MOLECULAR ECOLOGY WILEY

Connecting high-throughput biodiversity inventories: Opportunities for a site-based genomic framework for global integration and synthesis

Paula Arribas[1] | Carmelo Andújar[1] | Martin I. Bidartondo[2,3] | Kristine Bohmann[4] | Éric Coissac[5] | Simon Creer[6] | Jeremy R. deWaard[7,8] | Vasco Elbrecht[9] | Gentile F. Ficetola[5,10] | Marta Goberna[11] | Susan Kennedy[12,13] | Henrik Krehenwinkel[13] | Florian Leese[14,15] | Vojtech Novotny[16,17] | Fredrik Ronquist[18] | Douglas W. Yu[19,20,21] | Lucie Zinger[22] | Thomas J. Creedy[23] | Emmanouil Meramveliotakis[24] | Víctor Noguerales[24] | Isaac Overcast[22,25] | Hélène Morlon[22] | Alfried P. Vogler[2,23] | Anna Papadopoulou[24] | Brent C. Emerson[1]

[1]Island Ecology and Evolution Research Group, Instituto de Productos Naturales y Agrobiología (IPNA-CSIC), San Cristóbal de la Laguna, Spain

[2]Department of Life Sciences, Imperial College London, London, UK

[3]Comparative Plant and Fungal Biology, Royal Botanic Gardens, London, UK

[4]Section for Evolutionary Genomics, Faculty of Health and Medical Sciences, Globe Institute, University of Copenhagen, Copenhagen, Denmark

[5]Université Grenoble Alpes, CNRS, Université Savoie Mont Blanc, LECA, Laboratoire d'Ecologie Alpine, Grenoble, France

[6]School of Natural Sciences, Bangor University, Gwynedd, UK

[7]Centre for Biodiversity Genomics, University of Guelph, Guelph, Canada

[8]School of Environmental Sciences, University of Guelph, Guelph, Canada

[9]Centre for Biodiversity Monitoring (ZBM), Zoological Research Museum Alexander Koenig, Bonn, Germany

[10]Department of Environmental Sciences and Policy, University of Milano, Milano, Italy

[11]Department of Environment and Agronomy, INIA, Madrid, Spain

[12]Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology Graduate University, Onna-son, Japan

[13]Department of Biogeography, Trier University, Trier, Germany

[14]Aquatic Ecosystem Research, Faculty of Biology, University of Duisburg-Essen, Essen, Germany

[15]Centre for Water and Environmental Research (ZWU) Essen, University of Duisburg-Essen, Essen, Germany

[16]Biology Centre, Institute of Entomology, Czech Academy of Sciences, Ceske Budejovice, Czech Republic

[17]Faculty of Science, University of South Bohemia, Ceske Budejovice, Czech Republic

[18]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden

[19]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

[20]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China

[21]School of Biological Sciences, University of East Anglia, Norwich, UK

[22]Institut de Biologie de l'ENS (IBENS), Département de biologie, École normale supérieure, CNRS, INSERM, Université PSL, Paris, France

[23]Department of Life Sciences, Natural History Museum, London, UK

[24]Department of Biological Sciences, University of Cyprus, Nicosia, Cyprus

[25]Division of Vertebrate Zoology, American Museum of Natural History, New York, USA

Correspondence
Paula Arribas, Island Ecology and Evolution Research Group, Institute of Natural Products and Agrobiology (IPNA-CSIC), San Cristóbal de la Laguna, Spain.
Email: pauarribas@ipna.csic.es

Abstract

High-throughput sequencing (HTS) is increasingly being used for the characterization and monitoring of biodiversity. If applied in a structured way, across broad geographical scales, it offers the potential for a much deeper understanding of global biodiversity through the integration of massive quantities of molecular inventory data generated independently at local, regional and global scales. The universality, reliability and efficiency of HTS data can potentially facilitate the seamless linking of data among species assemblages from different sites, at different hierarchical levels of diversity, for any taxonomic group and regardless of prior taxonomic knowledge. However, collective international efforts are required to optimally exploit the potential of site-based HTS data for global integration and synthesis, efforts that at present are limited to the microbial domain. To contribute to the development of an analogous strategy for the nonmicrobial terrestrial domain, an international symposium entitled "Next Generation Biodiversity Monitoring" was held in November 2019 in Nicosia (Cyprus). The symposium brought together evolutionary geneticists, ecologists and biodiversity scientists involved in diverse regional and global initiatives using HTS as a core tool for biodiversity assessment. In this review, we summarize the consensus that emerged from the 3-day symposium. We converged on the opinion that an effective terrestrial Genomic Observatories network for global biodiversity integration and synthesis should be spatially led and strategically united under the umbrella of the metabarcoding approach. Subsequently, we outline an HTS-based strategy to collectively build an integrative framework for site-based biodiversity data generation.

KEYWORDS
biodiversity assessment, DNA metabarcoding, Genomic Observatories, harmonized data generation, high-throughput sequencing

1 | INTRODUCTION

High-throughput sequencing (HTS) is increasingly being used for the characterization and monitoring of ecosystems and holds the prospect for a much deeper understanding of global diversity on Earth (e.g. Bohan et al., 2017; Bush et al., 2017; Taberlet et al., 2018). Deep DNA sequencing of biodiversity conducted at particular sites (i.e., site-based biodiversity assessment) presents opportunities for both detailed characterization and monitoring of local biomes. If replicated across other sites, broader scale biodiversity patterns/trends may emerge. Studying these trends not only allows better understanding of the processes that generate them, but is also highly relevant for (i) developing conservation and management policies, (ii) mapping living resources to allocate funding and efforts to biodiversity research, and (iii) comparing and understanding ecosystems (Pedrós-Alió & Manrubia, 2016). We may be witnessing the emergence of a new scientific paradigm, in which the conjunction of massive publicly available data sets and computational power lead to patterns and underlying causes being sought directly, rather than through more traditional hypothetico-deductive methods (Hey et al., 2009). If such a transition is to occur in biodiversity science, HTS can play a central role. The universality, reliability and efficiency of HTS data can potentially facilitate linking data among assemblages from almost anywhere (including elusive domains referred to as last biotic frontiers), at multiple biological levels (genes, populations, species, lineages), for any taxonomic group and regardless of prior taxonomic knowledge (Bik et al., 2012; Deiner et al., 2017; Ji et al., 2013). Projects from local to global scales are currently generating massive quantities of biodiversity inventories (i.e., community-level data) derived from HTS. These offer unprecedented opportunities to better understand both the distribution and the dynamics of biodiversity, and to propose strategies for its conservation. However, collective international efforts are required to optimally facilitate global integration and synthesis. While integrative frameworks for site-based genomic science do exist in the microbial realm (e.g., Gilbert et al., 2010, 2014), such frameworks have yet to be extended to nonmicrobial biotas.

From the November 11 to 13, 2019, evolutionary geneticists, ecologists and biodiversity conservationists involved in diverse large-scale initiatives based on HTS biodiversity assessments met at the University of Cyprus for the Next Generation Biodiversity Monitoring Symposium. Funded through iBioGen, a European Commission-funded (H2020) project, the symposium was conceived to discuss how best to develop an integrative HTS framework for measuring and understanding global patterns of biodiversity. The main objectives were to: (i) assess the current state of site-based biodiversity science using DNA-sequence information; and (ii) identify the challenges and opportunities for implementing a network for global integration and synthesis, featuring HTS as a core tool and with particular emphasis on the nonmicrobial terrestrial realm. In this review we synthesize the diversity of ideas raised over the 3-day symposium, and discuss their relevance for a more integrative genomics-informed biodiversity science.

2 | DNA SEQUENCE DATA: FROM GLOBAL REPOSITORIES TO GENOMIC OBSERVATORIES FOR INTEGRATION AND SYNTHESIS IN BIODIVERSITY SCIENCE

The systematic archiving of DNA-sequence data within global molecular repositories by the International Nucleotide Sequence Database Collaboration (DDBJ, EMBL-EBI and NCBI) emerged from the need to guarantee the preservation of, and accessibility to, genetic sequence data, a universal data type that links all biodiversity. These public repositories have the potential to serve as reference databases, providing the bedrock for global integration and synthesis in biodiversity science, in a similar fashion and complementary to other repositories focused on species occurrence records (e.g., GBIF, http://www.gbif.org/). Although the potential of DNA sequence repositories to document global patterns of genetic and phylogenetic diversity has been demonstrated (Holt et al., 2013; Miraldo et al., 2016), important limitations have been identified, primarily linked to: (i) a lack of standardized metadata associated with DNA sequences and (ii) taxonomic, geographical and sampling biases associated with these public repositories (Deck et al., 2017; Pope et al., 2015; Yilmaz et al., 2011).

The Barcode of Life Data Systems (BOLD; Ratnasingham & Hebert, 2007; http://boldsystems.org/) emerged as a curated workbench for the storage of specific DNA sequences under the umbrella term DNA barcoding (Hebert et al., 2003), addressing, in part, the metadata gap within global molecular repositories. The value of a taxonomically accurate and georeferenced DNA sequence database, even if focused on a single locus, has been demonstrated by analyses revealing how genetic diversity is structured at multiple scales (deWaard et al., 2019; Muñoz, 2007; Theodoridis et al., 2020). Additionally, numerous local–regional studies have benefitted from the contextualization of local data across larger spatial scales (Ashfaq et al., 2017; Cicconardi et al., 2017). More recently, the Genomic Observatories Metadatabase (GEOME; http://www.

geome-db.org/) has been provided as an open access repository for metadata associated with genetic data (or biosamples) of any type, linking ecologically and evolutionarily relevant metadata with publicly archived DNA sequence data (Deck et al., 2017; Riginos et al., 2020). GEOME thus provides a data management workflow to facilitate the use of public molecular repositories for global integration and synthesis.

Developments to address the metadata gap linked to genetic data are encouraging, but most records within public repositories are incidental point records (i.e., single specimens) that lack information about co-observed species (Jetz et al., 2019). This also imposes important constraints for documenting global biodiversity patterns, even if such point estimates are well contextualized with metadata. These constraints are comparable to those described for the GBIF database (Faith et al., 2013), including important biases for taxonomic and sampling coverage that are difficult to resolve in the short to medium term with additional point records. Inventories (i.e., site-based recording of multiple species) are a key biodiversity data type harbouring much potential for model-supported assessment of spatial biodiversity and its change (Jetz et al., 2019). However, the use of site-based data for the assessment of biodiversity distribution and change has been more a promise than a reality until recently (but see Phillips et al., 2019; van den Hoogen et al., 2019), and further efforts on data generation, standardization, integration and re-use are needed (Guralnick et al., 2018; König et al., 2019). In the case of DNA sequence data, HTS provide the potential to shift from incidental observations toward biodiversity inventories by (i) generating community-scale biodiversity data in a consistent and comparable manner; (ii) stimulating development of automated processing pipelines; and (iii) radically expanding taxonomic coverage, and thus increasing potential for synthetic analyses of global biodiversity (Bush et al., 2017, 2019). A pioneering international effort to generate site-based molecular information for global biodiversity synthesis has been focused on microbes, primarily by the Earth Microbiome Project (EMP; Gilbert et al., 2010, 2014; https://earthmicrobiome.org). The EMP was founded in 2010, with the core objective to construct a global microbial map, and has been catalysed by the explosion of metagenomics studies characterizing microbiomes worldwide. It has grown exponentially, now comprising a collaborative network of more than 500 researchers from 161 international institutions. It provides protocols and standards, and supports prepublication data sharing and crowd-sourcing data analysis to enable universal patterns of microbial diversity to be explored (Shoemaker et al., 2017; Thompson et al., 2017).

In an analogous initiative to the EMP, but one which extends to complete ecosystems, Davies et al. (2012) advocated the establishment of an international network of Genomic Observatories (GOs), with the subsequent publication of a founding charter (Davies et al., 2014). Site-based genomic inventories were proposed as a fundamental data source for biodiversity science, but with less focus on broad spatial coverage. Rather, it was suggested that ecological understanding could advance by focusing massive sequencing efforts on a limited number of model ecosystems. Concept definitions

within the founding charter of the GO network are as follows. A GO is defined as an ecosystem or site subject to long-term scientific research, including (but not limited to) the sustained study of genomic biodiversity (the genetic variation found among genomes) from single-celled microbes to multicellular organisms. A GO network is a network of sites that fall within the GO definition, although the fuller definition incorporates researcher and institutional details (see Davies & Field, 2012; Davies et al., 2012, 2014). These definitions provide an abstract concept of site-based biodiversity measurement with three sampling dimensions: (i) genomic, (ii) temporal and (iii) spatial. The GO network ambition of a genome-centred characterization of biodiversity across both space and time is exciting. However, the GO concept itself is greatly ambitious, which may in part explain its limited adoption being restricted, so far, to initiatives primarily focused on marine meiofauna (Buttigieg et al., 2019; Kopf, 2015). The focus of the iBioGen meeting was to review the GO network concept and vision within the current state of the art and ongoing initiatives, and identify tangible opportunities to catalyse a site-based genomic network for global integration and synthesis of biodiversity data in the terrestrial realm.

3 | TOWARD A SITE-BASED TERRESTRIAL GENOMIC NETWORK FOR INTEGRATIVE AND SYNTHETIC BIODIVERSITY SCIENCE

Three major challenges have been identified for the implementation of GOs (Davies et al., 2012): (i) logistical constraints associated with field sampling, (ii) sequencing and data curation costs, and (iii) the collection of relevant metadata. In relation to metadata, it was sensibly suggested that GOs should integrate within existing research sites that are already rich in metadata. However, the logistical and financial challenges associated with field sampling and genomic data generation remain as much a limitation now as they were 8 years ago. Even if sequencing costs were to dramatically decrease, and technology were to dramatically improve, the ambition of the original GO concept would still remain beyond reach in the short to medium term. However, we believe that implementation of the conceptual framework of Davies et al. (2012, 2014) can be achieved with a realistic appreciation of logistical and financial constraints, and an attempt to integrate the three principal axes of the concept: space, time and genome.

3.1 | The spatial and temporal dimensions

Achieving a site-based genomic network, capable of providing data for global syntheses in biodiversity science, will require trade-offs with regard to the axes of space, time and genome. This is because integration across the three axes will vary depending on the question being asked. For example, estimating and mapping the number of species on Earth (*the Linnean challenge*), including taxonomic, phylogenetic and functional diversity, places heavy emphasis on

the spatial axis. Taxonomic information can be achieved with limited investment along the genomic axis (e.g., via metabarcoding or future alternatives), with more investment for phylogenetic information, and a much greater effort to examine functional diversity. Addressing the global estimation of species distributions and ecological niches (*the Wallacean challenge*) also falls most heavily along the spatial axis. Spatial sampling is important not only to tease apart the population- and community-level processes that impact macroecological and macroevolutionary patterns, but also to develop predictive models for the impact of global change on biodiversity and ecosystem functioning. The temporal axis is also particularly important here, as it can directly inform about resistance, resilience and dynamics of community change in the face of natural and anthropic disturbance. However, time series are extremely difficult to obtain and under some conditions time can even be substituted for space, such as using sediment cores across glacial retreat areas (Chen & Ficetola, 2020) or geologically dated chronosequences (Rominger et al., 2016).

The spatial dimension of GOs was implicit within the original definition of Davies et al. (2012) and was explicitly incorporated within the concept of the GO network (Davies et al., 2014). However, the genomic and temporal dimensions were given greater emphasis, with the spatial dimension assuming secondary importance. GOs were framed in the context of long time series for particular sites, or so-called "model ecosystems." Despite this earlier temporal emphasis, our discussions concluded that a sustained and coordinated effort is most likely to emerge through opportunities for a spatially based network, which will provide more immediate and broader opportunities for global integration. Site-based research across the ecoregions could efficiently integrate data into the Global Earth Observation System of Systems (http://www.geoportal.org/), one of the main objectives of the GO network as initially defined (Davies et al., 2014). Recent developments open the door to connect Earth Observation (EO) technologies to biodiversity and ecosystems (CEOBE; Bush et al., 2017) to derive broader scale correlations and inferences from continuous remote-sensing data with point-sample biodiversity data. This approach is anchored in the generation of site-based biodiversity data, which are then combined with EO variables to interpolate continuous maps of biodiversity variables. The generation of spatially dense, site-based biodiversity data across ecologically distinct regions is fundamental for parameterization within the model of Bush et al. (2017), and they propose HTS as a universally applicable tool for this purpose. We believe that a collaborative site-based genomic network, where spatial coverage is the primary focus, is a realistic ambition that bridges both the GO network and CEOBE visions to model continuous surfaces of biodiversity across whole regions, and eventually the planet.

We also see a more fundamental argument for a network of shallow (i.e., without extensive effort on the temporal axis) sampling points. Limited logistical constraints associated with shallow sampling mean that implementation is less prohibitive, enhancing the opportunity for independent uptake, and the development of a globally extensive network. We converged in opinion that the most effective

way to develop a GO network is to build upon and integrate among existing initiatives using HTS to generate inventory data at multiple spatial scales (see below for details). Additionally, international efforts for global microbial biodiversity syntheses, such as the EMP (Gilbert et al., 2014), the Ocean Sampling Day consortium (OSD; Kopf, 2015) and TARA Oceans (Karsenti et al., 2011), align with a spatially led vision for a GO network. These may serve as models for the development of best practices, methodological harmonization and analytical synthesis.

Reduced logistical challenges associated with data generation for a spatially led vision of the network are also likely to have positive statistical consequences. We see higher spatial replication as potentially reducing the degree of data harmonization required for global integration and comparison. Coverage for a large number of shallow sampled sites can correct for local-scale noise that would be more difficult to account for with a network of fewer temporally sampled sites. Finally, a GO network where shallow sampling of biodiversity is a cornerstone is a potential catalyst for a noncentralized, bottom-up network expansion, with potential for citizen-science to collect site-based biodiversity samples over large areas (e.g., Kopf, 2015).

3.2 | The genomic dimension

The concept of a GO network was originally envisaged as community-scale genome-level data capture, from single-celled microbes to multicellular organisms, through time (Davies et al., 2012, 2014). We agreed that this ambition of characterizing ecosystems at the deepest genomic dimension (whole-genome scale) is currently unrealizable, particularly within a spatially dense sampling framework for biodiversity synthesis. In the light of this, we sought to define a more plausible integration of genome sequencing technology and genome data within a global network for biodiversity observation. Agreement was reached that less exhaustive genomic sampling approaches can provide foundational site-based biodiversity information, where data standards can be relatively easily implemented. Of high importance is that relative cost efficiency helps to democratize implementation, facilitating network expansion in economically limited tropical countries, which are megadiverse, often under greater threats and much less sampled (Zinger et al., 2020).

A single-locus approach represents the lowest level of genomic investment for comparative site-based biodiversity assessment. DNA barcoding is a simple idea wherein short genomic regions that typically differ between species can be used to assign some level of taxonomic identity to an unknown specimen (Hebert et al., 2003). The extension of this method to more complex samples where multiple specimens/species are present has given rise to the term DNA metabarcoding (Taberlet et al., 2012; Yu et al., 2012) involving massively parallelized HTS characterization of taxonomic composition within a biological sample. This reduces the logistical and the financial costs associated with individualized sampling and sequencing of specimens (Ji et al., 2013). These are transversal benefits that are particularly relevant for hyperdiverse assemblages such

as those from soil (Arribas et al., 2016) or tropical canopies (Creedy et al., 2019). Individualized processing of specimens from such assemblages is a major bottleneck, and as such they are recurrently ignored in conventional biodiversity surveys. Metabarcoding can additionally be deployed to inventory diversity from degraded tissues or without direct sampling of organisms, as exemplified by "environmental DNA" (eDNA; Deiner et al., 2017) and "invertebrate-derived DNA" (Schnell et al., 2012), thus maximizing diversity recovery from biosamples.

DNA metabarcoding involves PCR (polymerase chain reaction)-coupled HTS of one or more DNA barcode markers, directly from bulk or environmental samples, represents the most cost-efficient approach for obtaining molecular community profiles (Porter & Hajibabaei, 2018), and is already established as a reference tool to map global microbial biodiversity (e.g., Karsenti et al., 2011; Kopf, 2015). It is highly scalable within a relatively simple methodological framework, facilitating increased sampling breadth and depth within biodiversity studies. Importantly, it offers many opportunities for standardization, reproducibility and sample handling efficiency.

Metabarcoding can generate biodiversity inventory data in which taxonomy can be assigned at a sufficiently informative taxonomic rank for the reliable, relative estimation of richness and composition of communities. Conventional processing of raw metabarcoding data clusters sequences based on their similarity, resulting in operational taxonomic units (OTUs), for which taxonomic assignment to species is possible when reference barcode sequences are available. However, even without species-specific reference libraries, assignment to some taxonomic level can be achieved using public repositories (e.g., GenBank, BOLD, SILVA and UNITE). The fundamental output of metabarcoding is ecological tables of samples by taxa (OTUs), with the key advantage that this can be achieved at an unprecedented scale and cost-efficiency. By generating species inventories across a network of sites, metabarcoding data can be used to estimate ecological/trophic networks through co-occurrence analysis (Bohan et al., 2017). Although there are important limitations associated with inferring interactions from co-occurrence data (Blanchet et al., 2020), geographically extensive data can provide hypothetical frameworks within which more directed sampling and sequencing (e.g., by the analyses of gut contents, Alberdi et al., 2019), can be undertaken within particular GOs (see Box 1).

While we reached a consensus that metabarcoding is the current candidate tool for a GO network, limitations were also recognized. Deriving abundance information from metabarcoding data remains a challenge, primarily due to inherent biases in the PCR step, but also because of variation in gene copy number, organelle number, and technical aspects of sampling, laboratory, sequencing and bioinformatic workflows (Deagle et al., 2019; Taberlet et al., 2018; Zinger et al., 2019). Generally speaking, it has been suggested that the most conservative approach is to treat DNA metabarcoding data as presence–absence data after applying a high filtering threshold based on read relative abundance to mitigate the impact of the false positives (Elbrecht et al., 2015; Taberlet et al., 2012). For general patterns of community assembly, it has been shown that ecological

Box 1 *SuperGOs*: extending the temporal and genomic axes of the GO network

Within the spatially led terrestrial GO network that we propose, where metabarcoding is at the core of data generation, both the temporal and the genomic axes can be more deeply sampled (Figure 1), consistent with the idea of "model ecosystems" (Davies et al., 2012, 2014). As well as potentially providing a deeper understanding of dynamics at a local scale, these additional layers of information may also serve as: (i) calibrations within the global network (e.g., for the assessment of inventory completeness, validation of diversity estimations using different sources of genomic information); (ii) sites where the interplay between different dimensions of diversity (e.g., genetic, taxonomic, phylogenetic, functional, interaction) can be assessed in depth; or (iii) sites where the periodic implementation of a particular module or modules could be undertaken to generate long-term HTS biodiversity data series. Due to the more intense nature of data generation associated with such sites, we refer to them as "SuperGOs." Temporal sampling could take advantage of sites with rich histories of data collection (e.g., sites within the Long-Term Ecosystem Research network, https://www.ilter.network/; or ICP Forests, http://icp-forests.net), or existing national biomonitoring frameworks (e.g., the UK Countryside Survey, https://countrysidesurvey.org.uk/; the German Environmental Specimen Bank, https://www.umweltprobenbank.de/en). Alternatively, sites with natural temporal records, such as lake sediment cores (e.g., Chen & Ficetola, 2020; Pansu et al., 2015) or organic deposits (e.g., in feco-urinary middens, Murray et al., 2012), may provide historical context for more extensive forward temporal sampling (reviewed in Bálint et al., 2018). SuperGOs could also be a focus for generating genomic resources (i.e., partial or complete genomes, microbiomes, diet) for well-characterized taxa, or species of specific interest within sites. For example, a series of spatially clustered sites, well characterized for their terrestrial arthropod fractions of biodiversity with metabarcoding, could be used to provide broader context for understanding how habitat structure, prey availability and feeding preferences influence vertebrate insectivores (Makiola et al., 2020). Increasing the depth of genomic sampling within a site, from barcode regions, to whole organelle genomes, to partial nuclear genomes and finally to whole nuclear genomes, will in itself generate new resources that may promote further synergy. As an example, for any annotated genome produced within a SuperGO, the associated barcode sequence metadata for that species allow its presence to be easily established in any other GO. If pertinent (e.g., due to inter-GO variation for climatic or biotic variables), targeted gene sequencing or whole genome resequencing could be implemented in additional GOs to explore potential adaptive variation.

indices using abundance or presence-only data perform similarly well (Beentjes et al., 2018; Creedy et al., 2019; Ranasinghe et al., 2012). Considering read abundance as a probability of presence has also been shown to provide alpha/beta diversity estimates that are closer to those obtained with traditional approaches (Calderón-Sanou et al., 2019). The use of taxon-specific correction factors and occupancy models shows great promise for improving abundance estimates from metabarcoding data (e.g., Kembel et al., 2012; Krehenwinkel et al., 2017) and for some taxa it has been shown that read abundance can be used to estimate actual abundance (Schenk et al., 2019). However, this requires high taxon-specific effort, and thus implementation will probably be limited with regard to complex and/or hyperdiverse communities, at least in the near future.

As well as recording species presence, metabarcoding can potentially provide a measure of haplotype variation within and across communities. Recent advances in the removal of both PCR and sequencing errors (Amir et al., 2017; Callahan et al., 2016; Edgar, 2016), and co-amplified nuclear copies of mitochondrial genes (Andújar et al., 2020), open the door for bypassing sequence clustering steps, thus yielding haplotype-level data (amplicon sequence variants) for downstream analyses (Elbrecht et al., 2018; Turon et al., 2020). Thus, alpha and beta diversity can be analysed at different hierarchical levels to understand how population- and community-level processes drive macroecological and macroevolutionary patterns (Arribas et al., 2020; Tsuji et al., 2020). This possibility of recording data in the form of haplotype-level community tables would improve comparability of biodiversity surveys from independent studies. This is because individual haplotypes are directly comparable between independently processed data sets, while OTUs are not (Callahan et al., 2017). We thus believe that appropriately denoised haplotype tables would be a substantial step in integration among independent metabarcoding data sets.

Metabarcode data provide, to some extent, phylogenetic information among community members, and this has been exploited to reveal phylodiversity patterns (alpha and beta) across communities, mostly in prokaryotes (e.g., Fierer & Jackson, 2006; Goberna et al., 2016), and evolutionary processes across entire lineages (e.g., diatoms, Lewitus et al., 2018). We recognize the inherent limitations of short metabarcode sequences for phylogenetic resolution, but note that this can be partly overcome by placing metabarcode sequences in backbone phylogenies (Balaban et al., 2018; Pérez-Valera et al., 2018). A phylogenetic framework can simultaneously (i) assist species delimitation and taxonomic assignment, through the phylogenetic placement of the massive quantities of unidentified OTUs/haplotypes, and (ii) contribute tip-level data to the Tree of Life (Hinchliff et al., 2015). We therefore see potential for synergy between a GO network built upon community-level metabarcoding, and other global initiatives developing phylogenomic (e.g., the Earth Biogenome Project, Lewin et al., 2018) and taxonomic resources (e.g., BOLD, Ratnasingham & Hebert, 2007). Organelle skimming of mixed community samples can directly provide the phylogenetic context of a local sample (Crampton-Platt et al., 2015) which, when combined with samples
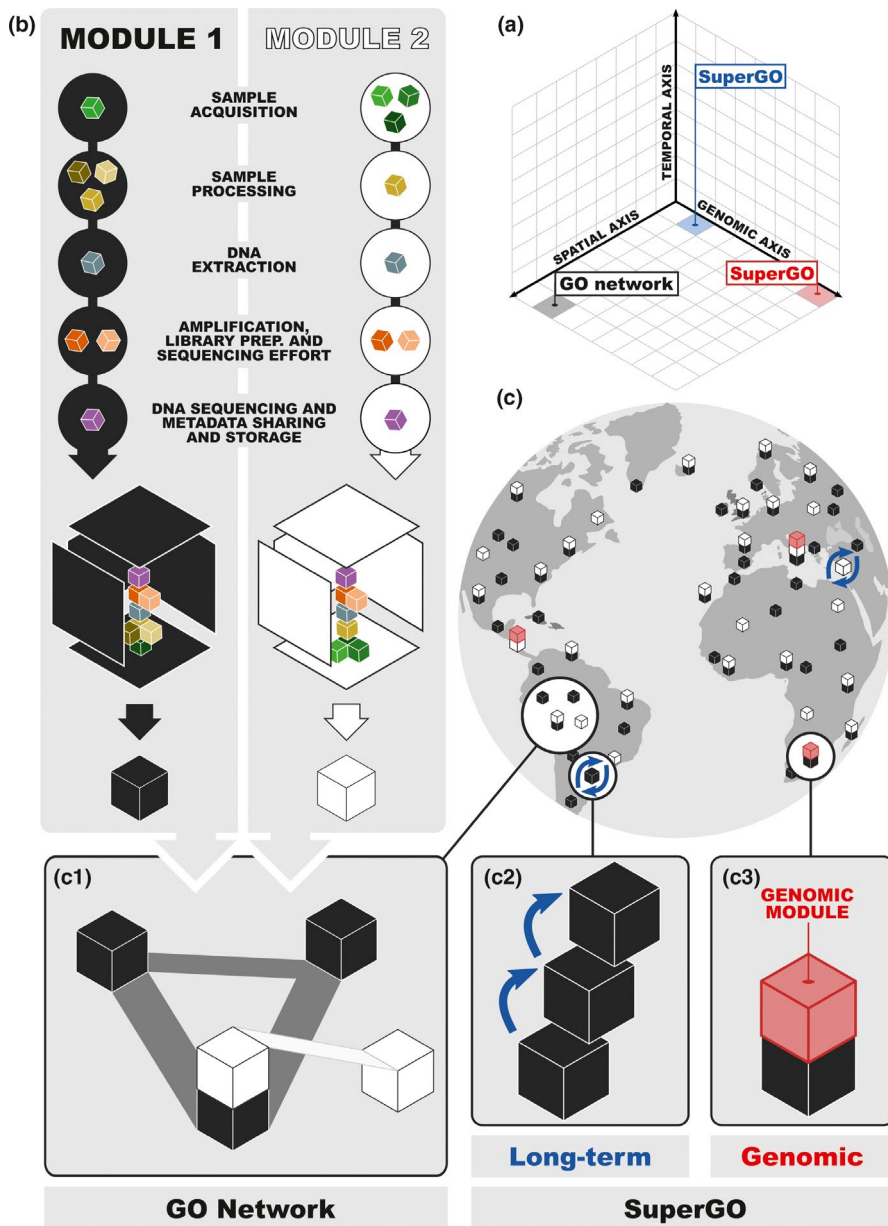
FIGURE 1 Conceptual representation of the modular framework for harmonized data generation within a spatially led Genomic Observatories (GOs) network. (a) Representation of the three principal axes of the GO concept: space, time and genome. (b) Representation of the modular structure for protocol harmonization across GOs. Modules, as basic building blocks of the framework, can provide simple, integrated and interoperable procedures for site-based characterization of extensive biodiversity fractions. Black block: module 1 (e.g., soil biodiversity); white block: module 2 (e.g., terrestrial arthropods); coloured blocks: protocols or submodules across the sequence of the five steps for a harmonized generation of inventory data within each module. (c) Map showing how modules can be implemented in: (c1) a spatially led network of sites for biodiversity inventory with shallow temporal and genomic efforts (spatial axis for the GO network concept); (c2) replicated through time within a site by repeated module implementation or historical samples (temporal axis for the SuperGO concept); and (c3) complemented with genomic resources (e.g., partial and complete genomes, microbiomes, gut contents) for specimens/assemblages within a site (genomic axis for the SuperGO concept)

from other sites, each contribute to an increasingly complete Tree of Life (e.g., SITE-100 project; www.site100.org). Such efforts to develop taxonomically curated and phylogenetically defined reference specimen repositories could be undertaken within SuperGOs (see Box 1).

With currently available DNA sequencing technology, other molecular techniques for site-based biodiversity analysis remain uncompetitive compared to metabarcoding, when simultaneously considering scalability, cost and tractability (Porter & Hajibabaei, 2018). Metagenomics (Handelsman et al., 1998) and metatranscriptomics (Poretsky et al., 2005) present a key advantage over metabarcoding for characterizing community samples, as their PCR-free protocols remove biases associated with the PCR amplification step of metabarcoding, thus making abundance estimates less challenging (e.g., Tang et al., 2015). These techniques sequence all DNA/RNA in a sample, and sequences

are then interrogated *in silico* for taxonomically and functionally informative gene sequences. However, for fundamental tasks of estimating species richness and taxonomic assignment, financial cost alone would dramatically limit uptake compared to metabarcoding. Additionally, while some bioinformatic skills are required for metabarcoding, the analysis of metagenomic data presents a more complex challenge, and would additionally act as a barrier to uptake. Nevertheless, metagenomics does enable a broader sampling along the genomic axis and could be deployed within a SuperGO (Box 1). Although metabarcoding will remain the dominant approach for quite a while, we have already witnessed transformative sequencing capacity change with existing HTS technology. In this sense, we also view the maintenance of bulk DNA extractions for sites (preferably using nondestructive protocols, e.g., Marquina et al., 2019; Nielsen et al., 2019), or even unprocessed sample replicates, as an important strategy (Jarman

et al., 2018). Doing so will ensure that site data can be updated as new HTS developments emerge, promoting long-term network efficiency and effectiveness.

Discussion within the symposium converged on the opinion that an effective, spatially led terrestrial GO network should place importance on maximizing the global distribution of sampling sites and that these should be strategically united under the umbrella of the single-locus metabarcoding approach (Figure 1a). Within this framework, we agreed that optimal comparison and integration among independently generated community data sets would be best served by the establishment of harmonized (cf. standardized, see Walters & Scholes, 2017) molecular approaches for site-based biodiversity surveys. The remainder of the symposium was thus focused on identifying how to collectively build an integrative framework for data generation, and the outcomes of this discussion are summarized in the following section.

4 | A FRAMEWORK FOR HARMONIZED DATA GENERATION WITHIN A SPATIALLY LED GO NETWORK

Generating biodiversity observations compliant with global data standards, and the need for efforts to develop the associated cyber-infrastructure, were an explicit element of early discussions regarding GOs and a GO network (Davies et al., 2014). Toward this goal, the Biocode Commons (http://biocodecommons.org/) was developed as an open community effort, to provide an open source platform to facilitate GO network conformity to existing biodiversity information standards in the taxonomic (Biodiversity Information Standards Organization, Wieczorek et al., 2012), genomic (Genomic Standards Consortium, Field et al., 2011) and data sharing (ISA Commons and BioSharing community, Field et al., 2009; Sansone et al., 2012) domains. Biocode Commons has helped focus efforts to solve the "metadata gap" associated with public molecular repositories, which has been addressed by the GEOME repository (Deck et al., 2017; Riginos et al., 2020). It allows data contributors to create customized yet standard-compliant spreadsheets that capture the temporal and geospatial context of each biosample (from incidental point records or inventories, and from Sanger sequences to HTS). GEOME represents a fundamental step toward global integration of DNA sequence-connected biodiversity data, where scientific reproducibility, synthesis and the potential for DNA sequence reuse are maximized (Riginos et al., 2020).

A spatially led, noncentralized GO network would require bottom-up growth, and many existing international, national and regional projects or initiatives demonstrate the potential for this through broadly shared goals of generating site-based biodiversity data using metabarcoding (Table 1). However, in addition to standardization for reporting taxonomic, genomic and metadata, we also see the need for both procedures and best practices for the generation of site-based metabarcoding data to be similarly standardized. Standardization is fundamental to maximize the potential

for integration and global syntheses, as limited integration of genomic approaches in biodiversity observation networks is mainly attributed to the lack of global, standardized and well-contextualized data sets and accompanying best practices (Canonico et al., 2019; Kahlert et al., 2019).

Within the marine realm, the recently constituted Global Omics Observatory Network (GLOMICON; https://glomicon.org/) promotes the alignment of protocols and information standards. These derive from the vast accumulated knowledge of large-scale surveys of ocean waters, to generate a framework of harmonized methods for long-term marine biodiversity observation (see Canonico et al., 2019). The terrestrial nonmicrobial realm has yet to experience such a development, but the time is right for the establishment of a standardized framework with metabarcoding at its core. The overarching goal of this framework should be to maximize user uptake and the global scale of the network through "harmonization" rather than "standardization" (Walters & Scholes, 2017). For that, various combinations of standards are favoured, and the minimum set of attributes of the framework is constrained, to allow intercalibrations, conversions and sorting of data among GOs. Such a framework could both maximize cross-GO integration of biodiversity data, and catalyse network growth. A key mechanism for site inclusion within the network should be the adoption of GO data generation standards.

A harmonized framework for metabarcoding data generation within a GO network can build on recent efforts on the harmonization of conventional inventory data. The Humboldt Core (Guralnick et al., 2018) has recently been proposed as a conceptual framework for capturing, in a standardized and general way, core information about processes underpinning inventory work. The objective of the Humboldt Core is to expand biodiversity data set discovery, interoperability and modelling utility for site-based biodiversity data, a data type essential for the assessment of biodiversity variation in space and time (Guralnick et al., 2018; Jetz et al., 2019). Although mostly focused on conventional inventories, it provides a useful scaffold for the development of a framework for a terrestrial site-led GO network by extending the philosophy of the Humboldt Core to metabarcoding data. The Humboldt Core places importance on the reporting of the broader scope of site-based data (spatial, temporal, taxonomic and environmental), and this is to some extent addressed within existing initiatives (GEOME, Deck et al., 2017; Riginos et al., 2020). However, it also emphasizes the need for reporting and harmonizing within the inventory process, methodology and effort. It is here, at the stage of generating biodiversity data, that the major gap for the implementation of the GO network exists, and thus where effort should be focused.

Metabarcoding analyses of site-based biodiversity are rapidly accumulating, but concern has been raised that the manner of this increase risks losing sight of the challenges in producing high-quality and reproducible data (Zinger et al., 2019). Indeed, development of best practices and standardized methods for the generation of metabarcoding data has been a recurrent topic in recent years across different communities (Andújar et al., 2018; Orgiazzi et al., 2015; Ransome et al., 2017). Efforts toward harmonization and best

TABLE 1 International, national and regional projects or initiatives with broadly shared goals with the proposed spatially led Genomic Observatories (GO) network and generating site-based biodiversity (terrestrial nonmicrobial) data using the metabarcoding approach. These initiatives were identified in the symposium with high relevance for building up the framework for harmonized data generation within the GO network

Project/initiative	Website	Institution/entity main holder	Spatial scope	Genomic approach	Target biodiversity fraction	Funding source
BIOSCAN	https://ibol.org/programs/bioscan/	University of Guelph	Global	Community metabarcoding—specimens microbiome metabarcoding	Terrestrial arthropods collected by Malaise traps	University of Guelph
SITE100	https://www.site100.org/	NHM London	Global	Community metabarcoding—mito metagenomics	Terrestrial arthropods collected by interception, pitfall and Malaise traps, soil arthropods extracted by FBF protocol	NHM Biodiversity Initiative
IceCommunities	https://cordis.europa.eu/project/id/772284	Universita degli studi di Milano	Global	eDNA metabarcoding	eDNA (bacteria, fungi, protists, plants, animals) from soil samples	European Research Council
GlobNet	https://leca.osug.fr/ANR-2021-2017-GlobNets	LECA	Global	eDNA metabarcoding	eDNA (bacteria, fungi, protists, plants, animals) from soil samples	ANR (2017–2021)
VIGILIFE	https://beauvalnature.org/en/conservation/programme/spygen-vigilife	SPYGEN	Global	eDNA metabarcoding	eDNA filtered water samples	WWF, CNRS, Bonneval Nature, Nat Geo, and others
GSSP	https://www.helsinki.fi/en/researchgroups/spatial-food-web-ecology/research/gssp/about-gssp	University of Helsinki	Global	eDNA metabarcoding	eDNA (fungi) filtered air samples	Academy of Finland
LIFEPLAN	https://www.helsinki.fi/en/projects/lifeplan	University of Helsinki	Global	Community metabarcoding—eDNA metabarcoding	eDNA (bacteria, fungi, protists, plants, animals) from soil samples, eDNA (fungi) filtered air samples, terrestrial arthropods collected by Malaise traps	European Research Council
LUCAS	https://esdac.jrc.ec.europa.eu/projects/lucas	EC Joint Research Centre	Europe	eDNA metabarcoding	eDNA (bacteria, fungi, protists, plants, animals) from soil samples	EC Joint Research Centre
DNAqua-Net	http://dnaqua.net/	University of Duisburg-Essen	Europe	Community metabarcoding—eDNA metabarcoding	Aquatic arthropods collected by mesh, eDNA (bacteria, fungi, protists, plants, animals) filtered water samples	COST
1000rivers	https://1000rivers.net/	NatureMetrics	Europe	eDNA metabarcoding—specimens microbiome metabarcoding	eDNA (fish) filtered water samples	NatureMetrics, Citizen groups

(Continues)

TABLE 1 (Continued)

Project/initiative	Website	Institution/entity main holder	Spatial scope	Genomic approach	Target biodiversity fraction	Funding source
Pacific Scanner		Okinawa Institute of Science and Technology	Pacific archipelagos	Community metabarcoding, gut content metabarcoding of individual specimens	Terrestrial arthropods collected by vegetation beating and leaf litter sifting	Okinawa Institute of Science & Technology (OIST) Kick-start fund
Insect Biome Atlas	https://insectbiomeatlas.org	Swedish Museum of Natural History	Sweden, Madagascar	Community metabarcoding	Terrestrial arthropods collected by Malaise traps and leaf litter sifting, eDNA (arthropods) from soil samples, insect-associated microbiomes	Knut & Alice Wallenberg Foundation
SCANDNAnet	https://www.syke.fi/en-US/Research__Developmen t/Research_and_devel opment_projects/Proje cts/SCANDANnet/SCAND NAnet(47361)	Finnish Environment Institute SYKE	Europe	Community metabarcoding—eDNA metabarcoding	Aquatic arthropods collected by mesh, eDNA (bacteria, fungi, protists, plants, animals) filtered water samples	Nordic Council of Ministers
CALeDNA	https://ucedna.com/	University of California	California	eDNA metabarcoding	eDNA (bacteria, fungi, protists, plants, animals) from soil samples	Catalyst Grant Program
SOILmesoDiv		IPNA-CSIC, NHM London	Spain, France, Canary Islands, Madeira, Madagascar	Community metabarcoding	Soil arthropods extracted by FBF protocol	The Royal Society, MINECO
BioWide	https://www.gbif.org/es/datas et/3b8c5ed8-b6c2-4264-ac52-a9d772d69e9f#descr iption	Århus University	Denmark	eDNA metabarcoding	eDNA (bacteria, fungi, protists, plants, animals) from soil samples	Villum Foundation
Dimensions		University of California	Hawaii	Community metabarcoding	Terrestrial arthropods collected by active searching	

practices have been described for some biodiversity fractions, such as microbiomes (Pollock et al., 2018), fungi (Kõljalg et al., 2013) or eDNA (Creer et al., 2016; Taberlet et al., 2018), but are much less developed for many others (e.g., for Metazoa community metabarcoding). The success of microbial initiatives has pivoted on harmonized protocols for microbial/seawater sampling, DNA extraction, library generation or sequencing (e.g., Caporaso et al., 2012; Marotz et al., 2017 for the EMP or Alberti, 2017; Gorsky et al., 2019; Kopf, 2015 for the TARA Oceans and the OSD). Discussion within the symposium identified the following five steps as important for harmonization across the inventory process, methodology and effort. (i) Sample acquisition—including sampling design (spatial scale, temporal scale and biological replicates), sampling method, sampling effort and sample storage. (ii) Sample processing (pre-DNA extraction)—potentially including protocols for sample cleaning, size selection, photo recording, extraction of vouchers, homogenization and sample storage. (iii) DNA xtraction—DNA extraction protocols, negative and positive controls, technical replicates and long-term storage. (iv) Amplification, library preparation and sequencing effort—including the target metabarcode fragment(s), sample screening, PCR and library preparation protocols, negative controls, mock standards for intercalibration, sequencing technology and minimum sequencing effort. (v) DNA sequence and metadata sharing and storage (Figure 1b). A challenge for the development of such a framework is that, while existing implementations are in general achievable with a single, or several, type(s) of site-based community sample (e.g., microscopic communities collected by filtering water/air or eDNA), capturing nonmicrobial biodiversity requires a more diverse sampling and sample processing approach.

We discussed the potentially complex sampling requirements that would typically be needed to capture all nonmicrobial fractions of terrestrial diversity, and concluded that a "modular" structure for data acquisition is the most promising approach. Modules, as basic building blocks of the framework, can provide simple, integrated and interoperable procedures for site-based characterization of broad biodiversity fractions. An appropriate module design would comprise a minimum set of protocols (or submodules) across the sequence of the five steps described above, which would facilitate the generation of inventory data across a wide taxonomic spectrum (if possible, complete biotas), in a simultaneous, integrated and cost-efficient way. Some modules would require a combination of multiple independent protocols (submodules) for the sample acquisition step, but a much simpler pipeline for further steps, for example in the case of terrestrial arthropods. However, other modules may be more efficiently executed with a single integrated sample acquisition protocol, but a more diverse structure for sample processing and DNA extraction, for example in the case of soil or freshwater biodiversity inventory (Figure 1b). Common submodules across different modules are also desirable to minimize overall complexity. A modular structure has been successfully incorporated in ForestGEO (https://forestgeo.si.edu), where individual trees are inventoried for multiple layers of biodiversity (e.g., arthropods, vertebrates or lianas) using harmonized pipelines.

We propose that similar success can be achieved across broader geographical and environmental scales, using metabarcoding as a shared element across modules designed for specific fractions of terrestrial biodiversity.

Biodiversity fractions for which (i) there is high potential for global integration from site-based data, and (ii) metabarcoding protocols, data sets or sampling programmes already exist (or are in an advanced stage of development), are obvious candidates for module development. Soil biodiversity emerged as a potential candidate, given that soils host the vast majority of life on Earth, provide ecosystem services that support above-ground food webs, and are also one of the last "biotic frontiers" of biodiversity (Bardgett & van der Putten, 2014). HTS approaches, and in particular metabarcoding, have been extensively applied to reveal community structure among microbial prokaryotes (bacteria and archaea) and fungi across most soil biomes, including global-scale initiatives (Delgado-Baquerizo et al., 2018; Tedersoo, 2014; Thompson et al., 2017). Thus, there is a solid foundation of methodological harmonization and data integration for the development of a soil biodiversity module. The application of metabarcoding to characterize the diversity of soil protist or fauna fractions has been limited (Arribas et al., 2016; Geisen, 2016), and approaches integrating whole soil biota are even more scarce (but see Zinger et al., 2019 using eDNA). Recent efforts to compile a guide of methods for obtaining taxonomic and functional profiles of soil biodiversity have identified metabarcoding as a fundamental tool (Geisen et al., 2019). Microbial, mesofauna, macrofauna and eDNA could be simultaneously sampled within a unified sampling design, including a combination of sample processing protocols to extract bulk communities (e.g., arthropods, nematodes, roots) and eDNA from the soil matrix. Methodological and logistical advances within ongoing large-scale soil biodiversity projects such as the LUCAS, GlobNet, LIFEPLAN or IceCommunities (see Table 1 for details) could be used to develop a soil biodiversity module.

Terrestrial arthropods also emerged as a biodiversity fraction with high potential for a modular approach, being a hyperdiverse and ubiquitous group with high potential for the application of HTS (Ji et al., 2013; Yu et al., 2012). Best practices and harmonization for metabarcoding of bulk arthropod samples are still in the early stages of development, but already there are multiple global initiatives that pivot on arthropod metabarcoding (e.g., the BIOSCAN initiative and its regional extensions such as the BioAlfa or the Kruger Malaise Program, the SITE-100 project, the Insect Biome Atlas Project, and LIFEPLAN; see Table 1). We view this critical mass, which is complemented by various national-level efforts, as a solid platform for the development of a terrestrial arthropod module, where passive sampling methods (e.g., pitfall, light and malaise traps, the last of which is already standardized within the BIOSCAN initiative) can be implemented individually, or in combination. Recent efforts toward developing best practices for COI community metabarcoding (e.g., Andújar et al., 2018; Elbrecht et al., 2019; Marquina et al., 2019) can be further developed for harmonized protocols for subsequent sample processing.

5 | CONCLUDING REMARKS

The value of large-scale site-based records for terrestrial invertebrates has been recently demonstrated for nematodes (van den Hoogen et al., 2019) and earthworms (Phillips et al., 2019), where synthetic analyses have yielded predictions of global diversity patterns within both groups. However, for other fractions of biodiversity, new initiatives are needed to provide such data, which presents both a challenge and at the same time an opportunity. We conclude that the establishment of an integrative HTS framework for measuring and understanding global patterns of biodiversity is now a realizable ambition. We further conclude that, by placing metabarcoding within the core of this framework, there is high potential for: (i) the establishment of harmonized protocols; (ii) fostering participation; (iii) capitalizing upon recent initiatives; and (iv) comparability and integration of data among network members. While work continues to address remaining challenges for some biodiversity metrics, such as abundance and biomass (Bista et al., 2018; Lamb et al., 2019), these challenges are balanced by unique opportunities from DNA-based biodiversity data. These include automated taxonomic assignment, spatial maps of beta diversity, and the ability to integrate phylogeny and genetic diversity with other environmental variables to understand not only how biodiversity is structured globally, but also the overarching processes that might explain the dynamics of community assembly (e.g., Theodoridis et al., 2020). Within such a framework, deeper genomic and temporal sampling can be developed within sites of interest, building upon the foundation of local metabarcode data, and the spatial context of broader regional and global sampling within the network.

AUTHOR CONTRIBUTIONS

All authors contributed to the ideas and discussion of this review. P.A. and B.C.E. coordinated the Symposium and led the writing with contributions from all authors.

DATA AVAILABILITY STATEMENT

No data were analysed in this meeting review.

ORCID

Paula Arribas https://orcid.org/0000-0002-0358-8271
Carmelo Andújar https://orcid.org/0000-0001-9759-7402
Jeremy R. deWaard https://orcid.org/0000-0001-9778-5454
Vasco Elbrecht https://orcid.org/0000-0003-4672-7099
Gentile F. Ficetola https://orcid.org/0000-0003-3414-5155
Marta Goberna https://orcid.org/0000-0001-5303-3429
Susan Kennedy https://orcid.org/0000-0002-1616-3985
Henrik Krehenwinkel https://orcid.org/0000-0001-5069-8601
Florian Leese https://orcid.org/0000-0002-5465-913X
Lucie Zinger https://orcid.org/0000-0002-3400-5825
Alfried P. Vogler https://orcid.org/0000-0002-2462-3718
Brent C. Emerson https://orcid.org/0000-0003-4067-9858

REFERENCES

Alberdi, A., Aizpurua, O., Bohmann, K., Gopalakrishnan, S., Lynggaard, C., Nielsen, M., & Gilbert, M. T. P. (2019). Promises and pitfalls of using high-throughput sequencing for diet analysis. *Molecular Ecology Resources*, 19, 327–348.

Alberti, A. (2017). To metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. *Scientific Data*, 4, 170093.

Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Zech Xu, Z., Kightley, E. P., Thompson, L. R., Hyde, E. R., Gonzalez, A., & Knight, R. (2017). Deblur rapidly resolves single-nucleotide community sequence patterns. *American Society for Microbiology*, 2, 1–7.

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, 3968–3975.

Andújar, C., Creedy, T. J., Arribas, P., Lopez, H., Salces-Castellano, A., Perez-Delgado, A., Vogler, A. P., & Emerson, B. C. (2020) Validated removal of nuclear pseudogenes and sequencing artefacts from mitochondrial metabarcode data. *bioRxiv*.

Arribas, P., Andújar, C., Hopkins, K., Shepherd, M., & Vogler, A. P. (2016). Metabarcoding and mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil. *Methods in Ecology and Evolution*, 7, 1071–1081.

Arribas, P., Andújar, C., Salces-Castellano, A., Emerson, B. C., & Vogler, A. P. (2020). The limited spatial scale of dispersal in soil arthropods revealed with whole-community haplotype-level metabarcoding. *Molecular Ecology*. https://doi.org/10.1111/mec.15591

Ashfaq, M., Akhtar, S., Rafi, M. A., Mansoor, S., & Hebert, P. D. N. (2017). Mapping global biodiversity connections with DNA barcodes: Lepidoptera of Pakistan. *PLoS One*, 12, 1–13.

Balaban, M., Sarmashghi, S., & Mirarab, S. (2018). APPLES: Fast distance-based phylogenetic placement. *bioRxiv*, 475566.

Bálint, M., Pfenninger, M., Grossart, H.-P., Taberlet, P., Vellend, M., Leibold, M. A., Englund, G., & Bowler, D. (2018). Environmental DNA time series in ecology. *Trends in Ecology and Evolution*, 33, 945–957.

Bardgett, R. D., & van der Putten, W. H. (2014). Belowground biodiversity and ecosystem functioning. *Nature*, 515, 505–511.

Beentjes, K. K., Speksnijder, A. G. C. L., Schilthuizen, M., Schaub, B. E. M., & van der Hoorn, B. B. (2018). The influence of macroinvertebrate abundance on the assessment of freshwater quality in The Netherlands. *Metabarcoding and Metagenomics*, 2, 1–8.

Bik, H. M., Porazinska, D. L., Creer, S., Caporaso, J. G., Knight, R., & Thomas, W. K. (2012). Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in Ecology and Evolution*, 27, 233–243.

Bista, I., Carvalho, G. R., Tang, M., Walsh, K., Zhou, X., Hajibabaei, M., Shokralla, S., Seymour, M., Bradley, D., Liu, S., Christmas, M., & Creer, S. (2018). Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. *Molecular Ecology Resources*, 18, 1020–1034.

Blanchet, F. G., Cazelles, K., & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. *Ecology Letters*, 23, 1050–1063.

Bohan, D. A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A. J., & Woodward, G. (2017). Next-generation global biomonitoring: Large-scale, automated reconstruction of ecological networks. *Trends in Ecology and Evolution*, 32, 477–487.

Bush, A., Compson, Z. G., Monk, W. A., Porter, T. M., Steeves, R., Emilson, E., Gagne, N., Hajibabaei, M., Roy, M., & Baird, D. J. (2019). Studying ecosystems with DNA metabarcoding: lessons from biomonitoring of aquatic macroinvertebrates. *Frontiers in Ecology and Evolution*, 7, 1–12.

Bush, A., Sollmann, R., Wilting, A., Bohmann, K., Cole, B., Balzter, H., Martius, C., Zlinszky, A., Calvignac-Spencer, S., Cobbold, C. A., Dawson, T. P., Emerson, B. C., Ferrier, S., Gilbert, M. T. P., Herold, M., Jones, L., Leendertz, F. H., Matthews, L., Millington, J. D. A., … Yu, D. W. (2017). Connecting Earth observation to high-throughput biodiversity data. *Nature Ecology and Evolution*, 1, 1–9.

Buttigieg, P. L., Janssen, F., Macklin, J., & Pitz, K. (2019). The Global Omics Observatory Network: Shaping standards for long-term molecular observation. *Biodiversity Information Science and Standards*, 3. https://doi.org/10.3897/biss.3.36712

Calderón-Sanou, I., Münkemüller, T., Boyer, F., Zinger, L., & Thuiller, W. (2019) From environmental DNA sequences to ecological conclusions: How strong is the influence of methodological choices? *Journal of Biogeography*, 47(1), 193–206. https://doi.org/10.1111/jbi.13681

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME Journal*, 11, 2639–2643.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13, 581–583.

Canonico, G., Buttigieg, P. L., Montes, E., Muller-Karger, F. E., Stepien, C., Wright, D., Benson, A., Helmuth, B., Costello, M., Sousa-Pinto, I., Saeedi, H., Newton, J., Appeltans, W., Bednaršek, N., Bodrossy, L., Best, B. D., Brandt, A., Goodwin, K. D., Iken, K., … Murton, B. (2019). Global observational needs and resources for marine biodiversity. *Frontiers in Marine Science*, 6, 1–20.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G., & Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME Journal*, 6, 1621–1624.

Chen, W., & Ficetola, G. F. (2020). Numerical methods for sedimentary-ancient-DNA-based study on past biodiversity and ecosystem functioning. *Environmental DNA*, 2, 115–129.

Cicconardi, F., Borges, P. A. V., Strasberg, D., Oromí, P., López, H., Pérez-Delgado, A. J., Casquet, J., Caujapé-Castells, J., Fernández-Palacios, J. M., Thébaud, C., & Emerson, B. C. (2017). MtDNA metagenomics reveals large-scale invasion of belowground arthropod communities by introduced species. *Molecular Ecology*, 26, 3104–3115.

Crampton-Platt, A., Timmermans, M. J. T. N., Gimmel, M. L., Kutty, S. N., Cockerill, T. D., Vun Khen, C., & Vogler, A. P. (2015). Soup to tree: The phylogeny of beetles inferred by mitochondrial metagenomics of a bornean rainforest sample. *Molecular Biology and Evolutioniology and Evolution*, 32, 2302–2316.

Creedy, T. J., Ng, W. S., & Vogler, A. P. (2019). Toward accurate species-level metabarcoding of arthropod communities from the tropical forest canopy. *Ecology and Evolution*, 9, 3105–3116.

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W. K., Potter, C., & Bik, H. M. (2016). The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution*, 7, 1008–1018.

Davies, N., & Field, D. (2012). Sequencing data: A genomic network to monitor Earth. *Nature*, 481, 145.

Davies, N., Field, D., Amaral-Zettler, L., Clark, M. S., Deck, J., Drummond, A., Faith, D. P., Geller, J., Gilbert, J., Glöckner, F. O., Hirsch, P. R., Leong, J.-A., Meyer, C., Obst, M., Planes, S., Scholin, C., Vogler, A. P., Gates, R. D., Toonen, R., … Zingone, A. (2014). The founding charter of the Genomic Observatories Network. *GigaScience*, 3, 1–6.

Davies, N., Meyer, C., Gilbert, J. A., Amaral-Zettler, L., Deck, J., Bicak, M., Rocca-Serra, P., Assunta-Sansone, S., Willis, K., & Field, D. (2012). A call for an international network of genomic observatories (GOs). *GigaScience*, 1, 1.

Deagle, B. E., Thomas, A. C., McInnes, J. C., Clarke, L. J., Vesterinen, E. J., Clare, E. L., Kartzinel, T. R., & Eveson, J. P. (2019). Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? *Molecular Ecology*, 28, 391–406.

Deck, J., Gaither, M. R., Ewing, R., Bird, C. E., Davies, N., Meyer, C., Riginos, C., Toonen, R. J., & Crandall, E. D. (2017). The Genomic Observatories Metadatabase (GeOMe): A new repository for field and sampling event metadata associated with genetic samples. *PLoS Biology*, 15, 1–7.

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. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26, 5872–5895.

Delgado-Baquerizo, M., Oliverio, A. M., Brewer, T. E., Benavent-González, A., Eldridge, D. J., Bardgett, R. D., Maestre, F. T., Singh, B. K., & Fierer, N. (2018). Bacteria found in soil. *Science*, 325, 320–325.

deWaard, J. R., Levesque-Beaudin, V., deWaard, S. L., Ivanova, N. V., McKeown, J. T. A., Miskie, R., Naik, S., Perez, K. H. J., Ratnasingham, S., Sobel, C. N., Sones, J. E., Steinke, C., Telfer, A. C., Young, A. D., Young, M. R., Zakharov, E. V., & Hebert, P. D. N. (2019). Expedited assessment of terrestrial arthropod diversity by coupling Malaise traps with DNA barcoding. *Genome*, 62, 85–95.

Edgar, R. (2016). UNOISE2: Improved error-correction for Illumina 16S and ITS Amplicon sequencing. *bioRxiv*, 081257.

Elbrecht, V., Braukmann, T. W., Ivanova, N. V., Prosser, S. W., Hajibabaei, M., Wright, M., Zakharov, E. V., Hebert, P. D., & Steinke, D. (2019). Validation of COI metabarcoding primers for terrestrial arthropods. *PeerJ*, 7, e7745.

Elbrecht, V., Leese, F., & Rockström, J. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—Sequence relationships with an innovative metabarcoding protocol. *PLoS One*, 10, e0130324.

Elbrecht, V., Vamos, E. E., Steinke, D., & Leese, F. (2018). Estimating intraspecific genetic diversity from community DNA metabarcoding data. *PeerJ*, 6, e4644.

Faith, D., Collen, B., Ariño, A., Patricia Koleff, P. K., Guinotte, J., Kerr, J., & Chavan, V. (2013). Bridging the biodiversity data gaps: Recommendations to meet users' data needs. *Biodiversity Informatics*, 8, 41–58.

Field, D., Amaral-Zettler, L., Cochrane, G., Cole, J. R., Dawyndt, P., Garrity, G. M., Gilbert, J., Glöckner, F. O., Hirschman, L., Karsch-Mizrachi, I., Klenk, H.-P., Knight, R., Kottmann, R., Kyrpides, N., Meyer, F., San Gil, I., Sansone, S.-A., Schriml, L. M., Sterk, P., … Wooley, J. (2011). The genomic standards consortium. *PLoS Biology*, 9, 8–10.

Field, D., Sansone, S.-A., Collis, A., Booth, T., Dukes, P., Gregurick, S. K., Kennedy, K., Kolar, P., Kolker, E., Maxon, M., Millard, S., Mugabushaka, A.-M., Perrin, N., Remacle, J. E., Remington, K., Rocca-Serra, P., Taylor, C. F., Thorley, M., Tiwari, B., & Wilbanks, J. (2009). Omics data sharing. *Science*, 326, 234–236.

Fierer, N., & Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 626–631.

Geisen, S. (2016). Thorough high-throughput sequencing analyses unravels huge diversities of soil parasitic protists. *Environmental Microbiology*, 18, 1669–1672.

Geisen, S., Briones, M. J. I., Gan, H., Behan-Pelletier, V. M., Friman, V.-P., de Groot, G. A., Hannula, S. E., Lindo, Z., Philippot, L., Tiunov, A. V.,

& Wall, D. H. (2019). A methodological framework to embrace soil biodiversity. *Soil Biology and Biochemistry*, *136*, 107536.

Gilbert, J. A., Jansson, J. K., & Knight, R. (2014). The earth microbiome project: Successes and aspirations. *BMC Biology*, *12*, 1–4.

Gilbert, J. A., Meyer, F., Antonopoulos, D., Balaji, P., Brown, C. T., Brown, C. T., Desai, N., Eisen, J. A., Evers, D., Field, D., Feng, W. U., Huson, D., Jansson, J., Knight, R., Knight, J., Kolker, E., Konstantinidis, K., Kostka, J., Kyrpides, N., … Stevens, R. (2010). Meeting report: The terabase metagenomics workshop and the vision of an earth microbiome project. *Standards in Genomic Sciences*, *3*, 243–248.

Goberna, M., Navarro-Cano, J. A., & Verdu, M. (2016). Opposing phylogenetic diversity gradients of plant and soil bacterial communities. *Proceedings of the Royal Society of London B*, *283*, 20153003.

Gorsky, G., Bourdin, G., Lombard, F., Pedrotti, M. L., Audrain, S., Bin, N., Boss, E., Bowler, C., Cassar, N., Caudan, L., Chabot, G., Cohen, N. R., Cron, D., De Vargas, C., Dolan, J. R., Douville, E., Elineau, A., Flores, J. M., Ghiglione, J. F., … Karsenti, E. (2019). Expanding Tara oceans protocols for underway, ecosystemic sampling of the ocean-atmosphere interface during Tara pacific expedition (2016–2018). *Frontiers in Marine Science*, *6*, 2016–2018.

Guralnick, R., Walls, R., & Jetz, W. (2018). Humboldt Core—Toward a standardized capture of biological inventories for biodiversity monitoring, modeling and assessment. *Ecography*, *41*, 713–725.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chemistry and Biology*, *5*, R245–R249.

Hebert, P. D. N. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, *270*, 313–321.

Hey, T., Tansley, S., & Tolle, K. (2009). *The fourth paradigm: Data-intensive scientific discovery.* : Microsoft Research.

Hinchliff, C. E., Smith, S. A., Allman, J. F., Burleigh, J. G., Chaudhary, R., Coghill, L. M., Crandall, K. A., Deng, J., Drew, B. T., Gazis, R., Gude, K., Hibbett, D. S., Katz, L. A., Laughinghouse, H. D., McTavish, E. J., Midford, P. E., Owen, C. L., Ree, R. H., Rees, J. A., … Cranston, K. A. (2015). Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 12764–12769.

Holt, B. G., Lessard, J.-P., Borregaard, M. K., Fritz, S. A., Araújo, M. B., Dimitrov, D., Fabre, P.-H., Graham, C. H., Graves, G. R., Jønsson, K. A., Nogués-Bravo, D., Wang, Z., Whittaker, R. J., Fjeldså, J., & Rahbek, C. (2013). An update of Wallace's zoogeographic regions of the world. *Science*, *17331*, 74–78.

Jarman, S. N., Berry, O., & Bunce, M. (2018). The value of environmental DNA biobanking for long-term biomonitoring. *Nature Ecology and Evolution*, *2*, 1192–1193.

Jetz, W., McGeoch, M. A., Guralnick, R., Ferrier, S., Beck, J., Costello, M. J., Fernandez, M., Geller, G. N., Keil, P., Merow, C., Meyer, C., Muller-Karger, F. E., Pereira, H. M., Regan, E. C., Schmeller, D. S., & Turak, E. (2019). Essential biodiversity variables for mapping and monitoring species populations. *Nature Ecology and Evolution*, *3*, 539–551.

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P. M., Woodcock, P., Edwards, F. A., Larsen, T. H., Hsu, W. W., Benedick, S., Hamer, K. C., Wilcove, D. S., Bruce, C., Wang, X., Levi, T., Lott, M., … Yu, D. W. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, *16*, 1245–1257.

Kahlert, M., Alfjorden, A., Apunte-Ramos, K., Bailet, B., Pérez Burillo, J., Carrera Gonzalez, A. G., Castro, D., Di Bernardi, C., Dully, V., Fekete, J., Frühe, L., González, R., Gratsia, E., Hanjalić, J., Kamberović, J., Kelly, A.-M., Meriggi, C., Nousiainen, I., Ørberg, S., … Weigand, A. M. (2019). New molecular methods to assess biodiversity. Potentials and pitfalls of DNA metabarcoding: a workshop report. *Research Ideas and Outcomes*, *5*. https://doi.org/10.3897/rio.5.e38915

Karsenti, E., Acinas, S. G., Bork, P., Bowler, C., De Vargas, C., Raes, J., Sullivan, M., Arendt, D., Benzoni, F., Claverie, J.-M., Follows, M., Gorsky, G., Hingamp, P., Iudicone, D., Jaillon, O., Kandels-Lewis, S., Krzic, U., Not, F., Ogata, H., … Wincker, P. (2011). A holistic approach to marine Eco-systems biology. *PLoS Biology*, *9*, 7–11.

Kembel, S. W., Wu, M., Eisen, J. A., & Green, J. L. (2012). Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Computational Biology*, *8*, 16–18.

Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F. S., Bahram, M., Bates, S. T., Bruns, T. D., Bengtsson-Palme, J., Callaghan, T. M., Douglas, B., Drenkhan, T., Eberhardt, U., Dueñas, M., Grebenc, T., Griffith, G. W., Hartmann, M., Kirk, P. M., Kohout, P., … Larsson, K.-H. (2013). Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology*, *22*, 5271–5277.

König, C., Weigelt, P., Schrader, J., Taylor, A., Kattge, J., & Kreft, H. (2019). Biodiversity data integration—The significance of data resolution and domain. *PLoS Biology*, *17*, 1–16.

Kopf, A. (2015). The ocean sampling day consortium. *GigaScience*, *4*, 27.

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*, 1–12.

Lamb, P. D., Hunter, E., Pinnegar, J. K., Creer, S., Davies, R. G., & Taylor, M. I. (2019). How quantitative is metabarcoding: A meta-analytical approach. *Molecular Ecology*, *28*, 420–430.

Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., Durbin, R., Edwards, S. V., Forest, F., Gilbert, M. T. P., Goldstein, M. M., Grigoriev, I. V., Hackett, K. J., Haussler, D., Jarvis, E. D., Johnson, W. E., Patrinos, A., Richards, S., Castilla-Rubio, J. C., … Zhang, G. (2018). Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, 4325–4333.

Lewitus, E., Bittner, L., Malviya, S., Bowler, C., & Morlon, H. (2018). Clade-specific diversification dynamics of marine diatoms since the Jurassic. *Nature Ecology and Evolution*, *2*, 1715–1723.

Makiola, A., Compson, Z. G., Baird, D. J., Barnes, M. A., Boerlijst, S. P., Bouchez, A., Brennan, G., Bush, A., Canard, E., Cordier, T., Creer, S., Curry, R. A., David, P., Dumbrell, A. J., Gravel, D., Hajibabaei, M., Hayden, B., van der Hoorn, B., Jarne, P., … Bohan, D. A. (2020). Key questions for next-generation biomonitoring. *Frontiers in Environmental Science*, *7*, 197.

Marotz, C., Amir, A., Humphrey, G., Gaffney, J., Gogul, G., & Knight, R. (2017). DNA extraction for streamlined metagenomics of diverse environmental samples. *BioTechniques*, *62*, 290–293.

Marquina, D., Andersson, A. F., & Ronquist, F. (2019). New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods. *Molecular Ecology Resources*, *19*, 90–104.

Marquina, D., Esparza-Salas, R., Roslin, T., & Ronquist, F. (2019) Establishing arthropod community composition using metabarcoding: Surprising inconsistencies between soil samples and preservative ethanol and homogenate from Malaise trap catches. *Molecular Ecology Resources*, *19*, 1516–1530.

Miraldo, A., Li, S., Borregaard, M. K., Flórez-Rodríguez, A., Gopalakrishnan, S., Rizvanovic, M., Wang, Z., Rahbek, C., Marske, K. A., & Nogués-Bravo, D. (2016). An antropocene map of genetic diversity. *Introduction to Conservation Genetics*, *353*, 41–65.

Muñoz, J. (2007). Biodiversity conservation including uncharismatic species. *Biodiversity and Conservation*, *16*, 2233–2235.

Murray, D. C., Pearson, S. G., Fullagar, R., Chase, B. M., Houston, J., Atchison, J., White, N. E., Bellgard, M. I., Clarke, E., Macphail, M., Gilbert, M. T. P., Haile, J., & Bunce, M. (2012). High-throughput sequencing of ancient plant and mammal DNA preserved in herbivore middens. *Quaternary Science Reviews*, *58*, 135–145.

Nielsen, M., Gilbert, M. T. P., Pape, T., & Bohmann, K. (2019). A simplified DNA extraction protocol for unsorted bulk arthropod samples that maintains exoskeletal integrity. *Environmental DNA*, *1*, 144–154.

Orgiazzi, A., Dunbar, M. B., Panagos, P., de Groot, G. A., & Lemanceau, P. (2015). Soil biodiversity and DNA barcodes: Opportunities and challenges. *Soil Biology and Biochemistry*, *80*, 244–250.

Pansu, J., Giguet-Covex, C., Ficetola, G. F., Gielly, L., Boyer, F., Zinger, L., Arnaud, F., Poulenard, J., Taberlet, P., & Choler, P. (2015). Reconstructing long-term human impacts on plant communities: An ecological approach based on lake sediment DNA. *Molecular Ecology*, *24*, 1485–1498.

Pedrós-Alió, C., & Manrubia, S. (2016). The vast unknown microbial biosphere. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 6585–6587. https://doi.org/10.1073/pnas.1606105113

Pérez-Valera, E., Verdú, M., Navarro-Cano, J. A., & Goberna, M. (2018). Resilience to fire of phylogenetic diversity across biological domains. *Molecular Ecology*, *27*, 2896–2908.

Phillips, H. R. P., Guerra, C. A., Bartz, M. L. C., Briones, M. J. I., Brown, G., Crowther, T. W., Ferlian, O., Gongalsky, K. B., van den Hoogen, J., Krebs, J., Orgiazzi, A., Routh, D., Schwarz, B., Bach, E. M., Bennett, J. M., Brose, U., Decaëns, T., König-Ries, B., Loreau, M., … Eisenhauer, N. (2019). Global distribution of earthworm diversity. *Science*, *366*, 480–485.

Pollock, J., Glendinning, L., Wisedchanwet, T., & Watson, M. (2018). The madness of microbiome: attempting to find consensus "best practice" for microbiome studies. *Applied and Environmental Microbiology*, *84*, 1–12.

Pope, L. C., Liggins, L., Keyse, J., Carvalho, S. B., & Riginos, C. (2015). Not the time or the place: The missing spatio-temporal link in publicly available genetic data. *Molecular Ecology*, *24*, 3802–3809.

Poretsky, R. S., Bano, N., Buchan, A., LeCleir, G., Kleikemper, J., Pickering, M., Pate, W. M., Moran, M. A., & Hollibaugh, J. T. (2005). Analysis of microbial gene transcripts in environmental samples. *Applied and Environmental Microbiology*, *71*, 4121–4126.

Porter, T. M., & Hajibabaei, M. (2018). Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. *Molecular Ecology*, *27*, 313–338.

Ranasinghe, J. A., Stein, E. D., Miller, P. E., & Weisberg, S. B. (2012). Performance of two southern California benthic community condition indices using species abundance and presence-only data: Relevance to DNA barcoding. *PLoS One*, *7*. https://doi.org/10.1371/journal.pone.0040875

Ransome, E., Geller, J. B., Timmers, M., Leray, M., Mahardini, A., Sembiring, A., Collins, A. G., & Meyer, C. P. (2017). The importance of standardization for biodiversity comparisons: A case study using autonomous reef monitoring structures (ARMS) and metabarcoding to measure cryptic diversity on Mo'orea coral reefs, French Polynesia. *PLoS One*, *12*, e0175066.

Ratnasingham, S., & Hebert, P. D. N. (2007). The barcode of life data system. *Molecular Ecology Notes*, *7*, 355–364.

Riginos, C., Crandall, E. D., Liggins, L., Gaither, M. R., Ewing, R. B., Meyer, C., Andrews, K. R., Euclide, P. T., Titus, B. M., Therkildsen, N. O., Salces-Castellano, A., Stewart, L. C., Toonen, R. J., & Deck, J. (2020). Building a global genomics observatory: Using GEOME (the Genomic Observatories Metadatabase) to expedite and improve deposition and retrieval of genetic data and metadata for biodiversity research. *Molecular Ecology Resources*, *20*, 1458–1469.

Rominger, A. J., Goodman, K. R., Lim, J. Y., Armstrong, E. E., Becking, L. E., Bennett, G. M., Brewer, M. S., Cotoras, D. D., Ewing, C. P., Harte, J., Martinez, N. D., O'Grady, P. M., Percy, D. M., Price, D. K., Roderick, G. K., Shaw, K. I., Valdovinos, F. S., Gruner, D. S., Gillespie, R. G., & Ricklefs, R. (2016). Community assembly on isolated islands: macroecology meets evolution. *Global Ecology and Biogeography*, *25*, 769–780.

Sansone, S.-A., Rocca-Serra, P., Field, D., Maguire, E., Taylor, C., Hofmann, O., Fang, H., Neumann, S., Tong, W., Amaral-Zettler, L., Begley, K., Booth, T., Bouguelleret, L., Burns, G., Chapman, B., Clark, T., Coleman, L.-A., Copeland, J., Das, S., … Hide, W. (2012). Toward interoperable bioscience data. *Nature Genetics*, *44*, 121–126.

Schenk, J., Geisen, S., Kleinbölting, N., & Traunspurger, W. (2019). Metabarcoding data allow for reliable biomass estimates in the most abundant animals on earth. *Metabarcoding and Metagenomics*, *3*, 117–126.

Schnell, I. B., Thomsen, P. F., Wilkinson, N., Rasmussen, M., Jensen, L. R., Willerslev, E., Bertelsen, M. F., & Gilbert, M. T. P. (2012). Screening mammal biodiversity using DNA from leeches. *Current Biology*, *22*, R262–R263.

Shoemaker, W. R., Locey, K. J., & Lennon, J. T. (2017). A macroecological theory of microbial biodiversity. *Nature Ecology and Evolution*, *1*, 0107.

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA for biodiversity research and monitoring.* Oxford University Press.

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology*, *21*, 1789–1793.

Tang, M., Hardman, C. J., Ji, Y., Meng, G., Liu, S., Tan, M., Yang, S., Moss, E. D., Wang, J., Yang, C., Bruce, C., Nevard, T., Potts, S. G., Zhou, X., & Yu, D. W. (2015). High-throughput monitoring of wild bee diversity and abundance via mitogenomics. *Methods in Ecology and Evolution*, *6*, 1034–1043.

Tedersoo, L. (2014). Disentangling global soil fungal diversity. *Science*, *346*, 1052–1053.

Theodoridis, S., Fordham, D. A., Brown, S. C., Li, S., Rahbek, C., & Nogues-Bravo, D. (2020). Evolutionary history and past climate change shape the distribution of genetic diversity in terrestrial mammals. *Nature Communications*, *11*, 1–11.

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., Prill, R. J., Tripathi, A., Gibbons, S. M., Ackermann, G., Navas-Molina, J. A., Janssen, S., Kopylova, E., Vázquez-Baeza, Y., González, A., Morton, J. T., Mirarab, S., Zech Xu, Z., Jiang, L., … Knight, R. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, *551*, 457–463.

Tsuji, S., Miya, M., Ushio, M., Sato, H., Minamoto, T., & Yamanaka, H. (2020). Evaluating intraspecific genetic diversity using environmental DNA and denoising approach: A case study using tank water. *Environmental DNA*, *2*, 42–52.

Turon, X., Antich, A., Palacín, C., Præbel, K., & Wangensteen, O. S. (2020). From metabarcoding to metaphylogeography: separating the wheat from the chaff. *Ecological Applications*, *30*, 1–19.

van den Hoogen, J., Geisen, S., Routh, D., Ferris, H., Traunspurger, W., Wardle, D. A., de Goede, R. G. M., Adams, B. J., Ahmad, W., Andriuzzi, W. S., Bardgett, R. D., Bonkowski, M., Campos-Herrera, R., Cares, J. E., Caruso, T., de Brito Caixeta, L., Chen, X., Costa, S. R., Creamer, R., … Crowther, T. W. (2019). Soil nematode abundance and functional group composition at a global scale. *Nature*, *572*, 194–198.

Walters, M., & Scholes, R. J. (2017). *The GEO handbook on Biodiversity Observation Networks*. : Springer Open.

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., & Vieglais, D. (2012). Darwin core: An evolving community-developed biodiversity data standard. *PLoS One*, *7*. https://doi.org/10.1371/journal.pone.0029715

Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J. R., Amaral-Zettler, L., Gilbert, J. A., Karsch-Mizrachi, I., Johnston, A., Cochrane, G., Vaughan, R., Hunter, C., Park, J., Morrison, N., Rocca-Serra, P., Sterk, P., Arumugam, M., Bailey, M., Baumgartner, L., … Glöckner, F. O. (2011). Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nature Biotechnology*, *29*, 415–420.

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*, *3*, 613–623.

Zinger, L., Bonin, A., Alsos, I. G., Bálint, M., Bik, H., Boyer, F., Chariton, A. A., Creer, S., Coissac, E., Deagle, B. E., De Barba, M., Dickie, I. A., Dumbrell, A. J., Ficetola, G. F., Fierer, N., Fumagalli, L., Gilbert, M. T. P., Jarman, S., Jumpponen, A., … Taberlet, P. (2019). DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology*, *28*, 1857–1862.

Zinger, L., Donald, J., Brosse, S., Gonzalez, M. A., Iribar, A., Leroy, C., Murienne, J., Orivel, J., Schimann, H., Taberlet, P., & Lopes, C. M. (2020). Advances and prospects of environmental DNA in Neotropical rainforests. *Advances in Ecological Research*, *62*, 331–373.