Decoding dissolved information: environmental DNA sequencing at global scale to monitor a changing ocean

Luke R. Thompson[a,b] and Peter Thielen[c]

[a]Northern Gulf Institute, Mississippi State University, 2 Research Blvd, Starkville, Mississippi, 39759, USA

[b]Ocean Chemistry and Ecosystems Division, Atlantic Oceanographic and Meteorological Laboratory, National Oceanic and Atmospheric Administration, 4301 Rickenbacker Cswy, Miami, Florida, 33149, USA

[c]Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, Maryland, 20723-6099, USA

Email: luke.thompson@noaa.gov (correspondence), peter.thielen@jhuapl.edu

Abstract

The use of environmental DNA (eDNA) technology for environmental monitoring is rapidly expanding, with applications for fisheries, coral reefs, harmful algal blooms, invasive and endangered species, and biodiversity monitoring. By enabling detection of species over space and time, eDNA fulfills a fundamental need of environmental surveys. Traditional surveys are expensive, require significant capital expenditure, and can be destructive; eDNA offers promise for cheaper, less invasive, and higher-resolution (i.e. genetic) assessments of environments and stocks. However, challenges in quantification, detection limits, biobanking capacity, reference databases, and data management and integration remain significant hurdles to efficient eDNA monitoring at global and decadal scale. Here, we consider the current state of eDNA technology and its suitability for the problems for which it is being used. We explore the current best practices, the logistical and social challenges that prevent eDNA from widespread adoption and benefit, and the emerging technologies that may address those challenges.

Abstract Figure Legend: Processing steps in eDNA analysis from water collection to DNA sequence analysis.

Introduction

Environmental DNA (eDNA) is the DNA of whole or partial organisms, organismal traces (e.g. skin, mucus, and feces), or microbes in an environmental sample. Broadly defined, marine eDNA includes DNA from marine microbiomes, plankton, and traces of marine animals (the latter being an alternate, more narrow definition of eDNA). The power of eDNA comes from the ability to detect both microorganisms (often as whole cells) and macroorganisms (often as fragments of organisms or free DNA) from the nucleic acids they contain or leave behind. Analysis of eDNA enables detection of essentially any organism via trace DNA evidence, assessment of the relative or absolute abundance of particular groups, and precise taxonomic assignment using DNA sequences.

The techniques used in eDNA monitoring (Fig. 1) represent a powerful toolkit for marine biomonitoring and biosurveillance and include shotgun metagenomics (sequencing of all DNA in short fragments), metabarcoding (sequencing a selected gene fragment from a group of organisms such as all fish or bacteria), and quantitative PCR (quantifying copies of a selected gene fragment using PCR and fluorescence). For microorganisms having small-to-moderately sized genomes (less than ~10 Mbp), assembly of metagenomic reads can yield whole genomes that reveal functional potential [1–4]. For animals as well as microorganisms, metabarcoding can reveal the presence (or absence), and in some cases relative abundance, of species in a water sample [5,6], and qPCR is able in some cases to quantify the number of individuals present or recently present in a given volume of water [7,8].

The applications of eDNA technology in marine research and management are myriad and are summarized only briefly here, with a focus on those that may benefit from technological innovations. Marine microbiology was one of the earliest applications of eDNA metabarcoding (amplicon sequencing), for example, the use of the small subunit ribosomal RNA gene since the early 1990s to profile marine communities and identify novel bacterial groups [9]. Shotgun sequencing of marine eDNA has allowed genome assembly of uncultured groups [2,10] and a description of global ocean diversity patterns [11]. Metabarcoding has expanded to include multiple loci, targeting all major extant lineages [12–14] and leveraging high-throughput sequencing to generate millions of reads per community [15]. Analyzing multiple major taxonomic groups is possible from a single eDNA sample, which has great value for environmental and biodiversity monitoring [16,17], but one of the challenges is how to analyze multiple datasets together. Metabarcoding of multiple loci gives separate relative abundance profiles of each group that must be reconciled, whereas shotgun metagenomics gives information about the community as a whole but can miss low-abundance groups and presents challenges in taxonomic classification.

For marine monitoring, another obstacle is how to adequately sample the vast ocean and do so regularly enough to monitor global change [18]. Collective sampling and analysis by multiple teams provides a solution (Box 1) but creates new challenges in harmonization of methods across many groups' sampling and in data sharing and management. Detection of animals,

such as invasive or endangered species, marine mammals, and fish populations [5,19], is a powerful use of eDNA technology but brings additional difficulties in detection limits (low DNA concentrations), fragmentation of DNA (harder to amplify, requiring shorter amplicons with less resolution), and DNA half-life considerations (how long ago was the organism present) [20]. Coral reef monitoring [21–23] and other localized monitoring require consideration of water currents moving DNA away from where it was released and linking of DNA detection to ecosystem status.

The legacy of eDNA sequence data is one of its biggest advantages: DNA sequences are information-rich and represent a genetic snapshot in time that has enduring value beyond a single study. Proper management of the sequence data and associated metadata enables value to be extracted after the original survey is completed [4,24,25]. Emphasizing the value of this approach, in genomics research of higher organisms, a new field called 'macrogenetics' has sprung up that leverages thousands of publicly accessible genomes to infer patterns of genetic variation [26]. qPCR data, as PCR cycle numbers rather than DNA sequences, do not endure in the same way as the other technologies, but they can provide critical quantitative information to environmental surveys, adding to the qualitative information of sequence data [7,27,28].

Several analytical and data interoperability challenges prevent eDNA from being adopted for biological characterization at a global scale, and they occur at each stage of the analysis workflow (Table 1, Figure 2). These include low biomass in the environment, especially in oligotrophic waters; limited sample collection, storage, and preservation capacity; scalability to analyze large numbers of samples in the laboratory and with current analytical tools; costs associated with sample processing and analyst time; and limited availability of genetic reference data for the diverse species present in marine environments. New approaches addressing these issues continue to emerge rapidly and are increasingly enabling new capabilities including autonomous sampling, efficient sample processing, and improved data analysis and metadata integration (Box 2).

Sample collection and biobanking

Collection of marine eDNA via filtration or precipitation is a simple process, but the way it is done has important implications for its analysis and enduring value. Several reviews have covered best practices for sampling and collection [29–31]. Here, we focus on critical or emerging aspects of eDNA collection, including the biobanking of collected material, the concurrent collection of environmental data and sampling metadata, and use of autonomous systems.

DNA is nature's digital information storage medium. It is copied in high fidelity as organisms reproduce, enabling inference of evolutionary history from small genetic changes accumulated and inherited over many generations. In addition to processing and sequencing eDNA samples for present-day analyses, there is value in storing physical biomass and purified DNA from the environment. Proactive collection and storage of eDNA samples (i.e. biobanking) for future use

would enable retrospective analysis of biodiversity as analysis technologies advance. Growing interest in DNA for data storage and recovery in nonnatural systems attests to these concepts [32,33]. In considering biobanking strategies, there are advantages and disadvantages to both unprocessed samples (allow for future DNA extraction technologies but are bulkier and typically require ultracold storage) and purified DNA (less bulky and more stable, but some DNA may have been fragmented or poorly extracted). The success of biobanking will depend on committed institutions with resources to store and archive both raw samples and purified DNA, including high-fidelity sample and metadata tracking, compliance reporting, and communication among interested parties. Minimal variables to report include the precise location and time of sampling, volume of water collected, method of filtration (e.g. filter type and pore size), time from collection to preservation and/or filtration, how samples were stored (e.g. preservatives, temperature), and (later) methods of DNA extraction.

In addition to standard eDNA sample collection methods, uncrewed or automated system demonstrations have recently shown promise for collection from challenging-to-access environments such as the deep ocean [28,34,35]. While early moored *in situ* capabilities integrated quantitative PCR and microarray technologies [36], new autonomous collection platforms are currently limited to collection and preservation. Development of a next-generation autonomous *in situ* eDNA analysis capability (e.g. extraction, sequencing, and/or quantification) would greatly accelerate data availability by generating data at the point of collection, rather than in dedicated laboratories. Such platforms would require seaworthy micro- or millifluidic preparation methods, along with embedded data analysis and efficient data distribution capabilities, to fully realize eDNA assessment potential. Such an analysis capability would enable future platforms to operate similarly to Argo floats over extended periods of time to directly describe biological communities, expanding the concept of biogeochemical Argo (BGC-Argo) floats that are currently deployed and collecting physical, chemical, and bio-optical oceanographic data throughout the global ocean [37].

DNA extraction and sequencing

Extracting eDNA from a filtered sample requires chemical, enzymatic, and/or physical disruption of biological material to release DNA followed by extraction and purification. Owing to the low biomass of many species' DNA in marine samples, this DNA is both precious and prone to contamination. Most laboratories establish precautions for eDNA sample handling and DNA extraction to prevent misinterpretation of results due to laboratory contamination. These practices include sterile technique, one-way flow of samples in the lab to separate pre- and post-PCR activities, and physical isolation of samples throughout the analysis process to prevent cross-contamination. As an extreme example of the utility of careful lab practices, ancient eDNA samples have been used to identify entire ecosystems from ancient DNA in Greenland [38].

Analysis of eDNA at the point of collection offers opportunities for rapid data acquisition and cost reduction but will require newer, simpler technologies to be widely and reproducibly

deployed. Early examples of these analysis capabilities include qPCR approaches [39], linear amplification, and shipboard eDNA metabarcoding using nanopore sequencing [40]. Future platforms for complete process automation, from collection to end-point analysis, would drastically reduce the hands-on requirements for eDNA analysis while enabling a new data stream for environmental managers.

Data analysis, management, and integration

A given eDNA project may generate tens to hundreds of gigabytes of DNA sequence data, a trove of information that requires careful curation for maximal utility. To get from sequence data to actionable knowledge, the data must be transferred to a computing resource, undergo primary processing (e.g. calling of amplicon sequence variants in the case of metabarcoding), undergo secondary processing (statistical analysis and data visualization), and then be deposited and integrated into portals of larger datasets. Challenges for eDNA data include computational and human time demands to analyze large DNA sequence datasets, incomplete reference databases, the need to standardize, and ensure data are shared and reusable, which will facilitate interoperability and integration for maximal benefit. Proper management of data can improve both the efficiency of individual data analyses and the integration of multiple and diverse datasets.

Rapid bioinformatics analysis of eDNA sequence data is being enabled by multiple emerging bioinformatics technologies. Algorithmic improvements, including the use of artificial intelligence and machine learning (AI/ML), are increasingly being explored for DNA sequence data [41]. Cloud computing is facilitating these additional processing capabilities (e.g. multiple CPUs and GPUs), and cloud storage is enabling greater access to data on multiple platforms in addition to critical storage and backup capacity, facilitated by command-line tools (e.g. rclone) to transfer data between cloud servers and HPC resources. Bespoke analyses (setting up each new analysis from scratch) are inefficient and are increasingly being replaced by bioinformatics workflows. Workflows using standard tools (e.g. Nextflow, Snakemake) allow analyses that are more rapid, iterable, and tractable [42–46]. Containers (e.g. Docker, Singularity) allow workflows to be run on multiple computing environments [45]. Workflows for metabarcoding are numerous [47,48] and hold great potential for metagenomics and statistical analysis and data visualization.

Reference databases for eDNA data are critical to assign sequences to taxa, both in metabarcoding [49,50] and metagenomic studies [51,52]. A significant fraction of marine organisms (whole microbial genomes and protist/metazoan mitochondrial genomes) are missing from reference databases or are incorrectly annotated. Sequencing more reference specimens is ongoing for both microbes and higher organisms [53]. However, parallel avenues of inquiry where reference databases are not explicitly required are also being used. In metabarcoding, sequences were once clustered into operational taxonomic units (OTUs) in a bespoke manner, such that taxonomic assignments were linked to the particular OTU assignments of a study. Newer methods of denoising (removing errors from) sequences are commonly used now [54,55], generating ASVs that are comparable across studies [24]. This means that sequences

can be analyzed without assigning each sequence to a reference, which can be added later when available. Various sequence-based analyses can also be performed without the need for taxonomic assignment. ASV software requires continued development to maintain its utility on newer sequencing platforms.

Standardization of data formats is a critical priority to ensure eDNA data are findable, accessible, interoperable, and reusable (Box 2). In many cases, sample-associated metadata are often incomplete or formatted differently depending on the data type or the persons who generated it, hindering large-scale meta-analysis. For example, latitude and longitude may be reported in degrees/minutes/seconds (human-readable) rather than decimal degrees (machine-readable). The use of data standards such as those from the International Organization for Standardization (ISO) and the GSC provides a way to standardize data to make it interoperable. In cases where data standards do not cover all variables or use cases or there are multiple competing standards, there is a need to expand and harmonize standards. Institutional repositories for environmental and DNA sequence data, described next, are in a position to enforce community standards such as those from ISO and GSC.

Sharing of eDNA data and environmental metadata (Box 1) is critical to enable researchers to leverage collective data and advance the field, but too often, data are not deposited into repositories in a timely fashion, are not well-documented, and associated environmental metadata are poorly described or not linked or provided. Repositories for sequence data (National Center for Biotechnology Information, NCBI), biodiversity data (Ocean Biodiversity Information System, OBIS, https://obis.org/), and environmental data (National Center for Environmental Information) need to be fully embraced and used by the community. To ensure data are uploaded in a timely fashion, funding agencies may put timelines on when data should be submitted to repositories. To ensure links between different datasets from the same project, persistent identifiers such as DOIs (digital object identifiers) should be used and cross-linked. Processed data can in principle be deposited to sequence-focused databases (e.g. taxa observation tables to NCBI) and biodiversity-focused repositories (e.g. species observation records to OBIS); continuing to develop and use guidance for these resources [56] will be important to improve the archiving of these important data. Finally, finding datasets relevant to a given question—even if they have been properly formatted and deposited—is often difficult. Metadata need to be complete to enable users to search by region (e.g. latitude, longitude), data type (metabarcoding, metagenomic, etc.), generating institution, sequences or taxa present, and environmental or other metadata.

Integration of eDNA data with other types of marine data, facilitated through data sharing, is critical because a single type of eDNA data (e.g. metabarcoding data for a given genetic locus) tells us only about the presence/absence or relative abundance of marine taxa captured by that method. By combining multiple data types, including both additional eDNA data (e.g. other metabarcoding loci or whole-genome shotgun data) and other forms of biological and environmental data, a much richer picture can emerge that allows more synergistic and holistic ecosystem understanding, including the development of predictive ocean biodiversity models.

Data types that may enrich marine eDNA data include bio-optical, pigment, chemical (e.g. nutrients), physical (e.g, temperature, salinity), acoustic, and satellite data [57]. For example, eDNA can tell us which species are/were present in a small number of discrete samples within a water mass, while acoustics can tell us how much biomass is present in the water mass as a whole, with multiple depths and size classes, and these two data types can be combined to obtain a much richer perspective on fish populations. Given that marine data take many forms—DNA sequences but also qPCR cycle numbers, images, diversity metrics, and various tabular formats—converting or summarizing data into a tabular format (e.g. tab-separated or comma-separated values) is a solution that enables diverse data types to be analyzed via a single data analysis platform (e.g. R or Python/Pandas). This requires decisions about how best to distill complex data into scalar, categorical, or boolean data that lends itself to the tabular format. With sufficient sampling density, frequency, and detail, something akin to 'genomic weather maps' is possible. For example, efforts such as Bio-GO-SHIP [57] are discovering global biogeographic patterns (e.g. of nutrient limitation) using metagenomic sequencing [58,59].

Conclusions

The use of eDNA technology in marine systems is rapidly expanding. For this technology to have broad utility as a tool for marine biological surveillance at a global scale, practices at each step of the eDNA workflow need to be better standardized, automated, and managed. Equally important, eDNA's technology development needs to be informed by its ecological applications—the practical work necessary to track fine- and broad-scale biological change. Wider development and use of standards for sampling, sequencing, analysis, and data formatting will enable the global and decadal scales of Earth's changes to be detectable and potentially predictable. Bioinformatics workflows, supported by containers and cloud resources, will accelerate data analysis, and proper data management will allow both data and results to give life to studies beyond initial publications. Finally, we encourage the biobanking of both raw environmental samples and purified DNA, which will enable future extraction and DNA sequencing methods to reveal even further the deep biological information stored in nature's information storage molecule.

Acknowledgements

We thank Ryan Kelly, Chris Meyer, and Francisco Chavez for helpful comments on the direction of the paper, Nastassia Patin and Jasmin John for reviewing the paper, Joaquin Trinanes for discussions on the special issue, and Jaime Martínez-Urtaza, April Nishimura, and the editorial staff at Elsevier for guidance throughout the process. This work was supported by NOAA's Office of Oceanic and Atmospheric Research, U.S. Department of Commerce (award NA21OAR4320190), and Department of Navy's Office of Naval Research (award N00014-21-1-2601).

Figures, Tables, Boxes

Figure 1. Marine eDNA may consist of organismal tissues and waste, whole intact cells, broken or decaying cells, and free nucleic acids. Total DNA is extracted to yield purified DNA, which can then be subjected to molecular analyses. qPCR uses species-specific primers (e.g. goliath grouper) and amplification cycle thresholds to quantify the abundance of individual taxa. Metabarcoding uses amplification with group-specific primers (e.g. bacteria, fish) and sequencing by synthesis (SBS) to profile whole communities of organisms. Shotgun metagenomics uses SBS or other next-generation sequencing technologies to sequence pieces of all of the DNA in a sample, providing either community profiles or assembled genomes.

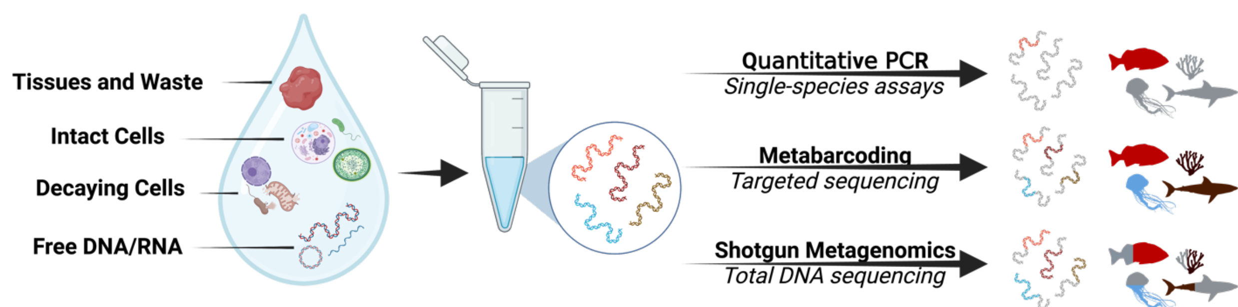

Figure 2. Processing steps in eDNA analysis from water collection to DNA sequence analysis.

Table 1. Processing steps in eDNA analysis from water collection to DNA sequence analysis. At each stage of the workflow are listed the best practices currently in use, logistical challenges, and emerging technologies that are being developed. We highlight major technical focus areas that need to be addressed for widespread adoption of eDNA as a basic research or biomonitoring toolset.

	Sample Collection	**Biomaterial Filtration & Biobanking**	**eDNA Extraction**	**Sequencing Preparation**	**DNA Sequencing**	**Data Analysis**
Current Best Practices	● Niskin/CTD sampling at sea ● Hand sampling ● Environmental data and metadata collection	● Disc/cartridge filters ● No size fractionation ● Ultracold storage or preservative	● Ultraclean lab practices ● One-way flow of samples ● Separation of pre-PCR from PCR/post-PCR	● Metabarcoding of multiple loci ● Whole-genome shotgun metagenomics ● qPCR ● Library QA/QC	● 2G platforms (SBS, ion semiconductor sequencing) ● Multiplexing of 100s of samples	● Independent tools (aligners, classifiers) ● End-to-end pipelines (open source)
Logistical Challenges	● Low biomass in the environment ● DNA dilution and degradation ● Spatial/temporal coverage limits	● Collection capacity ● Storage capacity	● Processing cost ● Processing time ● Contamination risk	● Processing cost ● Processing time	● Processing cost ● Incomplete genome assemblies	● Analyst time ● Reference databases ● Slow adoption of metadata standards
Emerging Technologies	● Autonomous systems ● Passive capture ● Ships of opportunity	● In situ preservation ● Biobanking	● Robotic DNA extraction ● Shipboard extraction	● Automated platforms ● Microfluidic devices ● Digital droplet PCR	● 3G platforms (nanopore sequencing, sequencing by binding) ● Improvements to 2G and 3G platforms	● AI-driven analysis ● Reference-free methods

Box 1. Collaboration

Coordination among organizations both nationally and internationally is critical to increasing efficiency and interoperability. While eDNA biomonitoring has been assessed as an inexpensive alternative to traditional survey approaches, the cost of sample collection, processing, and data analysis is unlikely to be translated into large-scale observational capabilities without motivating coordination within the resource management and research community. A problem is that each group has its own method of DNA collection, DNA extraction, sequencing, analysis, and metadata curation. Groups talk only occasionally, and they have different languages and vocabularies. As a result, data are minimally comparable.

Improved standardization of eDNA methods and interoperability of eDNA data can come through improved communication. It is unrealistic to expect every group to use the same methods for each stage of the process. A more reasonable approach is for groups to be transparent and communicate their methods frequently and in detail, through in-person and virtual meetings and rapid and clear publication of methods and practices. This is likely to lead to convergence of methodologies and development of best practices. Collaboration will be critical to supporting large, global efforts that could not be undertaken by a single research group or institution and to creating the standards that will enable researchers to integrate eDNA data from across the global ocean.

International organizations have formed around the goal of sharing and standardizing marine data, practices, and eDNA monitoring efforts. For ocean biodiversity data, the Ocean Biogeographic Information System (OBIS) has paved the way for large-scale data aggregation and dissemination and offers invaluable insight into next-generation systems [60]. OBIS is an ocean-focused participant in the Global Biodiversity Information Facility (GBIF, https://www.gbif.org/), uses the DarwinCore (DwC) standard for recording species observation events, and a guide now exists for publishing sequence-derived data there [56]. For practices, the Ocean Best Practices System [61], Earth System Information Practices, National Microbiome Data Collaborative, and Genomic Standards Consortium (GSC) are notable organizations bringing researchers together globally to develop standard methods. Monitoring efforts are being organized by Marine Life 2030 and the UN Decade of Ocean Sciences, the Atlantic Ocean Research Alliance, AtlanECO, the Tara Oceans Foundation, the Marine Biodiversity Observation Network, and the Ocean Biomolecular Observing Network. As an example of governmental coordination efforts, US Government agencies are coordinating to expand global-scale forensics and biosurveillance efforts, including monitoring of invasive species (US Geological Survey), bioterrorism (Department of Homeland Security), food security (US Department of Agriculture), port security (US Coast Guard), and climate change (National Oceanic and Atmospheric Administration).

Box 2. Accessibility

DNA sequencing remains a highly technical process, with significant capital investments required to outfit traditional laboratories with equipment and skilled personnel. Additionally, consumables and reagents used for eDNA extraction are a significant proportion of overall sample analysis costs; future platforms that enable laboratory process automation and miniaturization are likely to decrease personnel and material costs while providing time savings and expedited data availability. We envision autonomous collection systems that would integrate many steps of eDNA collection, archiving, and analyses to reduce the need for laboratory-based sample processing, shifting reliance from traditional laboratory facilities to equipment near the point of collection.

Analysis of eDNA data requires significant computational skill, especially as data volumes increase with adoption of newer DNA sequencing platforms. Future data meta-analysis efforts will benefit as methods for data generation and metadata recording are standardized, and these data are likely to increase the value of samples and datasets for future biomonitoring efforts. To improve data standardization and integration, we recommend continued community development of well-defined standards and submission requirements; training resources (data management guides, workshops, online websites, and wikis) and education about the FAIR data principles (findable, accessible, interoperable, and reusable) [62]; and positive incentives, including community ratings for metadata and data quality and citable data digital object identifiers (DOIs).

As environmental change accelerates in many parts of the world, attempts to expand eDNA biomonitoring in less developed or remote regions are likely to increase. The development of highly accessible methods is critical to the uptake of these technologies, especially in parts of the world that may be limited in their uptake of new technologies for ocean resource management due to cost or technical skill barriers [18].

Annotated References

Outstanding Interest

- [24] (Callahan et al. 2017) – Why amplicon sequence variants (ASVs) are superior to operational taxonomic units (OTUs) in representing exact sequence diversity and enabling cross-study meta-analyses.
- [6] (Djurhuus et al. 2020) – Demonstrates the application of multilocus amplicon sequencing, using metabarcoding of four loci to reveal seasonal patterns in Monterey Bay, California.
- [31] (Patin and Goodwin 2022) – A comprehensive review of eDNA methods focused on collection and preservation of microbial and eukaryotic eDNA.

Special Interest

- [56] (Finstad et al. 2020) – Provides detailed guidelines for submitting DNA-derived occurrence data into GBIF or OBIS using the DarwinCore standard, including example field entries for metabarcoding and quantitative PCR data.
- [53] (Hoban et al. 2022) – The genome skimming method described is an efficient method for reference marker gene sequencing and is now being deployed toward sequencing the mitochondrial genomes of all US marine fishes.
- [60] (Klein et al. 2019) – OBIS is a critical component in marine biodiversity data sharing, and is increasingly capable of storing genetic observational data.
- [26] (Leigh et al. 2021) – Highlights the history, challenges, and future potential of large-scale genetic analysis for multi-cellular organisms (e.g. macrogenomics).
- [29] (Minamoto et al. 2021) – An illustrated guide to eDNA collection methods from the Japanese eDNA Society.
- [3] (Sanders et al. 2019) – Provides a low-cost and high-throughput method for metagenomic sequencing and assembly using a combination of long and short reads.
- [62] (Thompson et al. 2020) – A guide to metadata and ontologies, which are critical for integrating large eDNA and microbiome datasets.

References

1. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF: **Community structure and metabolism through reconstruction of microbial genomes from the environment**. *Nature* 03 2004, **428**:37–43.

2. Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, et al.: **The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific**. *PLoS Biol* 2007, **5**:e77.

3. Sanders JG, Nurk S, Salido RA, Minich J, Xu ZZ, Zhu Q, Martino C, Fedarko M, Arthur TD, Chen F, et al.: **Optimizing sequencing protocols for leaderboard metagenomics by combining long and short reads**. *Genome Biol* 10 2019, **20**:1– 14.

4. Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, Wu D, Paez-Espino D, Chen I-M, Huntemann M, et al.: **A genomic catalog of Earth's microbiomes**. *Nat Biotechnol* 2020, doi:10.1038/s41587-020-0718-6.

5. Thomsen PF, Kielgast J, Iversen LL, Møller PR, Rasmussen M, Willerslev E: **Detection of a Diverse Marine Fish Fauna Using Environmental DNA from Seawater Samples**. *PLoS One* 08 2012, **7**:e41732.

6. Djurhuus A, Closek CJ, Kelly RP, Pitz KJ, Michisaki RP, Starks HA, Walz KR, Andruszkiewicz EA, Olesin E, Hubbard K, et al.: **Environmental DNA reveals seasonal shifts and potential interactions in a marine community**. *Nat Commun* 01 2020, **11**:1– 9.

7. Wilcox TM, McKelvey KS, Young MK, Jane SF, Lowe WH, Whiteley AR, Schwartz MK: **Robust Detection of Rare Species Using Environmental DNA: The Importance of Primer Specificity**. *PLoS One* 2013, **8**:e59520.

8. Ramón-Laca A, Wells A, Park L: **A workflow for the relative quantification of multiple fish species from oceanic water samples using environmental DNA (eDNA) to support large-scale fishery surveys**. *PLoS One* 2021, **16**:e0257773.

9. Giovannoni SJ, Britschgi TB, Moyer CL, Field KG: **Genetic diversity in Sargasso Sea bacterioplankton**. *Nature* 05 1990, **345**:60– 63.

10. Dupont CL, Rusch DB, Yooseph S, Lombardo M-J, Richter RA, Valas R, Novotny M, Yee-Greenbaum J, Selengut JD, Haft DH, et al.: **Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage.** *ISME J* 2011, doi:10.1038/ismej.2011.189.

11. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, et al.: **Structure and function of the global ocean microbiome**. *Science* 05 2015, **348**:1261359– 1261359.

12. Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM: **A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA Genes**. *PLoS One* 07 2009, **4**:e6372.

13. Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner H-W, Richards TA: **Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water**. *Mol Ecol* 2010, **19**:21–31.

14. Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, Minamoto T, Yamamoto S, Yamanaka H, Araki H, et al.: **MiFish, a set of universal PCR primers for metabarcoding**

environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Society Open Science* 2015, **2**:150088.

15. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, et al.: **Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms**. *ISME J* 03 2012, doi:10.1038/ismej.2012.8.

16. Muller-Karger FE, Miloslavich P, Bax NJ, Simmons S, Costello MJ, Pinto IS, Canonico G, Turner W, Gill M, Montes E, et al.: **Advancing Marine Biological Observations and Data Requirements of the Complementary Essential Ocean Variables (EOVs) and Essential Biodiversity Variables (EBVs) Frameworks**. *Frontiers in Marine Science* 2018, **5**:211.

17. Muller-Karger F, Kavanaugh M, Iken K, Montes E, Chavez F, Ruhl H, Miller R, Runge J, Grebmeier J, Cooper L, et al.: **Marine Life 2030: Forecasting changes to ocean biodiversity to inform decision-making: A critical role for the Marine Biodiversity Observation Network (MBON)**. *Mar Technol Soc J* 2021, **55**:84–85.

18. De León LF, Silva B, Avilés-Rodríguez KJ, Buitrago-Rosas D: **Harnessing the omics revolution to address the global biodiversity crisis**. *Curr Opin Biotechnol* 2023, **80**:102901.

19. Thomsen PF, Willerslev E: **Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity**. *Biol Conserv* 03 2015, **183**:4– 18.

20. Suarez-Bregua P, Álvarez-González M, Parsons KM, Rotllant J, Pierce GJ, Saavedra C: **Environmental DNA (eDNA) for monitoring marine mammals: Challenges and opportunities**. *Frontiers in Marine Science* 2022, **9**.

21. Polanco Fernández A, Marques V, Fopp F, Juhel J-B, Borrero-Pérez GH, Cheutin M-C, Dejean T, González Corredor JD, Acosta-Chaparro A, Hocdé R, et al.: **Comparing environmental DNA metabarcoding and underwater visual census to monitor tropical reef fishes**. *Environmental DNA* 2021, **3**:142–156.

22. Oka S-I, Doi H, Miyamoto K, Hanahara N, Sado T, Miya M: **Environmental DNA metabarcoding for biodiversity monitoring of a highly diverse tropical fish community in a coral reef lagoon: Estimation of species richness and detection of habitat segregation**. *Environmental DNA* 2021, **3**:55–69.

23. Sawaya NA, Djurhuus A, Closek CJ, Hepner M, Olesin E, Visser L, Kelble C, Hubbard K, Breitbart M: **Assessing eukaryotic biodiversity in the Florida Keys National Marine Sanctuary through environmental DNA metabarcoding**. *Ecol Evol* 2019, **9**:1029–1040.

24. Callahan BJ, McMurdie PJ, Holmes SP: **Exact sequence variants should replace operational taxonomic units in marker-gene data analysis.** *ISME J* 07 2017, doi:10.1038/ismej.2017.119.

25. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, et al.: **A communal catalogue reveals Earth's multiscale microbial diversity**. *Nature* 2017, **551**:457–463.

26. Leigh DM, van Rees CB, Millette KL, Breed MF, Schmidt C, Bertola LD, Hand BK, Hunter ME, Jensen EL, Kershaw F, et al.: **Opportunities and challenges of macrogenetic studies**. *Nat Rev Genet* 2021, **22**:791–807.

27. Doi H, Uchii K, Takahara T, Matsuhashi S, Yamanaka H, Minamoto T: **Use of Droplet Digital PCR for Estimation of Fish Abundance and Biomass in Environmental DNA Surveys**. *PLoS One* 2015, **10**:e0122763.

28. Den Uyl PA, Thompson LR, Errera RM, Birch JM, Preston CM, Ussler W, Yancey CE, Chaganti SR, Ruberg SA, Doucette GJ, et al.: **Lake Erie field trials to advance autonomous monitoring of cyanobacterial harmful algal blooms**. *Frontiers in Marine Science* 2022, **9**.

29. Minamoto T, Miya M, Sado T, Seino S, Doi H, Kondoh M, Nakamura K, Takahara T, Yamamoto S, Yamanaka H, et al.: **An illustrated manual for environmental DNA research: Water sampling guidelines and experimental protocols**. *Environmental DNA* 2021, **3**:8–13.

30. Buxton A, Matechou E, Griffin J, Diana A, Griffiths RA: **Optimising sampling and analysis protocols in environmental DNA studies**. *Sci Rep* 2021, **11**:11637.

31. Patin NV, Goodwin KD: **Capturing marine microbiomes and environmental DNA: A field sampling guide**. *Front Microbiol* 2023, **13**.

32. Church GM, Gao Y, Kosuri S: **Next-generation digital information storage in DNA**. *Science* 2012, **337**:1628.

33. Ceze L, Nivala J, Strauss K: **Molecular digital data storage using DNA**. *Nat Rev Genet* 2019, **20**:456–466.

34. Truelove NK, Patin NV, Min M, Pitz KJ, Preston CM, Yamahara KM, Zhang Y, Raanan BY, Kieft B, Hobson B, et al.: **Expanding the temporal and spatial scales of environmental DNA research with autonomous sampling**. *Environmental DNA* 2022, **4**:972–984.

35. Govindarajan AF, McCartin L, Adams A, Allan E, Belani A, Francolini R, Fujii J, Gomez-Ibañez D, Kukulya A, Marin F, et al.: **Improved biodiversity detection using a large-volume environmental DNA sampler with in situ filtration and implications for marine eDNA sampling strategies**. *bioRxiv* 2022, doi:10.1101/2022.01.12.475892.

36. Roman B, Scholin C, Jensen S, Marin R, Massion E, Feldman J: **The 2nd generation Environmental Sample Processor: Evolution of a robotic underwater biochemical laboratory**. In *Proceedings of OCEANS 2005 MTS/IEEE*. . 2005:1–9.

37. Bittig HC, Maurer TL, Plant JN, Schmechtig C, Wong APS, Claustre H, Trull TW, Udaya Bhaskar TVS, Boss E, Dall'Olmo G, et al.: **A BGC-Argo Guide: Planning, Deployment, Data Handling and Usage**. *Frontiers in Marine Science* 2019, **6**.

38. Kjær KH, Winther Pedersen M, De Sanctis B, De Cahsan B, Korneliussen TS, Michelsen CS, Sand KK, Jelavić S, Ruter AH, Schmidt AMA, et al.: **A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA**. *Nature* 2022, **612**:283–291.

39. Voelker CR, Ochoa AR, Armstrong-Spenrath L ’quita, Lott L, McDaniel JS, Blackburn AN, Cornell LE, Mahoney R, Asin SN: **Evaluating sensitivity and specificity of the Biomeme Franklin^{TM} three9 real-time PCR device and SARS-CoV-2 go-strips assay using clinical samples**. *J Clin Virol* 2022, **146**:105046.

40. Truelove NK, Andruszkiewicz EA, Block BA: **A rapid environmental DNA method for detecting white sharks in the open ocean**. *Methods Ecol Evol* 2019, **10**:1128–1135.

41. Liang Q, Bible PW, Liu Y, Zou B, Wei L: **DeepMicrobes: taxonomic classification for metagenomics with deep learning**. *NAR Genomics and Bioinformatics* 2020, **2**:lqaa009.

42. Keegan KP, Glass EM, Meyer F: **MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function**. In *Microbial Environmental Genomics (MEG)*. Edited by Martin F, Uroz S. Springer New York; 2016:207–233.

43. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, et al.: **KBase: The United States Department of Energy Systems Biology Knowledgebase**. *Nat Biotechnol* 2018, **36**:566–569.

44. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Tommaso PD, Nahnsen S: **The nf-core framework for community-curated bioinformatics pipelines**. *Nat Biotechnol* 2020, **38**:276–278.

45. Thompson LR, Anderson SR, Den Uyl PA, Patin NV, Lim SJ, Sanderson G, Goodwin KD: **Tourmaline: A containerized workflow for rapid and iterable amplicon sequence analysis using QIIME 2 and Snakemake**. *Gigascience* 2022, **11**.

46. Chen I-MA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, Hajek P, Ritter SJ, Webb C, Wu D, et al.: **The IMG/M data management and analysis system v.7: content updates and new features**. *Nucleic Acids Res* 2023, **51**:D723–D732.

47. Zafeiropoulos H, Viet HQ, Vasileiadou K, Potirakis A, Arvanitidis C, Topalis P, Pavloudi C, Pafilis E: **PEMA: a flexible Pipeline for Environmental DNA Metabarcoding Analysis of the 16S/18S ribosomal RNA, ITS, and COI marker genes**. *Gigascience* 2020, **9**.

48. Reiter T, Brooks† PT, Irber† L, Joslin† SEK, Reid† CM, Scott† C, Brown CT, Pierce-Ward NT: **Streamlining data-intensive biology with workflow systems**. *Gigascience* 2021, **10**.

49. McMurdie PJ, Holmes S: **Phyloseq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data.** *Pac Symp Biocomput* 2012,

50. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, et al.: **Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2**. *Nat Biotechnol* 2019, **37**:852–857.

51. Ounit R, Wanamaker S, Close TJ, Lonardi S: **CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers**. *BMC Genomics* 2015, **16**:236.

52. Wood DE, Lu J, Langmead B: **Improved metagenomic analysis with Kraken 2**. *Genome*

Biol 2019, **20**:257.

53. Hoban ML, Whitney J, Collins AG, Meyer C, Murphy KR, Reft AJ, Bemis KE: **Skimming for barcodes: rapid production of mitochondrial genome and nuclear ribosomal repeat reference markers through shallow shotgun sequencing**. *PeerJ* 2022, **10**:e13790.

54. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP: **DADA2: High-resolution sample inference from Illumina amplicon data**. *Nat Methods* 2016, **13**:581–583.

55. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Xu ZZ, Kightley EP, Thompson LR, Hyde ER, González A, et al.: **Deblur rapidly resolves single-nucleotide community sequence patterns**. *mSystems* 2017, **2**.

56. Finstad AG, Andersson A, Bissett A, Fossøy F, Grosjean M, Hope M, Kõljalg U, Lundin D, Nilsson H, Prager M, et al.: **Publishing sequence-derived data through biodiversity data platforms**. 2020, doi:10.35035/DOC-VF1A-NR22.

57. Clayton S, Alexander H, Graff JR, Poulton NJ, Thompson LR, Benway H, Boss E, Martiny A: **Bio-GO-SHIP: The Time Is Right to Establish Global Repeat Sections of Ocean Biology**. *Frontiers in Marine Science* 2022, **8**:767443.

58. Larkin AA, Garcia CA, Garcia N, Brock ML, Lee JA, Ustick LJ, Barbero L, Carter BR, Sonnerup RE, Talley LD, et al.: **High spatial resolution global ocean metagenomes from Bio-GO-SHIP repeat hydrography transects**. *Scientific Data* 2020, **8**:2020.09.06.285056.

59. Ustick LJ, Larkin AA, Garcia CA, Garcia NS, Brock ML, Lee JA, Wiseman NA, Moore JK, Martiny AC: **Metagenomic analysis reveals global-scale patterns of ocean nutrient limitation**. *Science* 2021, **372**:287–291.

60. Klein E, Appeltans W, Provoost P, Saeedi H, Benson A, Bajona L, Peralta AC, Bristol RS: **OBIS Infrastructure, Lessons Learned, and Vision for the Future**. *Frontiers in Marine Science* 2019, **6**.

61. Horstmann C, Buttigieg PL, Simpson P, Pearlman J, Karstensen J, Waite AM: **Towards a Best Practice for Developing Best Practices in Ocean Observation (BP4BP): Supporting Methodological Evolution through Actionable Documentation,**. *IOC Manuals and Guides* 2020, doi:10.25607/OBP-781.

62. Thompson L, Vangay P, Blumberg K, Christianson D, Dundore-Arias J, Hu B, Timme R, Wood-Charlson E: **Introduction to metadata and ontologies: Everything you always wanted to know about metadata and ontologies (but were afraid to ask)**. 2020, doi:10.25979/1607365.