biodiversity metrics that integrate multiple levels of biological organization, potentially better capturing interactions between ecological and evolutionary processes. Here, we propose the Community Genetic Distribution (CGD) as a novel biodiversity metric that leverages genetic data and integrates genetic diversity at both the within-species and among-species levels of organization. The CGD quantifies the distribution of intraspecific genetic variation across taxa within a community sample in a way that is conceptually identical to the construction of a Species Abundance Distribution (Matthews & Whittaker, 2015; Preston, 1948), while focusing on genetic diversity rather than counts of individuals. The CGD leverages historical population size change information contained in genetic data, reflecting both ecological and evolutionary processes that have led to the assembly of species. For example, whether the species have a history of local co-evolution, have colonized from a regional pool with co-evolution at a larger scale, or whether they are recent arrivals from a newly formed anthropogenic pool. Thus, the CGD provides insights into the organization and the long-term stability (Formoso-Freire et al., 2025), or instability, of a given local community and as such, it can provide valuable insights for research, policy, and conservation.

The CGD is constructed by calculating nucleotide diversity (π) (Nei & Li, 1979) within species (or operational taxonomic units [OTUs]), aggregating these values across all species within an ecological community sample, and analyzing their distribution (Fig. 1). By unifying historical information contained in genetic data with the ecological context of the community-scale data, the CGD provides a method for capturing information on the fundamental processes that shape ecological communities (i.e. dispersal, speciation, ecological drift, environmental filtering and species interactions; Vellend, 2016)). The CGD captures these processes by integrating information that impacts both community ecology and evolutionary history, making use of recent advances in molecular methodologies, notably high-throughput sequencing that can readily capture genetic signatures from every individual sampled to represent the composition of the community (Overcast, et al., 2023). Specifically, the distribution of intraspecific genetic diversity across taxa in a quantitative sample of individuals within a community can indicate the extent to which these taxa have shared a history at that site (Figure 1). In several important ways the CGD is analogous to the Species Abundance Distribution, but rather than characterizing the distribution of species abundances, it characterizes the distribution of both intra- and inter-specific genetic diversity across all species or OTUs in a community.

In this perspective, we outline the potential conservation utility of this novel biodiversity measure with the goal of centering community genetic diversity as an emergent property of an assemblage, which provides insight into the integrity and resilience of ecological communities. In the face of growing imperatives to monitor biodiversity as a “fabric” (Díaz et al., 2019) across taxonomic levels of organization and at multiple spatial and temporal scales, we argue that the community genetic distribution provides a rapid, cost-effective, and information-rich means for inventorying and monitoring biodiversity.

2 Advantages of Community Genetic Distributions for Conservation

2.1 Ecological communities as units of conservation

Historically, biodiversity assessment and conservation efforts have focused on individual species, for example through The Endangered Species Act in the US (16 U.S.C. § 1531) or the International Union for Conservation of Nature Red List (IUCN; iucnredlist.org) of Threatened Species. However, it is well understood that species are embedded within and depend on the complex biological matrix of their surrounding ecosystems (Díaz et al., 2019). This recognition has prompted the development of a suite of assessment metrics and tools above the species level, including several headline Indicators for the Kunming-Montreal Global Biodiversity Framework (GBF, 2024; e.g. Indicators A.1 IUCN Red List of Ecosystems and A.2 Extent of natural ecosystems). Focusing on conservation of the integrity of communities shifts the focus to the co-evolved properties of a community, including the interactions and dependencies that sustain both species of concern and those of least concern (for the moment). Focusing on the properties that emerge from entire communities also provides a means to identify stability, and hence infer continued delivery of ecosystem services in the face of environmental change and other anthropogenic disturbance. A key goal of the CGD is to leverage information that has historically been difficult to incorporate into biodiversity monitoring programs, particularly data that reflect community-level processes and long-term ecological dynamics.

2.2 Cost effective and efficient to collect with high throughput DNA surveys

Observing, identifying, and quantifying the enormity of global biodiversity has been a long-standing problem in monitoring biodiversity change. This task has become more achievable with the advent of high-throughput sequencing, including whole community metabarcoding and eDNA (Emerson et al., 2023; Ficetola & Taberlet, 2023; Taberlet et al., 2012). Metabarcoding in particular, based on the massive parallel sequencing of short and taxonomically informative DNA amplicons (DNA barcodes) to characterize diverse biological communities (Yu et al., 2012), is well suited to explore hyperdiverse ecosystems with little or no reference data available. Bulk community metabarcoding is commonly used with actual community samples, for example all insects from a Malaise trap (Gibson et al., 2014) or marine benthic communities (Leray & Knowlton, 2015), while environmental DNA (eDNA) metabarcoding uses DNA traces that organisms leave behind in their environment and has gained popularity as an entirely non-invasive approach to detect species and characterize the biotic composition of ecosystems (Bodawatta et al., 2025; Deiner et al., 2017). These approaches provide insights to the different facets of biodiversity, from genes to species to entire ecosystems (Krehenwinkel et al., 2022). For the purpose of constructing the CGD, metabarcoding data is ideal and has already been used in the recovery of community-wide patterns of genetic diversity (Overcast et al., 2021; Overcast, et al., 2023). Although metabarcoding is not free of flaws, such as taxonomic biases in PCR amplification efficiency, sequencing errors, or the co-amplification of paralogs of the barcode marker gene (see Hartig et al., 2024), it allows rapid and straightforward assessment of the composition of an entire assemblage of species (Graham et al., 2023). Moreover, approaches are continuously being refined to ensure that the genetic data mirrors the actual composition of taxa in a given community in terms of both diversity and abundance. Thus, while the inherent biases during the laboratory procedures, related to the quality and quantity of the DNA extractions or the biases in PCR amplifications precludes the estimation of absolute species abundances, metabarcoding datasets can provide an approximation of relative abundances of individual species in a community (Krehenwinkel et al., 2017; Lim et al., 2022; Luo et al., 2023).

2.3 Quantifying total biodiversity, including undescribed taxa

Biodiversity monitoring efforts, such as those led by the International Barcode of Life (iBOL) Consortium, use large-scale barcoding and metabarcoding to identify species through DNA sequencing. These approaches rely on reference databases to match DNA fragments (amplicons) to known taxa (Ratnasingham & Hebert, 2007), but such databases face challenges related to completeness, resolution, and biases in geographic and taxonomic coverage. Creating comprehensive reference databases is time-consuming and costly, and most described species remain unsequenced (e.g. (Moraes Zenker et al., 2024), as are the vast majority of cryptic, unknown and undescribed taxa, limiting the effectiveness of metabarcoding for characterizing the diversity of whole communities, in particular in the tropics. Additionally, taxonomic instability complicates database reliability, as names and classifications frequently change. Given these issues, the CGD, by design, does not require taxonomic information at all (it is 'taxonomy agnostic'), though it can be enhanced by a complete barcode sequence reference database. The ability to quantify the distribution of genetic variation within communities, without reference to taxonomy, greatly expands the scope of its utility by applying in equal measure to well-studied and fully described communities, and those that are undescribed and understudied (i.e. those that are most in need of quantification and assessment).

2.4 Information rich

The CGD contains information at multiple levels of biological organization, providing insights into the assembly history of a community, including its structure, historical population dynamics, and connectivity of component species across the landscape. At the community level, in the same way as SAD (Matthews & Whittaker, 2015), the CGD provides an estimate of species richness, represented by the number of OTUs within the focal community, and uncertainty in the estimation of species richness can be quantified using rarefaction curves

(Leray & Knowlton, 2015; Lim et al., 2022). At the species level, nucleotide diversity serves as a robust summary statistic of intraspecific genetic diversity which, when aggregated across species within a community, provides a snapshot of the community on a population genetic timescale. Unlike other biodiversity metrics that fluctuate with short-term changes in species abundances (e.g., ecological drift), the CGD reflects longer-term patterns of community stability/condition and turnover, providing a complementary dimension to traditional community ecology metrics

2.5 Efficient sampling properties of the CGD

The effect of sampling effort on ecological and biodiversity inference has been a focus of concern for decades (Martinez et al., 1999; Moreno & Halffter, 2001). Insufficient sampling can lead to biased conclusions, usually because of undersampling rare species. Ecological inventories focused on species richness may use rarefaction curves to quantify the uncertainty of species richness estimates, and similar techniques can be used to standardize sampling effort for abundance surveys. Rarefaction will necessarily underestimate biodiversity patterns and may additionally obscure true differences in ecological processes underlying such patterns. Rather than the total number of individuals in a sample, information in the CGD relies on sampling a sufficient number of DNA sequences. Nucleotide diversity is classically robust to sampling, requiring only 5-10 (haploid) samples for a given species to capture a reliable estimate of within-species genetic diversity at a given genetic locus (Tajima, 1983). This makes the CGD less sensitive to sampling effort than metrics like the SAD, for example. Furthermore, the probability of sampling a sufficient number of haplotypes for a given species is dramatically increased by using metabarcoding/eDNA studies, with typical samples containing tens of thousands of DNA sequences or more. This benefit also applies to large-scale barcoding approaches, such as 'megabarcoding' studies, where bulk samples (e.g. arthropods; Kitson et al., 2018) are first sorted into morphospecies and selected vouchers are then sequenced individually, rather than sequencing the entire batch, saving effort and cost while retaining information necessary for constructing a reliable CGD.

2.6 Responds on a timescale that is useful for conservation

Biodiversity measures that reflect the stability of a local assemblage over intermediate timescales – typically years to decades, over which changes in biodiversity resulting from conservation/remediation efforts are expected to become detectable – are particularly useful for monitoring change over time and hence support evidence-based decision-making. Other community-level measures of large-scale environmental disturbance, in particular SADs, can vary substantially over short time periods (Dornelas et al., 2018; Magurran, 2007). In addition, measures such as phylogenetic diversity (PD) focus on processes that occur over extended timescales (and the focus of PD is on a biogeographic area rather than a local community). As a long-term average of effective population size, the genetic diversity contained within a community carries information on intermediate timescales, and so the CGD is largely insensitive to short term fluctuations in community state.

3 Empirical conservation utility of CGDs

The CGD offers a sensitive metric for detecting the ecological impacts of environmental disturbances (e.g., habitat destruction, fires, droughts), as well as evaluating the effectiveness of conservation measures (e.g., protected areas, translocations) (Reynolds et al., 2012). To illustrate the conservation relevance and responsiveness of the CGD, we reanalyzed datasets from four previously published case studies that used whole-community quantitative metabarcoding data (Figure 2). In each case we downloaded the published data from publicly available repositories, reproduced the original bioinformatic analysis, and calculated and plotted the CGD - a pattern that was not investigated in any of the case studies. We tested different aspects of the CGD. First, we investigated whether the CGD can reflect co-evolutionary history, using spider communities across the Hawaiian archipelago (Graham et al., 2023). Here we observed a clear pattern of increasing CGD with increasing age of the community, as reflected by the geological age of the different islands, and consistent with the expected increase in co-evolutionary history of these communities over time. Second, we examined whether habitat type and fragmentation can influence the CGD using arthropod communities from five forest habitats on the Troodos mountain range in Cyprus (Noguerales et al., 2021). We found a strong effect, with arthropod communities from more geographically wide-spread habitats showing increased CGD, potentially as an effect of reduced environmental constraint and increased connectivity. Finally, we tested the extent to which the CGD might detect anthropogenic disturbance, using terrestrial arthropod communities from different islands (Kennedy et al., 2023) and marine metazoan communities from coastal South Africa (Holman et al., 2021). In both terrestrial and marine cases, we observed a strong and consistent reduction in CGD at disturbed sites, highlighting its potential as a reliable indicator of ecological degradation. Jupyter notebooks for reproducing all analyses are provided in the GitHub repository (https://github.com/isaacovercast/IMEMEBA-BCI).

Quantitative evaluation of the change in CGD shape can be obtained by calculating summary statistics on the distributions and contrasting these between community samples from comparable sites. Standard community ecological tools for summarizing SADs can be readily applied to the CGD with similar behaviors and interpretations (Magurran, 1988). For example, facets of the heterogeneity of genetic diversity values within a community can be quantified with indices derived from information theory, such as Shannon diversity (the first-order Hill number (Hill, 1973), and genetic diversity evenness. For community-scale genetic data, Shannon diversity provides a measure of species diversity that is sensitive to within-species genetic variation (Sherwin et al., 2017). Evenness is simply Shannon diversity scaled by species richness, and in this context it expresses the similarity in genetic diversity among all species in the community. Within a given empirical system, the combination of these two measures can detect the impact on community genetic diversity of ecologically relevant processes like habitat availability and environmental disturbance (Fig. 3).

4 Conservation utility of the CGD: Power, potential & provisos

4.1 Potential relevance to policy frameworks

Community genetic diversity captures information that is relevant across a range of domains: Scientists care about the stability of ecosystems, which is in turn related to resilience/resistance, and intactness (Van Meerbeek et al., 2021); Local, regional, and national governments care about ecosystem services and human health and wellbeing; Businesses care about reducing overhead, increasing revenue, and managing their public image. As a novel biodiversity metric, the CGD offers a means for designing rapid, cost-effective, and information-rich biodiversity indicators suitable for implementation within the corporate sustainability reporting directive (CSRD European Commission, 2022), global reporting initiative (*GRI - Home*, n.d.), Science-Based Targets Network (Biodiversity, 2020). Thus, the CGD can inform existing regulatory frameworks that require the continuous monitoring of biodiversity over time. While EBVs have been proposed for these frameworks, the CGD may reflect stability of an entire community because it captures demographic and historical processes shared across taxa - information not captured by existing EBVs.

Because the CGD provides operationalization of community stability and intactness, one application would be to quantify biodiversity so that it can be purchased in biodiversity credit markets (Ducros & Steele, 2022; Wunder et al., 2024). The purpose of such markets is to use private finance to compensate landowners for the direct and opportunity costs of carrying out restoration and conservation projects. Companies purchase biodiversity credits in order to offset negative impacts or to effect positive impacts on biodiversity. The potential benefits of such reporting include improved access to finance, reduced physical, litigation and regulatory risks, and enhanced corporate reputation. However, making biocredit markets work requires a difficult transformation: reducing the high-dimensional complexity of biodiversity to one-dimensional credits representing units of biodiversity status or change (Wauchope et al., 2024). It is also necessary to verify biodiversity improvements over time, because the value of the credits to buyers is reputational. The problem is that verification involves high transaction costs that reduce the funds available for conservation action. For example, in the Wallacea Trust (2023) scheme, one credit per hectare is earned for each 1% step towards convergence with a counterfactual reference site. Convergence is measured as the median percentage change in a suite of five or more metrics with a determined set of properties that may be chosen on a per project basis, one of which must measure habitat structural complexity and the rest the abundances of functionally important taxa (e.g. pollinators), with higher weights given to threatened species. All the taxon metrics require costly repeated measurements in the field. Similar basket-of-metrics approaches are used for other biodiversity credits, with different trade-offs between cost, comprehensibility, and credibility (Wunder et al., 2024). The potential advantage of a future CGD-derived metric is that it could reduce transaction costs by providing a single, direct, and efficient measure of the underlying goal of all biocredit projects, which is to conserve or restore ecosystems.

4.2 Technical caveats

Given the focus on capturing information from the entire community, the vast majority of metabarcoding studies target single-locus protein coding sequences (e.g. cytochrome oxidase (COI) for metazoa; Andújar et al., 2018). By construction, the CGD is not limited to single locus datasets; it is only a matter of practicality and what is currently readily available. At the same time, it is important to note some limitations specific to single-locus metabarcoding that should be taken into consideration when applying the CGD to these kinds of empirical datasets. Without a carefully curated reference database, overmerged haplotypes from distinct species would generate chimeric OTUs with inflated nucleotide diversity, potentially skewing the CGD. Further technical limitations include bias in amplification among all members of a community, amplification of nuclear pseudogenes (Andújar et al., 2020), and issues of contamination, though these are currently being addressed with more refined methodological approaches (Krehenwinkel et al., 2024). Additionally, there is some disagreement about the extent to which nucleotide diversity within single-locus mitochondrial protein coding genes (e.g. COI) corresponds to genome-wide nuclear diversity (Schmidt & Garroway, 2021, but see Allio et al., 2017). Bearing this in mind, conservation implications of the CGD should focus on how nucleotide diversity from single-locus markers may serve as a rough proxy for long-term effective population size, while recognizing that it does not reliably reflect genome-wide adaptive capacity. Looking to the future, some limitations of single-locus metabarcoding data can be overcome by leveraging multi-locus metabarcode approaches (Weitemier et al., 2021), RADSeq (Andrews et al., 2016), and/or whole genome sequencing (i.e. 'metagenomics'; Sleator et al., 2008). In principle, the CGD can be calculated using such expanded genetic datasets, providing a more robust picture of intraspecific genetic diversity across the community.

4.3 Context and need for comparison

As shown in the above examples, the CDG requires a comparative framework in order to measure shifts – changes across gradients in the same approximate area, and across space and/ or time. Importantly, in its current framing, the CGD cannot provide an absolute measure of stability or integrity of a community. For example, it cannot tell us the relative stability of the fish in a lake in Amazonia relative to arthropods in a forest in Cyprus; rather, it can tell us (for example, how different levels of pollution in the Amazonian lakes have affected community assembly patterns and hence composition of the Amazonian fish; or how arthropod change in response to shifts in forest composition (Fig. 3). In the future we hope to develop the CGD as a standalone measure that will provide information on the co-evolved properties of a community overall.

5 Concluding Remarks

The distribution of genetic variation in ecological communities is an informative yet largely unstudied biodiversity pattern with significant conservation relevance. It could be particularly useful for improving the functioning and oversight of the biodiversity finance and impact-disclosure policy instruments that have been proposed in Target 19 of the Kunming-Montreal Global Biodiversity Framework, as well as being generally useful as research variables in ecology and conservation. Beyond the advantages of studying the CGD that we have enumerated, the data necessary to construct the CGD already exists in thousands of published metabarcoding and eDNA studies and simply awaits to be unlocked.

8 References

Allio, R., Donega, S., Galtier, N., & Nabholz, B. (2017). Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: Implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Molecular Biology and Evolution*, *34*(11), 2762–2772.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews. Genetics*, *17*(2), 81–92.

Andújar, C., Arribas, P., Yu, D. W., Vogler, A. P., & Emerson, B. C. (2018). Why the COI barcode should be the community DNA metabarcode for the metazoa. *Molecular Ecology*, *27*(20), 3968–3975.

Andújar, C., Creedy, T. J., Arribas, P., López, H., Salces-Castellano, A., Pérez-Delgado, A., Vogler, A. P., & Emerson, B. C. (2020). Validated removal of nuclear pseudogenes and sequencing artefacts from mitochondrial metabarcode data. In *Cold Spring Harbor Laboratory* (p. 2020.06.17.157347). https://doi.org/10.1101/2020.06.17.157347

Biodiversity. (2020, May 5). Science Based Targets Network. https://sciencebasedtargetsnetwork.org/about/hubs/biodiversity/

Bodawatta, K., Madsen, A., Holman, L., Rasmussen, J., Johnson, M., Frøslev, T., Lynggaard, C., & Bohmann, K. (2025). *Advanced airborne eDNA sampling allows robust spatiotemporal characterisation of vertebrate communities*. https://doi.org/10.21203/rs.3.rs-5693469/v1

CSRD European Commission. (2022). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022L2464

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., de Vere, N., Pfrender, M. E., & Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, *26*(21), 5872–5895.

Díaz, S., Settele, J., Brondízio, E. S., Ngo, H. T., Agard, J., Arneth, A., Balvanera, P., Brauman, K. A., Butchart, S. H. M., Chan, K. M. A., Garibaldi, L. A., Ichii, K., Liu, J., Subramanian, S. M., Midgley, G. F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., … Zayas, C. N. (2019). Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science (New York, N.Y.)*, *366*(6471), eaax3100.

Dornelas, M., Antão, L. H., Moyes, F., Bates, A. E., Magurran, A. E., Adam, D., Akhmetzhanova, A. A., Appeltans, W., Arcos, J. M., Arnold, H., Ayyappan, N., Badihi, G., Baird, A. H., Barbosa, M., Barreto, T. E., Bässler, C., Bellgrove, A., Belmaker, J., Benedetti-Cecchi, L., … Hickler, T. (2018). BioTIME: A database of biodiversity time series for the Anthropocene. *Global Ecology and Biogeography: A Journal of Macroecology*, *27*(7), 760–786.

Ducros & Steele. (2022). *Biocredits to finance nature and people*. iied.org. https://www.iied.org/sites/default/files/pdfs/2022-11/21216IIED.pdf

Emerson, B. C., Borges, P. A. V., Cardoso, P., Convey, P., deWaard, J. R., Economo, E. P., Gillespie, R. G., Kennedy, S., Krehenwinkel, H., Meier, R., Roderick, G. K., Strasberg, D., Thébaud, C., Traveset, A., Creedy, T. J., Meramveliotakis, E., Noguerales, V., Overcast, I., Morlon, H., … Andújar, C. (2023). Collective and harmonized high throughput barcoding of insular arthropod biodiversity: Toward a Genomic Observatories Network for islands. *Molecular Ecology*, *32*(23), 6161–6176.

Ficetola, G. F., & Taberlet, P. (2023). Towards exhaustive community ecology via DNA metabarcoding. *Molecular Ecology*, *32*(23), 6320–6329.

Formoso-Freire, V., Baselga, A., & Gómez-Rodríguez, C. (2025). Climatic stability predicts the congruence between species abundance and genetic diversity. *Ecography*, *2025*(3), e07200.

Gibson, J., Shokralla, S., Porter, T. M., King, I., van Konynenburg, S., Janzen, D. H., Hallwachs, W., & Hajibabaei, M. (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(22), 8007–8012.

Graham, N. R., Krehenwinkel, H., Lim, J. Y., Staniczenko, P., Callaghan, J., Andersen, J. C., Gruner, D. S., & Gillespie, R. G. (2023). Ecological network structure in response to community assembly processes over evolutionary time. *Molecular Ecology*, *32*(23), 6489–6506.

GRI - Home. (n.d.). Retrieved May 4, 2025, from https://www.globalreporting.org/

Hartig, F., Abrego, N., Bush, A., Chase, J. M., Guillera-Arroita, G., Leibold, M. A., Ovaskainen, O., Pellissier, L., Pichler, M., Poggiato, G., Pollock, L., Si-Moussi, S., Thuiller, W., Viana, D. S., Warton, D. I., Zurell, D., & Yu, D. W. (2024). Novel community data in ecology-properties and prospects. *Trends in Ecology & Evolution*, *39*(3), 280–293.

Hill, M. O. (1973). Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology*, *54*(2), 427–432.

Holman, L. E., de Bruyn, M., Creer, S., Carvalho, G., Robidart, J., & Rius, M. (2021). Animals, protists and bacteria share marine biogeographic patterns. *Nature Ecology & Evolution*, *5*(6), 738–746.

Humphries, C. J., Williams, P. H., & Vane-Wright, R. I. (1995). Measuring biodiversity value for conservation. *Annual Review of Ecology and Systematics*, *26*(1), 93–111.

Kennedy, S., Calaor, J., Zurápiti, Y., Hans, J., Yoshimura, M., Choo, J., Andersen, J. C., Callaghan, J., Roderick, G. K., Krehenwinkel, H., Rogers, H., Gillespie, R. G., & Economo, E. P. (2023). Richness and resilience in the Pacific: DNA metabarcoding enables parallelized evaluation of biogeographic patterns. *Molecular Ecology*, *32*(23), 6710–6723.

Kitson, J. J. N., Warren, B. H., Thébaud, C., Strasberg, D., & Emerson, B. C. (2018). Community assembly and diversification in a species-rich radiation of island weevils

(Coleoptera: Cratopini). *Journal of Biogeography*, *45*(9), 2016–2026.

Krehenwinkel, H., Hans, J., Junk, I., Mahla, L., Melcher, A., Stothut, M., & Kennedy, S. (2024). *Metabarcoding for biodiversity estimation*. https://scholar.google.com/citations?user=1vCvXzgAAAAJ&hl=en&oi=sra

Krehenwinkel, H., Weber, S., Broekmann, R., Melcher, A., Hans, J., Wolf, R., Hochkirch, A., Kennedy, S. R., Koschorreck, J., Künzel, S., Müller, C., Retzlaff, R., Teubner, D., Schanzer, S., Klein, R., Paulus, M., Udelhoven, T., & Veith, M. (2022). Environmental DNA from archived leaves reveals widespread temporal turnover and biotic homogenization in forest arthropod communities. *Elife*, *11*, e78521.

Krehenwinkel, H., Wolf, M., Lim, J. Y., Rominger, A. J., Simison, W. B., & Gillespie, R. G. (2017). Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Scientific Reports*, *7*(1), 17668.

Kunming-Montreal Global Biodiversity Framework. (2024, October 1). Secretariat of the Convention on Biological Diversity. https://www.cbd.int/gbf

Leray, M., & Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(7), 2076–2081.

Lim, J. Y., Patiño, J., Noriyuki, S., Cayetano, L., Gillespie, R. G., & Krehenwinkel, H. (2022). Semi-quantitative metabarcoding reveals how climate shapes arthropod community assembly along elevation gradients on Hawaii Island. *Molecular Ecology*, *31*(5), 1416–1429.

Luo, M., Ji, Y., Warton, D., & Yu, D. W. (2023). Extracting abundance information from DNA-based data. *Molecular Ecology Resources*, *23*(1), 174–189.

Magurran, A. E. (1988). Diversity indices and species abundance models. In *Ecological Diversity and Its Measurement* (pp. 7–45). Springer Netherlands.

Magurran, A. E. (2007). Species abundance distributions over time. *Ecology Letters*, *10*(5), 347–354.

Martinez, N. D., Hawkins, B. A., Dawah, H. A., & Feifarek, B. P. (1999). Effects of sampling effort on characterization of food-web structure. *Ecology*, *80*(3), 1044–1055.

Matthews, T. J., & Whittaker, R. J. (2015). REVIEW: On the species abundance distribution in applied ecology and biodiversity management. *The Journal of Applied Ecology*, *52*(2), 443–454.

Moraes Zenker, M., Portella, T. P., Pessoa, F. A. C., Bengtsson-Palme, J., & Galetti, P. M., Jr. (2024). Low coverage of species constrains the use of DNA barcoding to assess mosquito biodiversity. *Scientific Reports*, *14*(1), 7432.

Moreno, C. E., & Halffter, G. (2001). On the Measure of Sampling Effort Used in Species Accumulation Curves. *Journal of Applied Ecology*, *38*(2), 487–490.

Noguerales, V., Meramveliotakis, E., Castro-Insua, A., Andújar, C., Arribas, P., Creedy, T. J., Overcast, I., Morlon, H., Emerson, B. C., Vogler, A. P., & Papadopoulou, A. (2021). Community metabarcoding reveals the relative role of environmental filtering and spatial processes in metacommunity dynamics of soil microarthropods across a mosaic of montane forests. *Molecular Ecology*. https://doi.org/10.1111/mec.16275

on Biological, C. (n.d.). *Kunming‐Montreal global biodiversity framework*.

Overcast, I., Achaz, G., & Aguilée, R. (2023). Towards a genetic theory of island biogeography: Inferring processes from multidimensional community‐scale data. *Global Ecology and*. https://onlinelibrary.wiley.com/doi/abs/10.1111/geb.13604

Overcast, I., Noguerales, V., Meramveliotakis, E., Andújar, C., Arribas, P., Creedy, T. J., Emerson, B. C., Vogler, A. P., Papadopoulou, A., & Morlon, H. (2023). Inferring the ecological and evolutionary determinants of community genetic diversity. *Molecular Ecology*. https://doi.org/10.1111/mec.16958

Overcast, I., Ruffley, M., Rosindell, J., Harmon, L., Borges, P. A. V., Emerson, B. C., Etienne, R. S., Gillespie, R., Krehenwinkel, H., Mahler, D. L., Massol, F., Parent, C. E., Patiño, J., Peter, B., Week, B., Wagner, C., Hickerson, M. J., & Rominger, A. (2021). A unified model of species abundance, genetic diversity, and functional diversity reveals the mechanisms structuring ecological communities. *Molecular Ecology Resources*, *21*(8), 2782–2800.

Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., Bruford, M. W., Brummitt, N., Butchart, S. H. M., Cardoso, A. C., Coops, N. C., Dulloo, E., Faith, D. P., Freyhof, J., Gregory, R. D., Heip, C., Höft, R., Hurtt, G., Jetz, W., … Wegmann, M. (2013). Ecology. Essential biodiversity variables. *Science (New York, N.Y.)*, *339*(6117), 277–278.

Preston, F. W. (1948). The Commonness, And Rarity, of Species. *Ecology*, *29*(3), 254–283.

Purvis, A., & Hector, A. (2000). Getting the measure of biodiversity. *Nature*, *405*(6783), 212–219.

Ratnasingham, S., & Hebert, P. D. N. (2007). bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, *7*(3), 355–364.

Reynolds, L. K., McGlathery, K. J., & Waycott, M. (2012). Genetic diversity enhances restoration success by augmenting ecosystem services. *PloS One*, *7*(6), e38397.

Schmidt, C., & Garroway, C. J. (2021). The conservation utility of mitochondrial genetic diversity in macrogenetic research. *Conservation Genetics*, *22*(3), 323–327.

Sherwin, W. B., Chao, A., Jost, L., & Smouse, P. E. (2017). Information theory broadens the spectrum of molecular ecology and evolution. *Trends in Ecology & Evolution*, *32*(12), 948–963.

Sleator, R. D., Shortall, C., & Hill, C. (2008). Metagenomics. *Letters in Applied Microbiology*. https://academic.oup.com/lambio/article-abstract/47/5/361/6700850

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8), 2045–2050.

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, *105*(2), 437–460.

Van Meerbeek, K., Jucker, T., & Svenning, J.-C. (2021). Unifying the concepts of stability and resilience in ecology. *The Journal of Ecology*, *109*(9), 3114–3132.

Vellend, M. (2016). *The Theory of Ecological Communities (MPB-57)*. Princeton University Press.

Wauchope, H. S., Zu Ermgassen, S. O. S. E., Jones, J. P. G., Carter, H., Schulte To Bühne, H., & Milner-Gulland, E. J. (2024). What is a unit of nature? Measurement challenges in the emerging biodiversity credit market. *Proceedings. Biological Sciences*, *291*(2036), 20242353.

Weitemier, K., Penaluna, B. E., Hauck, L. L., Longway, L. J., Garcia, T., & Cronn, R. (2021). Estimating the genetic diversity of Pacific salmon and trout using multigene eDNA metabarcoding. *Molecular Ecology*, *30*(20), 4970–4990.

Wunder, S., Fraccaroli, C., Bull, J. W., Dutta, T., Eyres, A., Evans, M. C., Thorsen, B. J., Jones, J. P. G., Maron, M., Muys, B., & al., E. (2024). Biodiversity credits: learning lessons from other approaches to incentivize conservation. In *osf.io*. https://doi.org/10.31219/osf.io/qgwfc

Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution / British Ecological Society*, *3*(4), 613–623.

Figures

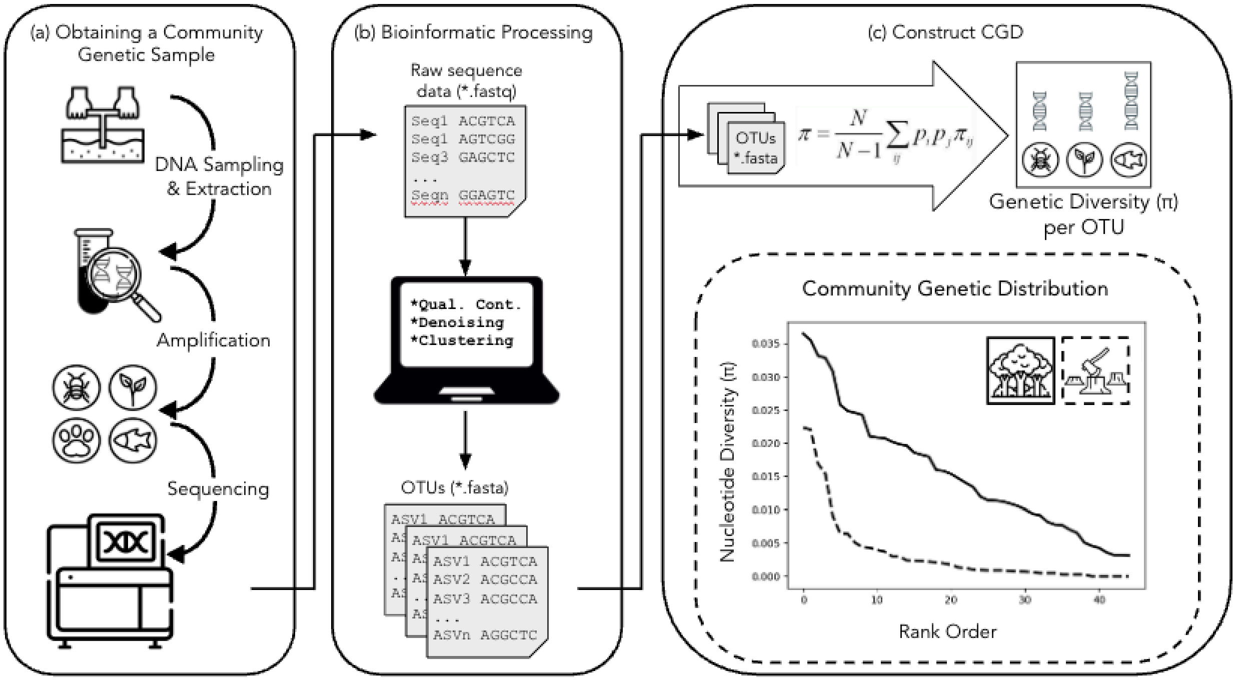

Figure 1 - Overview of the workflow for constructing the Community Genetic Distribution. The process begins with a) the collection of a sample of community-scale DNA following one of several approaches for obtaining metabarcode or environmental DNA samples. As examples, environmental DNA samples can be taken from numerous media including soil, water, or air, and metabarcode samples could include taking bulk arthropods from traditional field based collections using malaise or pitfall traps (for reviews see (Ruppert et al. 2019; Emerson et al. 2023). DNA is then extracted from the samples and amplified for one or more markers for the focal group of interest, followed by high-throughput sequencing. b) Raw sequence data is then subject to bioinformatic processing, which includes steps for quality control (filtering low quality data, removing adapter contamination), denoising (identifying Amplicon Sequencing Variants [ASVs] by statistically distinguishing between sequencing errors and true variants), and finally clustering ASVs into Operational Taxonomic Units (OTUs) using a given sequence similarity threshold (3% is common in the literature). Bioinformatic processing of metabarcode data involves many complexities which are beyond the scope of this perspective (for reviews see (Creedy et al. 2021; Hakimzadeh et al. 2024), but the overall goal is to produce a set of data files (in fasta format) containing all the DNA sequences for a given OTU. c) With this data in hand, one can construct the Community Genetic Distribution (CGD) by calculating within-OTU nucleotide diversity (π; (Nei and Li 1979) and aggregating π values across all OTUs in the sample. Similar to the Species Abundance Distribution, the CGD can be represented in several ways, which emphasize different properties of the distribution. Here we illustrate rank-ordered π values (akin to Rank Abundance Distributions) for two hypothetical community samples from intact (solid line) and disturbed systems (dashed line).

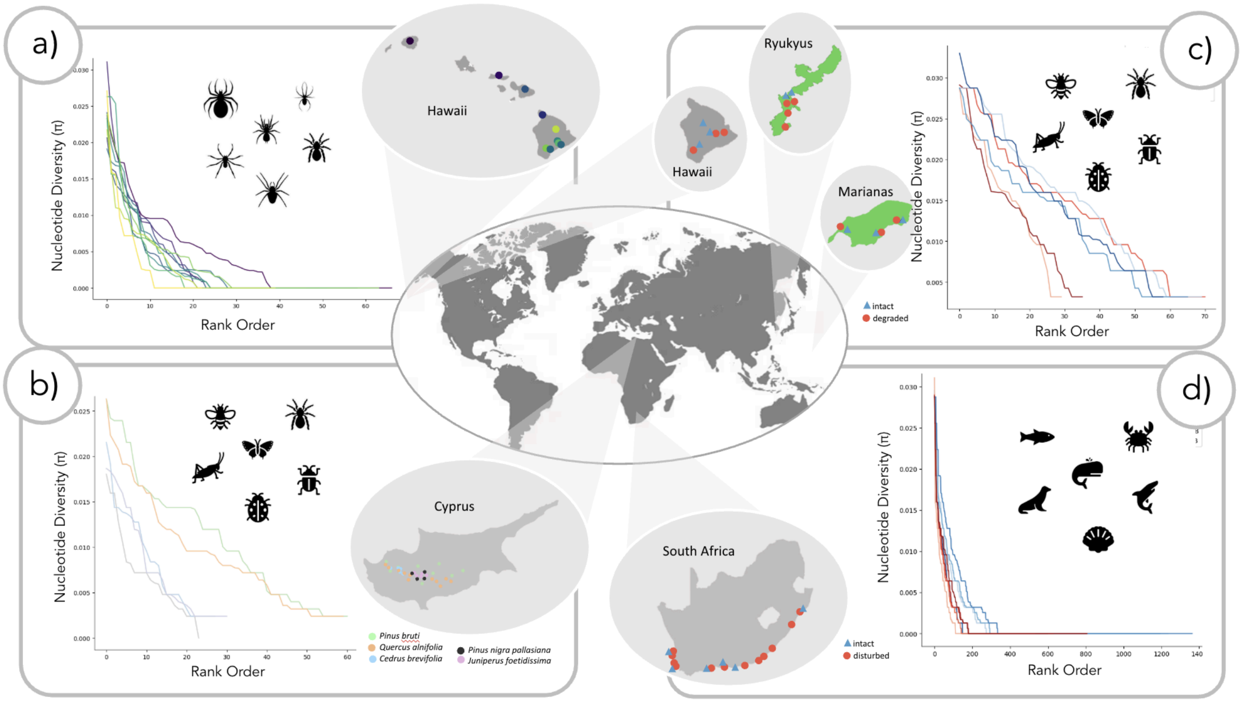

Figure 2 - Empirical examples of CGD across disparate spatial and taxonomic scales
Empirical examples from four different systems across the globe illustrate the effects of different ecological and evolutionary contexts on community genetic diversity composition: a) community age and co-evolutionary time; b) habitat type; and disturbance regimes in c) terrestrial and d) marine systems. a) The CGD for spider communities from Hawaii show an increasing pattern of CGD magnitude with increasing community age. The color of each CGD curve indicates substrate age from which the community was sampled, with age increasing from yellow (youngest) to dark purple (oldest). b) Soil microarthropod communities from montane forests on Cyprus sampled from five different habitat types show a pattern of increased CGD for more widely distributed pine (green) and oak (orange) forests, with reduced CGD in narrowly restricted cedar (blue) and high-elevation pine/juniper (purple and black, respectively) forests. Both panels on the right show the effect of different disturbance regimes: c) forest dwelling arthropod communities and; d) marine eukaryote communities. In both cases blue curves show stable and intact community CGDs and red curves show disturbed sites.

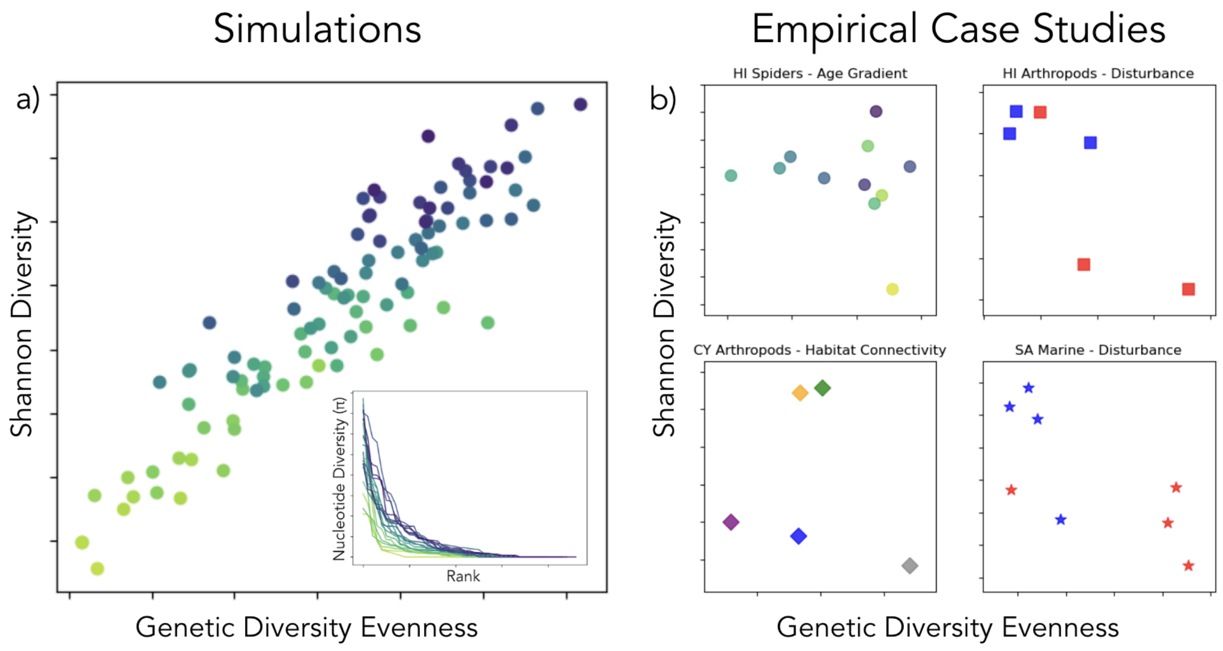

Figure 3 - Indicator metrics calculated for simulated and empirical CGDs
Calculating summary statistics on the shape of CGD curves can illuminate changes in community-scale genetic diversity among samples from comparable systems. Panel a) Eco-evolutionary simulations of a community assembling through time (inset; darker colors indicate older communities) reproduce the pattern of increasing genetic diversity with increasing age that are observed in the empirical spider communities on Hawaii (Fig. 2a). Shannon diversity and genetic diversity evenness are positively correlated in these simulations which emulate a natural (undisturbed) community assembling through time. Panel b) Summary statistics calculated for empirical case study CGDs show diverse changes that illustrate how the CGD can respond to different ecological processes. The strongest and most consistent pattern is observed under disturbance regimes where pristine communities of Hawaiian arthropods (blue squares) and marine eukaryotes (blue stars) show increased Shannon diversity and decreased genetic diversity evenness with respect to paired disturbed sites (red). Disturbance increases evenness through local extinction of rare species which show predominantly low genetic diversity. In undisturbed systems Shannon diversity and genetic diversity evenness show a trend of increasing with increasing community age (Hawaiian spiders; circles) and increasing habitat connectivity (Cyprus arthropods; diamonds).