Conservation macrogenetics

Authors: Chloé Schmidt¹,²,³*, Sean Hoban⁴, Walter Jetz¹,²

Affiliations:
1 Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA
2 Center for Biodiversity and Global Change, Yale University, New Haven, CT, USA
3 German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
4 The Center for Tree Science, The Morton Arboretum, Lisle, IL, USA

ORCIDs:
CS: 0000-0003-2572-4200
SH: 0000-0002-0348-8449
WJ: 0000-0002-1971-7277

*Correspondence to:
Chloé Schmidt
German Center for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig
Puschstrasse 4
Leipzig 04103
Germany
email: chloe.schmidt@idiv.de

Keywords: genetic diversity, biodiversity, macroecology, SGDCs, landscape genetics, essential biodiversity variables, biogeography
Abstract

Genetic diversity is a core aspect of biodiversity that has been underrepresented in global conservation policy but is gaining rapidly increasing recognition. Conservation geneticists have traditionally focused on identifying, managing, and safeguarding the adaptive potential of specific populations or species. However, for almost all species, conservation relevant, population-level genetic data is lacking. This limits the extent to which genetic diversity can be monitored, reported, and used for conservation policy and decision-making. Fortunately, rapid growth of open access repositories of genetic data holds great promise for conservation applications. Macrogenetics is an emerging discipline that explores patterns of, and processes underlying, population genetic composition at broad taxonomic and spatial scales by aggregating and reanalyzing thousands of previously published genetic datasets. Here we explain how focusing macrogenetic tools on conservation needs, or “conservation macrogenetics”, offers new opportunities to support genetic monitoring and decision-making for conservation practice. Conservation macrogenetics also provides an empirical basis for considering how anthropogenic drivers and policy decisions jointly affect multiple levels of biodiversity (genes, species, ecosystems) to better understand the complexity and resilience of biological systems.
1. Introduction

Due to its central importance for maintaining fitness and adaptive potential, intraspecific genetic diversity underlies the functioning and resilience of populations, species, communities, and ecosystems[1,2]. Alongside species and ecosystem biodiversity, genetic biodiversity (Box 1) is designated for protection by numerous national governments’ endangered species legislations, and by the Convention on Biological Diversity (CBD), a global treaty on conservation and sustainable use[3]. Genetic diversity is increasingly recognized as essential for understanding the ecological and evolutionary forces that shape biodiversity across organizational levels and improving the predictability of biological responses to environmental change for conservation and sustainability applications[1,4,5].

Despite its value for conservation, genetic diversity is not well integrated into national (for most countries) or global conservation policy[6–8], partly because genetic data have historically been difficult and expensive to collect. Where genetic data are available, they tend to be used for species-specific conservation actions[9]. Global and national policy for protecting genetic diversity have typically prioritized economically important species and ex situ conservation strategies (e.g., seed banks, captive breeding programs)[10]. Initiatives to conserve genetic diversity in unmonitored wildlife populations with limited to no data availability are therefore difficult to implement. Synthesizing knowledge on the intrinsic and extrinsic drivers of genetic diversity and population differentiation across species would be a valuable resource for conservationists to systematically consider genetic biodiversity in decision-making with, or without, genetic data.

Macrogenetics is an emerging field that repurposes existing genetic data to uncover population genetic patterns across taxa, time, and space[11]. We argue that developing macrogenetics
research in directions that are relevant for conservation applications and policies, or conservation macrogenetics, is needed, and timely (Fig. 1). Here, we will give a brief overview of conservation genetics and the benefits of leveraging multispecies data, explore conservation applications of macrogenetics, and close with forward-looking perspectives on the purview of conservation macrogenetics.

2. **A macro view of conservation genetics**

Genetic data are typically used in conservation to assess species-specific population parameters such as genetic diversity, inbreeding, demography, and isolation, and to delineate management units or evolutionarily significant units[2]. However, the data underlying these metrics are not often synthesized to inform the genetic status of species lacking genetic data, or used for multispecies conservation planning, though their potential has been recognized[4,9,12].

In contrast, conservation approaches used for species-level conservation have drawn on established statistical relationships between factors such as abundance, body size, environments, and traits; and centralized data and information resources, such as the International Union for the Conservation of Nature (IUCN) Red List, Map of Life (MOL), and the Global Biodiversity Information Facility (GBIF), to inform policy and decision-making. Knowledge from the field of macroecology—including island biogeography, scaling relationships, and niche theory—has provided baseline context for conservation decisions[13]. The leap from macroecological theory to conservation-relevant research was partly enabled by the rapid growth of biological and environmental databases[13] (e.g., those for occurrence records, life-history traits, demography, and remote sensed data), and analytical tools[13,14].
At present, our understanding of the broad-scale geographic distribution and patterning of genetic diversity, how patterns scale from within to across species, and their relationships with environments and other aspects of biodiversity are poorly understood. Furthermore, the present storage and access requirements of open genetic data and metadata are not conducive to routine use by researchers or conservation practitioners. As such, conservation planning, action, and policy have not integrated the genetic component of biodiversity to the same extent as species biodiversity because the accumulation of genetic data has long lagged behind other species-level data types.

Conservation macrogenetics is conceptually related to the interface of conservation and macroecology[13], which focuses on how general principles inferred from pattern-first, top-down analyses of biodiversity and ecological data can inform conservation, and conservation biogeography, the application of biogeographical principles underlying species’ distributional dynamics to conservation goals[14] (Box 2). Macrogenetics has already begun to reveal fundamental processes producing and maintaining biodiversity[11], insights which can be directed towards achieving conservation goals cohesively across biodiversity levels (Box 2). Well-defined statistical and mechanistic frameworks uniting variation across genes, species, and environments will support decision-making by enhancing predictive capacity and enabling practitioners to fill genetic data gaps by borrowing strength from other available data types. A strong basis for how population genetics fits into existing conservation practices based on macroecological principles would not only better integrate genetic diversity into conservation policy, but enable conservationists to emphasize the protection of biodiversity processes in addition to biodiversity states[9,15] across levels of organization. Below are key research questions for conservation macrogenetics:
1. How large an area is needed to conserve a defined minimum threshold of genetic diversity?

2. Does connectivity that supports species movements also support gene flow?

3. Are the same areas important for protecting genes and species; if not, where do these elements of biodiversity align?

4. How can we categorize species and populations by their conservation threat level using genetics (e.g. similar to the Red List?)

5. How can countries report genetic status and trends for international policy, including by using proxies?

6. How interrelated are the different components of genetic diversity, recently defined as Essential Biodiversity Variables, and thus which need to be measured in future studies?

### 3. Conservation applications for macrogenetics

The conservation biologist Michael Soulé was among the first to repurpose genetic data for new research questions[16]. The development of molecular markers (based on protein variants called allozymes) in the 1960s made estimating genetic diversity in natural populations feasible and routine. Soulé mined this rapidly growing literature to empirically demonstrate that high genetic diversity was maintained in large populations with long intervals between bottlenecks and low divergence rates. Although the usefulness of macrogenetics research in conservation has been recognized since Soulé’s time[11,17], practical applications have yet to be enumerated clearly. In this section we explore several areas where macrogenetics could inform conservation decision-making and practice.
3.1 Monitoring and predicting biodiversity change

Estimating effective population size decline. Widespread data archiving for conservation macrogenetics would be critical for reporting on genetic indicators where datasets are sufficient to estimate genetic summary statistics such as allelic richness or effective population size (Ne). However, data gaps and the difficulty associated with obtaining new data means that indicators of genetic composition may often be impossible to estimate directly from data. Multispecies genetic data, and a strong understanding of how genetic diversity components vary across species, can help fill these gaps. For example, macrogenetic syntheses led to the conservation rule of thumb that Ne tends to be roughly one-tenth to one-third of the census population size (Nc)[18,19]. Ne is an important genetic parameter in conservation that estimates the strength of genetic drift eroding genetic diversity in a population, and the relative ability of a population to track environmental change via adaptation. This 0.1 “rule of thumb” underlies a genetic biodiversity indicator (proportion of populations with Ne >500) leveraging abundance data[17] that was adopted by the Kunming-Montreal Global Biodiversity Framework, the commitments by 196 countries to achieve for nature by 2030[20].

Multispecies data can also be used to refine taxon-specific guidelines for genetic indicators. For instance, Ne in animals can vary across 3 to 4 orders of magnitude[21], meaning that a minimum Nc sufficient for some species (e.g., polar bears) is much too small for others (e.g., seagrasses or bumblebees). Practical use of the Ne>500 indicator relies on using Nc and the Ne/Nc ratio. Targeted macrogenetic analyses examining variation in Ne/Nc can lead to Nc threshold modifications for some species, for example by taking into account taxonomic group, or traits such as ploidy and reproductive mode[22,23]. This will be vital for correctly applying this indicator and reporting to the CBD, as well as for ensuring national and regional conservation
action targeted at species’ populations below the Ne = 500 threshold (for example by captive breeding and reintroduction). We note that this indicator can also be calculated from Ne directly, rather than using the Ne/Nc ratio. Macrogenetics can be used for reporting Ne in areas rich in genetic data such as North America and Europe.

Indicators that require genomic data are also valuable, and will become feasible for more species as data accumulate. For example, Peart et al.[24] used whole genomes to estimate residual variation in Ne/Nc across 17 pinniped species after controlling for species-specific demographic history (Tajima’s D). Positive Ne/Nc residuals reflected species whose contemporary abundance was lower than would be expected from historical Ne, and suggest declining population sizes. These Ne/Nc residuals were correlated with species Red List status.

Beyond the Ne/Nc ratio, conservation macrogenetics research could generate knowledge of how genetic composition varies with environments and other aspects of biodiversity, such as species distributions used for monitoring species populations[25], or phenology for monitoring traits. Effective proxies for indicators of genetic composition that leverage environmental and biological data are likely achievable at regional scales[5,26,27].

**Genetic diversity – area relationship.** The species area relationship (SAR), and its associate endemics area relationship (EAR), are foundational concepts in ecology that address how the numbers of any (SAR) or regionally restricted (EAR) species scale with area. In the absence of other data, the relationship has been used to estimate the magnitude of species richness declines following habitat loss[28,29], though the theoretical underpinnings have been debated[30,31]. Parallels between species and alleles (Box 2) and growing empirical and theoretical evidence suggest the existence of genetic diversity area relationships[1,32–36]. By extension of the SAR
and EAR applications to estimating species loss, this could be relevant for approximating levels of genetic diversity retained or vanished as areas are conserved or lost.

In both species and alleles, diversity accumulates rapidly and slows with increasing area, with a close fit to an exponential. The precise mathematical parameters of genetic diversity area relationships, including scaling exponents, can vary depending on the diversity metric[36], species[33–35], population structure (beta diversity), dispersal means, and the distribution of genetic diversity across a species range[1,35]. Additionally, genetic diversity is less well-explained by area than species richness[33], probably due to among and within population components of genetic subdivision and non-spatial evolutionary processes. Our understanding is still developing, and this suggests other important factors in genetic diversity area relationships have yet to be identified. Applying conservation macrogenetics to SAR and genetic diversity area relationships will help mobilize these scaling relationships for global conservation genetic policy.

Mimura et al.[1] suggest that genetic diversity area relationships using scaling exponents based on population differentiation could form the basis for a report card on genetic diversity loss due to area loss. Using an intermediate scaling exponent across species, Exposito-Alonso et al. suggest that species have already lost 10% of genetic diversity in terms of alleles since the industrial revolution[34]. This estimate aligns with an average 6% loss during the same period, estimated from datasets quantifying temporal genetic change in 91 species[37]. Macrogenetics using data from many more species will allow further refinement of these estimates.

**Habitat loss and species genetic diversity loss.** Recent advances in data integration and spatial, remote-sensing supported modelling are delivering information about species distributions in greater detail for an increasing range of taxa[38]. Data on species habitat preferences linked to
remotely-sensed land cover are enabling assessments of habitat-suitable ranges and their potential change over time[39–41]. Occurrence pixels that are assessed for suitable habitat may offer a proxy for population size that can be refined using, e.g. allometrically derived estimates of individual area requirements[42–44]. With remotely-sensed landcover products gaining in quality and precision, this creates an opportunity to go beyond SARs in estimating the consequences of habitat loss for genetic diversity. Remotely sensed changes in habitat-suitable range, e.g. characterized at 30m to 1000m spatial resolution, allow a more direct assessment of where habitat loss might cause population fragmentation, reduction in genetic diversity, or where populations may fall below critical thresholds for retaining genetic health and adaptive potential (e.g., [45]). Spatially explicit habitat loss data allow capturing changes in habitat connectivity in addition to area, and in some cases estimates of remaining population size[46]. This combined assessment of species range-wide changes in habitat area and connectivity has recently been empirically implemented at scale in MOL (https://mol.org/indicators/habitat). Individual species habitat scores contribute to the Species Habitat Index, a component indicator for Global Biodiversity Framework for Goal A that measures changes in species extinction risk and population size. Such spatially explicit indicators also have the potential to provide decision-support: for example, conservationists need accurate predictions of the amount of genetic diversity safeguarded by a protected area or lost to habitat destruction.

**Endangered species listing.** A clear avenue for repurposing publicly archived genetic data is for endangered species assessment. The IUCN Red List is one of the most widely used assessments of species extinction risk. The Red List considers information on species range area and fragmentation, abundance, and population trends in its evaluation. Decisions are not currently informed by genetic data; indeed, Red List status cannot be predicted from genetic data
Adequately testing for genetic diversity declines that may elevate extinction risk requires spatiotemporal data at a finer resolution than is generally available at present (i.e., population-level sampling over time) [11, 37, 47]. Red List status has previously been proposed as an indicator of genetic diversity status [48] including for use by the CBD. However, this species-level indicator is not informative for genetic diversity trends [23, 47]. Increased data availability will help determine the extent to which changes in genetic composition across space and time relate to ecological factors associated with Red List extinction risk (Fig. 2).

At a national level macrogenetics, especially of population differentiation metrics, may help with endangered species listing and management. This could include designation of “critical habitat” or “distinct population segments” (an important issue for implementation of genetic diversity protection under the U.S. Endangered Species Act [49]) and other policy decisions. Conservation macrogenetics may help enable a genetic diversity Red List or possibly criteria for triggering Red List status based on genetic diversity or genetic threats. At minimum, genetic diversity knowledge, including predictions from macrogenetics, should be summarized in Red List assessments even if not used for decision assessment [47].

3.2. Spatial conservation planning

Protecting multispecies genetic diversity. Regional to global maps of single- and multi-species genetic data would provide essential opportunities to visualize generalizable diversity patterns and identify hot and coldspots for genetic diversity, as is already done for species richness. Genetic diversity maps could be operationalized for specific taxonomic groups (e.g., trees or
mammals) or multiple taxa in a region (e.g., a single state) by leveraging knowledge from other
taxa and regions, depending on the spatial and taxonomic scale of the data.

Macrogenetics can be informative for the quantity and quality of protected areas and other
effective area-based conservation measures. For example, strategies addressing the “30 by 30”
Target of the new Kunming-Montreal Framework that aims to protect 30% of land and sea by
2030 are more likely to preserve common, but not most, alleles present in a population[50].

Without prior knowledge of genetic diversity or population structure, more populations require
protection to capture a majority (~86-91%) of alleles and heterozygosity, corresponding to ~50%
protected area[50,51], in line with recommendations to protect 90-95% of genetic diversity of
domestic and wild populations to prevent genetic erosion[52]. Including multispecies genetic
diversity in protected area decision-making can be more effective than decisions based on single
species or the presence or absence of species or habitats alone[4,9].

Balancing conflicting prioritization goals. A deeper understanding of relationships between
alpha and beta genetic diversity (Box 2) and species richness is vital for spatial conservation
planning. Multispecies measures of population genetic connectivity can provide empirical
estimates of gene flow to enhance our understanding of functional connectivity across protected
area networks beyond using species movement data, which does not necessarily correspond to
gene flow[53].

However, variable relationships between multispecies genetic diversity and species richness
could cause conflicting decisions for spatial conservation. For example, Schmidt et al.[5,54]
generated multispecies maps of neutral, genome-wide genetic diversity and differentiation for
mammals and amphibians in the United States and Canada. The locations of genetic diversity
coldspots differed across classes and tended to be in species richness hotspots. Genetic coldspots
for mammals and amphibians were located in the southwestern and southeastern US, respectively. Maps have also been produced for mitochondrial genetic diversity, where patterns differ from those of neutral nuclear genetic diversity[55–58]. Hanson and colleagues[26] have used several multispecies datasets to plan protected areas, which seems to be very useful in some places but less successful in regions of highly complex biogeographic history such as the Iberian Peninsula (e.g. where multiple diverged lineages have mixed).

Setting protected areas in hotspots for species richness or environmental heterogeneity may thus protect beta, but not alpha genetic diversity. This strategy could risk inadvertently protecting small, isolated populations that are at higher risk of genetic erosion[5,12,54]. Maximizing beta species diversity may also not capture beta genetic diversity for species not represented at multiple locations in a protected area network, ultimately limiting species’ potential for long-term persistence. Understanding and balancing alpha and beta macrogenetic diversity patterns is therefore vital for delivering options to maintain genetic diversity for most species across sites[9].

4. Data needs for realizing conservation macrogenetics

A major barrier to the widespread adoption of conservation macrogenetics is the lack of genetic data for most species. Like other aspects of biodiversity, genetic data have higher coverage of spaces and species in North America and Europe[55,56,59]. Poorly documented sample metadata also significantly limit the reusability of archived genetic data[60]. Macrogenetics so far has relied on opportunistically repurposing publicly available genetic data, but this is not sufficient to build a catalogue of genetic diversity using different methods, markers, and taxa. The research community must collaborate on and support efforts to revive unpublished data and collect new
data in undersampled areas, annotate existing data, and fund and build infrastructure that ensure data are openly and easily accessible for both research and conservation applications.

4.1. Data types

The most common publicly available genetic markers are microsatellites, mitochondrial DNA sequences, and single nucleotide polymorphism (SNP) data. To date, macrogenetics has largely capitalized on mitochondrial DNA\cite{55-58,61} and microsatellites\cite{5,54,59,62-65}. Its wide availability notwithstanding, reliable inferences about macrogenetic patterns of genome-wide genetic diversity or population structure cannot be made from mitochondrial DNA alone\cite{66}. Mitochondrial genetic diversity is not correlated with genome-wide (i.e., nuclear) genetic diversity, and does not generally represent fitness or adaptive potential\cite{66}. These shortcomings cannot be overcome by increasing sample size\cite{67,68}.

Nuclear markers (microsatellites or SNPs) are needed to estimate genetic diversity or composition that is conservation-relevant. Microsatellites approximate genome-wide genetic diversity well\cite{69}, and reflect population demography and neutral evolutionary processes (genetic drift and gene flow). They are abundant and still in wide use in landscape genetics and conservation genetics\cite{70}. Whole genome SNP data are the most versatile marker type, providing information about adaptive and neutral processes over contemporary and historical periods. Despite these advantages, the availability of SNPs has not yet matched the long accumulation of microsatellites in public repositories\cite{70}. Although sequencing costs are falling, SNPs are still not equally an option across the globe due to financial restrictions, availability of local sequencing facilities, and informatic expertise. Thus, the continued and valuable use of microsatellites where appropriate should be supported in the near future. In the meantime,
protocols for the management and macrogenetic synthesis of SNP data processed with different bioinformatic pipelines still need to be developed[11].

Different marker types are also typically used to address different types of research questions, affecting sampling design, downstream analyses, and practical applications in macrogenetics. For instance, the first global map of genetic diversity, published in 2016, was based on mitochondrial DNA[55]. Although including an impressive nearly 93,000 sequences, this study and others like it[56–58,61] have revealed insights about the structure of publicly available genetic sequence (mitochondrial or genomic) data that may affect their reusability for conservation macrogenetics. Generally, a dearth of population-level data for mtDNA data—i.e., multiple individuals sampled at a specific location to estimate population parameters—in sequence repositories limits possibilities for intraspecific analysis[61]. With low replication within species, individual sequences were often pooled to estimate genetic diversity across species and spatiotemporal scales that may sometimes be too large to be relevant or meaningful (e.g., grid cells of ~150000 km²)[68]. As SNPs become more routinely used and metadata standards continue to improve[11], researchers will be increasingly able to overcome these issues and conduct analyses at the population level, and where needed, allow data to be flexibly pooled into realistic populations (see e.g. [61]).

4.2. Data repositories and standards

Storage practices also differ across data types. Mitochondrial sequences and raw SNP sequence data are typically stored in programmatically accessible databases like GenBank (https://www.ncbi.nlm.nih.gov/genbank/), or BOLD[71]. Microsatellite data and processed SNP data (i.e., final variant calls used in analysis) are often stored in general-purpose repositories
such as DRYAD (https://datadryad.org/)[54,63]. Raw genomic data are large files that require specific expertise to process. However, called SNPs used in analysis may have been processed with different bioinformatic pipelines that make datasets incomparable. These storage conventions can complicate data aggregation and reusability. Indeed, most macrogenetics work that mobilize raw data use mitochondrial DNA due to its ease of access and straightforward synthesis across datasets[11]. In lieu of data access, macrogenetics has often relied on genetic summary statistics harvested from the literature[59,72], which may already be useful for conservation. However, data access, and the flexibility it gives users in the choice of summary statistics and analytical methods, is invaluable for conservation macrogenetics.

Moving forward, a single, standardized, queryable repository for publicly available genetic data of all types will be a valuable resource for conservation genetics and macrogenetics. This could also take the form of a platform that integrates data stored in different repositories, such as the Data Observation Network for Earth (https://www.dataone.org/), GBIF (https://www.gbif.org/), or MOL (https://mol.org/). The Genomic Observatories MetaDatabase (GEOME; https://geome-db.org/)[73,74] has taken an essential step in this direction by linking genomic data to sample metadata. Metadata were retrieved from the literature in “datathon” events, one of which was estimated to have rescued approximately US$ 2.1 million worth of metadata, representing 2300 hours of work by 25 data curators[75]. These types of initiatives are crucial for repurposing genetic data, and for maximizing return on investment by ensuring data longevity.

Concluding remarks
Macrogenetics, with its broad taxonomic and spatial perspectives on genetic diversity, is well-suited for integration into global conservation policy. By mobilizing existing data sampled from wildlife populations, macrogenetics directly tackles key gaps in global policy schemes [6,17]: 1) a focus on domestic or economically important species; 2) emphasis on *ex situ* management action, and monitoring genetic diversity in single species and species with DNA data; 3) a need to develop and test easily quantifiable genetic indicators; and 4) a lack of genetic data in many species and regions. Conservation macrogenetics applications such as those outlined in this article will help bring genetic diversity to the global policy stage. Integrating conservation macrogenetics with species and ecosystem conservation can also support holistic conservation and management policies to efficiently conserve all levels of biodiversity.

**Acknowledgements**

We thank Ivan Paz Vinas and Colin Garroway for feedback on manuscript drafts. This work was conducted as a part of the Standardizing, Aggregating, Analyzing and Disseminating Global Wildlife Genetic and Genomic Data for Improved Management and Advancement of Community Best Practices Working Group supported by the John Wesley Powell Center for Analysis and Synthesis, funded by the U.S. Geological Survey. SH acknowledges funding from NSF award 1759759.

**References**

5. Schmidt, C. et al. (2022) Genetic and species-level biodiversity patterns are linked by demography and ecological opportunity. Evolution 76, 86–100


49. Hoban, S. et al. (2021) Genetic diversity is considered important but interpreted narrowly in country reports to the Convention on Biological Diversity: Current actions and indicators are insufficient. *Biol. Conserv.* 261, 109233


The importance of timely metadata curation to the global surveillance of genetic diversity. *Conserv. Biol.* DOI: 10.1111/cobi.14061


*Introduction to Quantitative Genetics,* (4th edn), Longman

*Conserv. Biol.* DOI: 10.1111/cobi.14061


*Ecography.* 44:206–218

*Ecology.* 73:262–277


*Ecol. Lett.* 7:1225–1241

*The Theory of Ecological Communities,* Princeton University Press


*The Unified Neutral Theory of Biodiversity and Biogeography,* Princeton University Press


The contribution of species–genetic diversity correlations to the understanding of community assembly rules. *Oikos* 126, 759–771

Genetic diversity in widespread species is not congruent with species richness in alpine plant communities. *Ecol. Lett.* 15, 1439–1448

Latitudinal variation in genetic divergence of populations and the potential for future speciation. *Evolution* 58, 938–945


Box 1. Components of genetic diversity

*Genetic diversity* typically refers to the diversity of a local population that is relevant for evolutionary potential. Specific aspects of genetic diversity have different metrics, implications, and planning needs, especially when moving from single species to multispecies conservation (Table 1).

**Neutral and adaptive diversity.** Neutral genome-wide diversity is a central metric in conservation genetics[76]. It is positively correlated with functional genetic diversity[77]—though weakly[69]—and with individual fitness, and it is straightforward to estimate. Adaptive genetic diversity is diversity at loci underlying traits that affect fitness. Targeting adaptive genetic diversity is currently uncommon in applied conservation genetics[78]; this is difficult without knowledge of the traits, and genes, underlying adaptation to a given environment. However, diversity patterns in genes of known function have exciting potential for testing hypotheses about how species assemblages may be shaped by selective pressures acting in common across species. For example, Yiming et al.[79] studied latitudinal patterns of diversity and the strength of selection on the mammalian major histocompatibility complex, which may be related to parasite defense. Diversity patterns in protein-coding mitochondrial genes likely also have intriguing relationships with biodiversity gradients because mitochondrial genes are important to climatic adaptation[80]. The underlying drivers of adaptive and neutral genetic diversity fundamentally differ, and considering them separately in conservation macrogenetics will be important. For example, Xuereb et al.[81] showed in California sea cucumber that southern regions are prioritized to maintain genome-wide diversity, while northern populations are prioritized for climate change adaptation.
**Genetic diversity and differentiation.** Like species biodiversity, genetic biodiversity can be partitioned into alpha (genetic diversity of a local population) and beta diversity (population differentiation), both of which are necessary conservation considerations\[82\]. Alpha diversity governs population adaptive potential, while beta diversity represents diversity accumulated across a network of potentially locally adapted populations. Both are important for conservation, however may require different approaches. High alpha diversity is necessary to ensure sufficient genetic variation to avoid reduced population fitness due to inbreeding, and enable genetic adaptation to environmental change\[83\]. Beta diversity, however, is highest when populations are most isolated—such populations likely have lower genetic diversity, but may harbor rare alleles or adaptations specific to their habitat. A balance of population connectedness is critical: gene flow increases population resiliency, but local adaptation in response to spatially varying selection is most likely when rates of gene flow are low\[84\].
Box 2. Conceptually bridging genes and species

Many conceptual overlaps exist across the study of biodiversity at genetic and species levels[85–87]. Ecology typically takes a top-down approach, inferring process from pattern, while population genetics builds from the bottom-up. Joining these complementary approaches is a powerful way to understand the processes producing and maintaining biodiversity.

Species-genetic diversity correlations (SGDCs). The SGDC concept frames relationships between the genetic diversity of populations and the species diversity of communities[85]. SGDCs apply the four basic evolutionary forces of population genetics—gene flow, selection, genetic drift, and mutation—to communities, drawing parallels to migration, species-level selection, ecological drift, and speciation, respectively[85]. SGDCs were predicted to be generally positive and strongest when assessed with neutral markers in discrete habitats like oceanic islands[86]. However, many factors could decouple population and community sizes to generate negative SGDCs (Box Fig. 1). The evolutionary process in question (adaptive or neutral) could generate positive or negative SGDCs of varying strength depending on the focal species, its traits, and environmental context[12,88–90]. Therefore, much finer understanding of SGDCs is needed to understand if, for instance, protected areas and migration corridors will affect genetic diversity in the same way they protect species (See 3.2 Spatial conservation planning).

Diversity gradients. Species richness patterns, particularly the latitudinal richness gradient, have inspired several hypotheses about the processes generating biodiversity. Latitudinal gradients in genetic diversity and correlations to species richness have been commonly tested in macrogenetics[5,54–58,91–94]. The hypotheses[94] most commonly tested at the genetic level relate to ecological limits (resource-based limits on the sizes of populations and communities an
area can support) and evolutionary speed (varying speciation and extinction rates). Based on neutral theory in population genetics and macroecology[87], ecological limits hypotheses rely on parallel evolutionary processes acting on populations and communities as determined by resource availability and abundance. The global conservation policy implication of a latitudinal trend in genetic diversity would be that more funding and resources are needed in central latitude countries for protecting genetic biodiversity—a key topic in global negotiations of the recent Convention on Biological Diversity[95,96]. A related gradient are urbanization gradients. Recent studies have found that urban areas exhibit lower genetic diversity[63]—which has implications for Target 12 of the CBD calling for high quality green space for all people in urban areas.

**Figure I.** SGDC concept. Neutral evolutionary forces (ecological/genetic drift, left and migration/gene flow, right) acting on population and community levels are not always parallel. Mismatches can occur between genetic and ecological drift depending on the sizes of populations within communities (left). Migration (right) may not result in gene flow which will differentially affect community composition, in addition to the strength of drift on population and community levels. SGDCs can thus be positive, negative, or nonexistent; this limits our ability to predict genetic diversity from species richness alone.
Table 1. Moving beyond single species conservation genetics requires considerations for sampling, interpretation, and policy. Spatial conservation planning and indicators for monitoring are two policy areas that could benefit from macrogenetics research to effectively protect genetic diversity using the best available knowledge and data. Conservation targets are presented with example genetic metrics, and additional factors that should be considered when working with multispecies data.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Target</th>
<th>Metric</th>
<th>Multispecies considerations</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spatial conservation planning</strong></td>
<td>$N_e$, genetic diversity, genetic</td>
<td>Contemporary and coalescent $N_e$, GD, AR, $\pi$; $F_{ST}$, $G_{ST}$, population-specific $F_{ST}$, Jost’s D</td>
<td>Spatial prioritizations can differ when based on genetic data from one versus many species. Spatial patterns of multispecies genetic diversity and evidence of variable SGDCs in the literature suggest species richness is not a suitable proxy for the genetic diversity of populations in an area. Trade-offs between diversity and differentiation must also be considered, especially when isolated populations may be evolutionarily significant.</td>
<td>[4,5,9,12,90]</td>
</tr>
<tr>
<td><strong>Directing conservation funding</strong></td>
<td>$N_e$, genetic diversity</td>
<td>Contemporary and coalescent $N_e$; GD, AR, $\pi$</td>
<td>Regions where many species have especially high or low genetic diversity may warrant specific protection. Identifying national and international patterns (e.g., latitudinal gradients) of multispecies genetic diversity to can factor into decisions about where to direct conservation resources.</td>
<td></td>
</tr>
<tr>
<td><strong>Indicators</strong></td>
<td>$N_e$</td>
<td>Proportion of populations with $N_e &gt; 500$</td>
<td>Designed to protect and maintain populations with large $N_e$ to reduce rate of genetic diversity loss from genetic drift. However, $N_e = 500$ may be very low or high depending on species. Simple adjustments to this threshold have been suggested that are based on species ploidy. Multispecies genetic data can be used to empirically derive recommended thresholds for different taxonomic groups.</td>
<td>[17,23,97]</td>
</tr>
<tr>
<td>Genetic variation in $N_e/N_c$</td>
<td>Identifies species with lower abundance than would be expected given long-term effective population size. Based on multispecies data and is suitable for interspecies comparison because it accounts for baseline differences in effective population size across species. However, this metric is data intensive and requires whole genomes to estimate Tajima’s $D$ and coalescent $N_e$.</td>
<td>[24]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genetic diversity</td>
<td>Allelic richness</td>
<td>Based on allele counts, and changes more quickly following population size change than evenness metrics (e.g., gene diversity), making it informative for conservation and monitoring. Allele counts should be standardized with rarefaction to account for sample size differences before comparing across populations and species.</td>
<td>[11,17,98]</td>
<td></td>
</tr>
<tr>
<td>Species Habitat Index</td>
<td>Estimates changes in the sizes and connectivity of populations based on fine-grain species distribution information and remote sensing-supported capture of changes in suitable habitat. In the absence of range-wide genetic sampling for many species, the Index could potentially be a proxy for contemporary trends in genetic diversity for a large and representative portion of biodiversity. This possibility is not yet tested.</td>
<td>[25,39]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IUCN Red List status</td>
<td>Commonly used as a proxy for genetic diversity, yet is not reliably related to genetic diversity. IUCN status more likely reflects species-level characteristics like range size that lead to genetic diversity variation across species. As a species-level designation, it does not consider intraspecific variation, is biased by differences in genetic diversity at mutation-drift equilibrium across species, and does not capture genetic diversity declines.</td>
<td>[6,24,47,99]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Interspecific averages of genetic diversity metrics across species are influenced by species composition and differences in average genetic diversity across species. Depending on the data, the resolution of aggregation can create unrealistic populations or communities. Community-averaged genetic diversity values have unclear interpretations. [55–58,67,68]

**Genetic differentiation**

$G_{ST}$ metric extended to multiallelic markers like microsatellites. Pairwise $F_{ST}$ for biallelic markers like SNPs varies between 0 and 1 and is comparable across species. The maximum value of $G_{ST}$ is the average homozygosity of the measured populations. $G_{ST}$ cannot be directly compared across species, nor should thresholds be applied across species to define differentiated populations. [11,100,101]

Abbreviations: AR = allelic richness; EBV = Essential Biodiversity Variable; $F_{ST}$ = fixation index; GD = gene diversity; $G_{ST}$ = fixation index for multiallelic markers; $N_e$ = effective population size, $N_c$ = census population size; $\pi$ = nucleotide diversity; SGDC = species-genetic diversity correlation; SNP = single nucleotide polymorphism
Figure legends

Figure 1. Integrating macrogenetics into existing conservation frameworks. Conservation macrogenetics complements single-species, local-scale conservation genetics by synthesizing and generalizing species-specific inferences for global conservation policy. In turn, macrogenetics approaches can lead to top-down policies to protect genetic diversity, and can be used to identify regions or taxa at risk of genetic erosion that warrant species-specific focus.

Multispecies genetic diversity could be considered simultaneously with other levels of biodiversity including, but not limited to, species richness, phylogenetic, functional, or ecosystem diversity to enhance spatial conservation planning.

Figure 2. Practical applications for conservation macrogenetics. Leveraging openly accessible genetic data can support genetic monitoring and conservation decision making. Strengthening our understanding of how genetic diversity is related to species traits, distributions, environments will pave the way for model-based prediction approaches to fill in genetic data gaps by borrowing strength from other available data. Spatial genetic data can inform decision making for spatial prioritization, and genetic data sampled over time are a resource for tracking changes in populations that are not formally monitored or managed.
Figure 1
Figure 2