A vision for global eDNA-based monitoring in a changing world

Philip Francis Thomsen[1,*], Mads Reinholdt Jensen[1,2] and Eva Egelyng Sigsgaard[1]

[1]Department of Biology, Aarhus University, Ny Munkegade 116, DK-8000 Aarhus, Denmark
[2]Norwegian College of Fishery Science, UiT – The Arctic University of Norway, Tromsø, Norway
[*]Correspondence: pfthomsen@bio.au.dk

Abstract
Environmental DNA (eDNA) has opened promising avenues for establishing standardized, cost-efficient monitoring of biodiversity. However, comprehensive and systematic implementation is urgently needed to address the current biodiversity crisis. We here envision a global eDNA biomonitoring scheme, which could potentially revolutionize the understanding and conservation of life on Earth.

Introduction
In the last two decades, we have seen the dawn of a new approach in biology for obtaining information on species and communities. We have learned that our planet is brimming with microscopic traces of its past and present inhabitants in the form of DNA. As life on Earth unfolds, DNA is left behind in the environment by organisms spanning from bacteria to blue whales, and such environmental DNA (eDNA) fills the soils, lakes, rivers, oceans and even the air[1,2]. The specific environment in which the DNA is deposited determines how quickly it is degraded by chemical and biological processes, and thereby for how long we can detect it. This time frame ranges from days or weeks in aquatic systems to at least two million years in deep sediments, based on current estimates[1,3]. Due to a recent revolution in high-throughput sequencing technology, eDNA can be studied very thoroughly, targeting entire communities at once. Although the application of eDNA analyses encompasses exciting opportunities in paleoecology (ancient eDNA)[3], the scope for contemporary biodiversity monitoring and conservation is of urgent relevance and will therefore be the focus of this commentary. The Earth's biodiversity is in decline, currently surpassing pre-human extinction rates by an estimated 100-1,000 times[4]. Species that roamed the Earth long before us are leaving it at an untimely rate. And with each of them leaves not only a source of human inspiration and wonder, but also important biological functions, as well as potential scientific breakthroughs and discoveries.

To this date, more than a thousand scientific studies on eDNA from contemporary eukaryotic biodiversity have been conducted. So, what have we learned? From a pessimistic angle, one could claim: not a lot. For macro-organisms, we have largely confirmed well-established species distributions and community patterns. Of course, such confirmation studies have been necessary to optimize the eDNA approach and validate it against established monitoring methods. But while they have led to discoveries on the state, fate, and transport of eDNA in various settings, we have hardly discovered anything substantial or surprising about biodiversity – with the recent discovery of Micrognathozoa being an example of a rare exception[5]. With this perspective in mind, it is important to ask: what's next?

Environmental DNA metabarcoding refers to amplicon sequencing of communities within a specific taxonomic group and is currently the preferred approach for eDNA analyses[1]. But with advancing DNA sequencing technology, one can only imagine the depth of analyses possible within the next decades. This might well include completely novel molecular

approaches, coupled with artificial intelligence for data analyses. However, at least for the foreseeable future, it will most likely also involve further development of long-read sequencing technologies and shotgun sequencing, which avoid the PCR-related biases of metabarcoding[3,6]. Combined with improved reference databases, it might thus become possible to achieve an almost complete taxonomic characterization of all genomic material present in an eDNA sample across the tree of life[6]. We know from eDNA metabarcoding that with a reasonable amount of sampling, a good estimate of the geographical distributions of targeted taxa (e.g. marine fishes) can be obtained even at large spatial scales[7]. Thus, assuming representative sampling of relevant substrates, it is in theory "only" a matter of ultradeep, taxonomically unbiased sequencing before such information can be obtained for all taxa in a sampling area. Such comprehensive data could potentially yield i) high-quality information on presence/absence of species (richness), ii) genome-wide data on individuals represented in the sample (population genetics), iii) more accurate abundance data compared to e.g. eDNA metabarcoding (population sizes) and even iv) characterization of the collective functions of the community. With shotgun sequencing, relative read counts are less taxonomically biased, and with sufficiently deep sequencing, information on genetic variation within a species will be genome-wide, allowing for e.g. better estimates of the number of individuals represented in a sample[8]. A co-evolution in computer and data storage technology would be necessary for metagenomic analyses of eDNA samples to be feasible on a large scale. However, just like sequencing technologies, these fields are advancing fast.

A vision

While novel ways of analyzing eDNA will surely improve eDNA-based monitoring, we believe that realizing the method's full potential within conservation will also require improving its implementation. Environmental DNA biomonitoring is standardizable, scalable, and cost-efficient compared to most traditional monitoring methods, and these advantages remain to be fully leveraged. Building on the concept of Genomic Observatories[9], we here envision an ambitious, systematic implementation of the now scientifically well-established eDNA approach to inform management and conservation initiatives. Namely, a comprehensive eDNA monitoring scheme, that can capture and preserve genomic information across taxonomic groups and ecosystems on a global-scale and in a highly standardized manner (Fig. 1). This would not only be immensely valuable for conservation efforts and basic biological research, but also for extensive documentation of Earth's biological heritage. We envision a reappraisal of the historic focus by our predecessors on describing the natural wonders of the Earth - this time not only out of the curiosity of humankind, but also because biodiversity conservation depends upon it.

Developing a long-term and large-scale eDNA monitoring scheme could mitigate two fundamental and closely linked inadequacies in current biomonitoring. Firstly, the global lack of taxonomic expertise, which represents an odd contrast to the proportion of contemporary species on Earth still undescribed (perhaps as high as 80%) and the urgency of the biodiversity crisis. And secondly, the universal lack of long-term, standardized biomonitoring data, which are crucial for data-driven conservation efforts and for combating the societal "shifting baseline syndrome" (current generations' normalized view of a depleted biosphere compared to natural levels). Even if we succeed in increasing taxonomic expertise in our societies, it will not be sufficient to cover the massive diversity of life awaiting discovery and scientific description. However, the genomic data generated in a global eDNA biomonitoring programme will unveil many undescribed life forms and might in fact be the only feasible way to start closing the knowledge gap on Earth's biodiversity. A catalogue of environmental genomic diversity across the globe cannot replace the work of traditional taxonomists.

Descriptions of physical specimens still constitute the backbone of taxonomy, but they require a much longer time frame than DNA-based descriptions, which are therefore increasingly supplementing traditional taxonomy (e.g. the species hypotheses of the UNITE database, www.unite.ut.ee).

The generated data would also provide more accurate information on the biological diversity on Earth, including the locations of biodiversity hotspots, and better estimates of temporal trends in the diversity and abundance of understudied taxonomic groups. Similarly, they would aid in disentangling which environmental parameters that best explain biodiversity patterns. Existing ambitious projects like TARA Oceans have already generated massive amounts of genomic data on marine plankton[10], and can serve as inspiration for studies on other taxa and ecosystems. Genome-centric approaches like the one used in TARA Oceans for creating metagenome-assembled genomes (MAGs) from environmental samples and single-cell genomes (SAGs) are likely our best candidates for filling in the major taxonomic gaps and for combating the underrepresentation of most bacteria, viruses and microeukaryotes in genomic reference databases. At least for prokaryotes, the use of MAGs allows eDNA samples to serve both as genetic references and as observational data simultaneously. The long-standing research in environmental microbiology might additionally serve as valuable inspiration for the standardization of protocols (e.g. the Earth Microbiome Project, https://earthmicrobiome.org/) and for developing online tools for analyses (e.g. MGnify, https://www.ebi.ac.uk/metagenomics).

Importantly, eDNA analyses can also provide information on genetic diversity[8]. Very little is known about global patterns of genetic diversity in wild populations, although genetic diversity is one of the three levels of biological diversity recognized by the Convention on Biological Diversity (www.cbd.int), the other two levels being species and ecosystems. There are however challenges associated with obtaining accurate measures of genetic diversity from eDNA, and in the coming years it will be critical to establish robust denoising workflows that can tease real variation apart from sequencing artefacts. Massive eDNA monitoring data can also help direct future research towards potentially unknown biological functions by providing genomic data on a range of understudied organisms[10]. A high coverage metagenomic dataset potentially provides the possibility to perform analyses of ecological function at the community level, e.g. by identifying genes associated with known metabolic, immunological, or reproductive functions and comparing these across environmental and spatial gradients. Lastly, trophic interactions and food web properties, which are often very difficult to obtain using traditional data, can potentially be obtained more efficiently and at a larger spatial scale using eDNA in connection with databases of species traits and known trophic interactions[11].

Realizing the vision

A global eDNA monitoring scheme could be structured as local sampling, regional sequencing and sample storage, and global analyses and sharing of data in open access repositories. The data could be used for diverse purposes at all structural levels, including local implementation of conservation efforts (Fig. 1).

Sampling

The sampling part of the monitoring scheme should be based on automated samplers, but could be supplemented by large-scale citizen science initiatives (Fig. 1). Although some further development and optimization is needed, automated eDNA samplers allow for highly standardized sampling, a reduced number of person hours required per sample, and

deployment in remote localities[2,12]. Similarly, the evolution of drone technology allows for sampling in otherwise inaccessible areas, and while their use likely entails greater ethical and legal restraints and more person hours, their lower cost and greater portability provides potential for a more representative sampling of the globe. Citizen science-based sampling is restricted to more accessible areas, but is less dependent on technology and specialized equipment, and has the added benefit of spreading knowledge and awareness of biodiversity in local communities. Importantly, eDNA coupled with citizen science has already demonstrated potential for documenting species distribution patterns of fishes on a national scale[7]. As many eDNA studies rely on sparse scientific personnel, sufficient sampling over an environmental gradient may take weeks or months. Automated samplers, drones and large groups of citizen scientists on the other hand can sample a local, regional, or potentially global area at nearly identical time points, thereby eliminating temporal bias. This type of standardized sampling allows for highly authentic community comparisons across space and time, which are necessary for distinguishing signatures of global change from natural variation in community compositions. Standardizing the timing of eDNA sampling is more important than one might think, since even diel variation can be captured by eDNA samples[13]. Importantly, an expanding suite of other successful automated monitoring methods, such as image and acoustic recorders, could supplement the biodiversity data derived from eDNA, while information from LiDAR (Light Detection and Ranging) and satellites could provide associated environmental and biotic data[14].

Storage and analyses

Long-term storage of both samples and associated metadata would be crucial, and regional facilities for standardized sample storage and sequencing would make this feasible, allowing for the same samples to be robustly re-analysed when new questions or technologies arise. We also imagine that large regional facilities will be necessary to allow consistent high-throughput sequencing of the samples collected locally. However, local sequencing may also become a possibility, and although fully automated fieldable eDNA sequencing is still in its infancy, it is certainly achievable[14]. Analyses of the sequencing data may be run locally, regionally, or globally depending on the scope, using standardized pipelines that are becoming more user-friendly and accessible (e.g. eDNA Explorer, www.ednaexplorer.org). Standardisation and quality assurance of sequencing data, metadata and bioinformatic pipelines is vital for documentation and reproducibility, and for the reliability of conclusions drawn from the data. While it is not yet clear what this should entail, widely used guidelines exist[14,15], and worldwide experience is already being gained through large-scale eDNA projects such as CALeDNA, which analyses California's biodiversity using eDNA samples collected by citizens (www.ucedna.com). We envision that a large public database such as the Global Biodiversity Information Facility (www.gbif.org) could facilitate long-term storage of the eDNA data, but the exact resource and standards to be followed for such storage should be discussed further.

Making use of the data

Turning results derived from eDNA monitoring into conservation action is obviously not straightforward. It will take years before sufficient data have been gathered to allow for robust analyses of temporal biodiversity trends. However, in the short term, detection of endangered species of high conservation concern as well as of invasive species found outside their known distribution ranges could be used directly in conservation planning. In the long term, detected changes in species abundances or distribution patterns can be coupled with data on known environmental stressors to assess and model their impacts on biodiversity.

Challenges are plenty, but the timing is right

Environmental change will continue to drive massive reorganization of the Earth's biota. In fact, we have probably only seen "the tip of the iceberg" when it comes to species extinctions and changes in distribution patterns, due to an expected extinction debt (the future extinction of species due to past events) and an increasing intensity of global change[4]. And with current, highly insufficient levels of biomonitoring, changes in species distributions and population sizes as well as many species extinctions will probably for the most part go undetected. With limited resources for taxonomy and biodiversity monitoring, we need to look for alternative approaches to properly address the biodiversity crisis.

We are not alone in speculating that global eDNA-based monitoring programmes are achievable in the foreseeable future. Indeed, ambitious endeavours like the one envisioned here have been proposed by others. For example, a massive global DNA barcoding initiative has been established (https://ibol.org/bioscan/). Also, a recent initiative for global eDNA monitoring of vertebrates in freshwater habitats (www.ebioatlas.org) involves a partnership between the International Union for the Conservation of Nature (IUCN) as an intergovernmental body, and a private company. Furthermore, non-DNA-based large-scale citizen science projects have achieved impressive results during the last two decades. For instance, the project eBird by Cornell Lab of Ornithology (www.ebird.org) has gathered global-scale information on the distribution of bird species, which has yielded many new insights into e.g. bird migration patterns. Similarly, the iNaturalist image recognition algorithms have piqued the interest of millions of citizens in documenting biodiversity globally (www.inaturalist.org). A major challenge for realizing the vision proposed here is obviously the high, continuous operating costs and the need for a global consortium to facilitate the standardized data collection. We can only encourage funding bodies, wealthy visionaries and/or worldwide governmental institutions to find a common ground in agreeing that to document and protect life on Earth, funding must be allocated towards continuous biomonitoring of our planet. Importantly, we do not encourage that eDNA should replace existing or other emerging monitoring methods, nor should it consume resources already allocated to well-functioning biomonitoring and conservation initiatives.

Obviously, there are also ethical challenges associated with establishing a successful global eDNA initiative. An essential task for the global sampling consortium would for example be to ensure adherence to Nagoya protocols (https://www.cbd.int/abs) and to facilitate the sharing of samples and results across borders. Also, several technical caveats of the eDNA approach exist, such as contamination, PCR and extraction biases and difficulties in linking biomass or population sizes to eDNA data[1,13]. However, correct identification of eDNA sequences remains perhaps the most important technical challenge for large-scale implementation. Although rapidly expanding, reference databases, such as the Barcode of Life Data System (BOLD), National Center for Biotechnology Information (NCBI) and Genome Taxonomy Database (GTDB), contain genetic data representing only a fraction of the expected millions of species on Earth and exhibit large gaps in nuclear regions. Combined with challenges like sequencing errors, synonymy, and specimen misidentifications, this can result in low taxonomic resolution or wrongly assigned eDNA sequences. While ecological inferences and biomonitoring are possible even if sequences are not identified to lower-level taxonomy, the full potential of an eDNA sample can only be reached with a high taxonomic resolution. Thus, an initiative such as the current one will have to be justified partly on the mere establishment of standardized sampling, documentation, and storage of information on global biological diversity. The eDNA samples and data will serve an important function as a "molecular museum" of life forms that might disappear before we even discover them. But

over time, it will provide increasingly valuable data for the conservation efforts needed to restore our planet's biodiversity to sustainable levels. Ultimately, a future global eDNA database will provide fascinating stories about the living world and perhaps lead to new scientific breakthroughs – just like classic natural history museums have done for centuries.

Figure 1. Outline of a global eDNA biomonitoring vision. The sampling part is here based on automated samplers and supplemented by large-scale citizen science initiatives. Sample types are here limited to seawater (blue), freshwater (red) and air (yellow), which are likely the easiest to standardize, but they could be supplemented by other substrate types. Long-term, standardized storage and sequencing of eDNA samples is enabled by large regional facilities, allowing for the same samples to be re-analysed when new questions or technologies arise. Analyses of data are performed in a global network using open access bioinformatics. Results are shared with the scientific community for biodiversity research, as well as with local communities for direct conservation applications. Graphic design by vahle+nikolaisen.

Acknowledgments
P.F.T is grateful for support from the Carlsberg Foundation (grant CF18-0949), the Velux Foundations (grant 21517), and the Independent Research Fund Denmark (grant 1054-00033B), that all helped to inspire this work.

Declaration of interests
The authors have no interests to declare.

References
1. Thomsen, P.F. and Willerslev, E. (2015). Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. Biol. Conserv. 183, 4–18. https://doi.org/10.1016/j.biocon.2014.11.019.

2. Lynggaard, C., Bertelsen, M.F., Jensen, C.V., Johnson, M.S., Frøslev, T.G., Olsen, M.T., and Bohmann, K. (2022). Airborne environmental DNA for terrestrial vertebrate community monitoring. Curr. Biol. 32, 701–707. https://doi.org/10.1016/j.cub.2021.12.014.

3. Kjær, K.H., Pedersen, M.W., De Sanctis, B., De Cahsan, B., Korneliussen, T.S., Michelsen, C.S., Sand, K.K., Jelavić, S., Ruter, A.H., Schmidt, A.M.A., et al. (2022). A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. Nature 612, 283–291. https://doi.org/10.1038/s41586-022-05453-y.

4. Cowie, R.H., Bouchet, P., and Fontaine, B. (2022). The Sixth Mass Extinction: fact, fiction or speculation? Biol. Rev. 97, 640–663. https://doi.org/10.1111/brv.12816.

5. Giribet, G., Wangensteen, O.S., Garcés-Pastor, S., Møller, P.R., and Worsaae, K. (2023). Using eDNA to find Micrognathozoa. Curr. Biol. 33, R743–R759. https://doi.org/10.1016/j.cub.2023.04.075.

6. Manu, S., and Umapathy, G. (2023). Deep sequencing of extracellular eDNA enables total biodiversity assessment of ecosystems. Ecol. Indic. 156, 111171. https://doi.org/10.1016/j.ecolind.2023.111171.

7. Agersnap, S., Sigsgaard, E.E., Jensen, M.R., Avila, M.D.P., Carl, H., Møller, P.R., Krøs, S.L., Knudsen, S.W., Wisz, M.S., and Thomsen, P.F. (2022). A national scale "BioBlitz" using

citizen science and eDNA metabarcoding for monitoring coastal marine fish. Front. Mar. Sci. 9, 824100. https://doi.org/10.3389/fmars.2022.824100.

8. Andres, K.J., Lodge, D.M., Sethi, S.A., and Andrés, J. (2023). Detecting and analysing intraspecific genetic variation with eDNA: From population genetics to species abundance. Mol. Ecol. 32, 4118–4132. https://doi.org/10.1111/mec.17031.

9. Davies, N., Field, D., Amaral-Zettler, L., Clark, M.S., Deck, J., Drummond, A., Faith, D.P., Geller, J., Gilbert, J., Glöckner, F.O., et al. (2014). The founding charter of the Genomic Observatories Network. GigaSci 3, 2. https://doi.org/10.1186/2047-217X-3-2.

10. Delmont, T.O., Gaia, M., Hinsinger, D.D., Frémont, P., Vanni, C., Fernandez-Guerra, A., Eren, A.M., Kourlaiev, A., d'Agata, L., Clayssen, Q., et al. (2022). Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. Cell Genom. 2, 100123. https://doi.org/10.1016/j.xgen.2022.100123.

11. Pereira, C.L., Ersoy, Z., Gilbert, M.T.P., Gravel, D., Araújo, M. B., and Matias, M.G. (2023). Future-proofing environmental DNA and trait-based predictions of food webs. BioScience, biad089. https://doi.org/10.1093/biosci/biad089.

12. Hansen, B.K., Jacobsen, M.W., Middelboe, A.L., Preston, C.M., Marin III, R., Bekkevold, D., Knudsen, S.W., Møller P.R., and Nielsen, E.E. (2020). Remote, autonomous real-time monitoring of environmental DNA from commercial fish. Sci. Rep. 10, 13272. https://doi.org/10.1038/s41598-020-70206-8.

13. Jensen, M.R., Sigsgaard, E.E., Ávila, M.D.P., Agersnap, S., Brenner-Larsen, W., Sengupta, M.E., Xing, Y., Krag, M.A., Knudsen, S.W., Carl, H., et al. (2022). Short-term temporal variation of coastal marine eDNA. Environ. DNA 4, 747–762. https://doi.org/10.1002/edn3.285.

14. Besson, M., Alison, J., Bjerge, K., Gorochowski, T.E., Høye, T.T., Jucker, T., Mann, H.M.R., and Clements, C.F. (2022). Towards the fully automated monitoring of ecological communities. Ecol. Lett. 25, 2753–2775. https://doi.org/10.1111/ele.14123.

15. Wilkinson, M., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J-W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018. https://doi.org/10.1038/sdata.2016.18.