Received: 9 March 2021 | Revised: 12 April 2021 | Accepted: 15 April 2021

DOI: 10.1111/1755-0998.13399

MOLECULAR ECOLOGY RESOURCES WILEY

EDITORIAL

Biodiversity monitoring using environmental DNA

Monitoring biodiversity is essential to protect, preserve and restore ecosystems, particularly in the context of current challenges such as climate change, habitat destruction and globalization (Baird & Hajibabaei, 2012). Biomonitoring is needed for developing biotic indices for assessing ecological status, measuring impacts of anthropogenic activities in natural ecosystems, evaluating biodiversity loss, surveying nonindigenous species, conservation, and identifying cryptic species (Balvanera et al., 2006; Fišer et al., 2018). Thus, spatially and temporally structured biomonitoring activities provide a powerful tool for the implementation of regional, national and international regulations, directives and policies for nature conservation. However substantial impediments exist including access to remote locations, limited specialist taxonomic knowledge, cost, slow pace of human-driven data analyses, and typically low sensitivity for detection of rare and elusive species (Zinger et al., 2020). These drawbacks are often translated into expensive monitoring activities with limited spatial, temporal and taxonomic coverage. In this context, new approaches for biomonitoring are being explored, among which advanced DNA-based technologies are emerging (Kissling et al., 2018). The field of biodiversity monitoring through the analysis of the pool of DNA isolated from environmental samples, referred to as environmental DNA or eDNA (Pawlowski et al., 2020; Taberlet et al., 2012), is rapidly growing. This growth is being driven through improved approaches for sampling, data generation and analyses, and with recent advances on how eDNA should be interpreted for biodiversity assessments (Bohmann et al., 2014). The success of eDNA-based biomonitoring is reflected in exponential growth of publications within this area and increasing submissions to *Molecular Ecology Resources* in particular (Figure 1). *Molecular Ecology Resources* aims to publish high quality eDNA studies that serve as broad resources, including innovative methodologies for DNA sampling, enhanced laboratory protocols for data generation, or new computer programs and statistical advances for data analyses. Thus, the aim of this editorial is to contribute to producing good quality DNA data-derived essential biodiversity variables (EBVs) (Kissling et al., 2018) by providing guidance to the community submitting articles on the subject. For that purpose, we have summarized best practices established in published literature related to the different phases involved in the process, namely sampling, laboratory work, bioinformatic analyses and data interpretation (Figure 2).

1 | SAMPLING: THE IMPORTANCE OF WHERE, WHEN AND HOW

Environmental DNA samples used for biomonitoring can be widely heterogeneous spanning terrestrial, aquatic and aerial habitats and representing contemporary or historical (e.g., permafrost or sediment cores) communities (Deiner et al., 2017; Jarman et al., 2018). These can be collected either directly from the environments (e.g., soil or water) that contain the organisms (e.g., microbes, microfauna) or traces of them (e.g., extracellular DNA, tissue, mucus), or from other samples types such as stomach contents or scats. In both cases, a variety of sampling techniques can be applied using simple (e.g., bucket of water) or more sophisticated gear (e.g., Niskin bottles), including artificial (e.g., Autonomous Reef Monitoring Structures [Pearman et al., 2020]) or biological samplers (e.g., sponges [Mariani et al., 2019], leeches [Schnell et al., 2012]), or even automatic on-site sampling devices (Scholin et al., 2017). Thus, sampling design and methods for sample collection will be different from study to study depending upon the research questions. Yet, in all cases, the sampling strategy should be designed to produce a representative picture of the focal community across the desired geographical range and time frame. Additionally, sampling should account for environmental and species distribution heterogeneity in space or time by including an appropriate number of spatial and/or temporal replicates. For example, employing multiple nested levels of sampling allows for the estimation of uncertainty associated with imperfect detection (e.g., false negatives and false positives) (Doi et al., 2019; McClenaghan et al., 2020) that is valuable in evaluations of species distributions and metapopulation dynamics (Martel et al., 2020). In the interest of correct data interpretation, a detailed description of the sampling process, including sample collection coordinates and associated metadata, amount of collected material, procedures followed to avoid sample contamination, storage conditions, and other relevant features to enable study replication, should be provided according to sample type, reporting the core information specified by the minimum information about any (x) sequence (MIxS) framework (https://gensc.org/mixs/; [Yilmaz et al., 2011]) and Darwin Core (Wieczorek et al., 2012), as recently recommended (Berry et al., 2020). Description of standard procedures can be omitted, but these should be always referred to with their

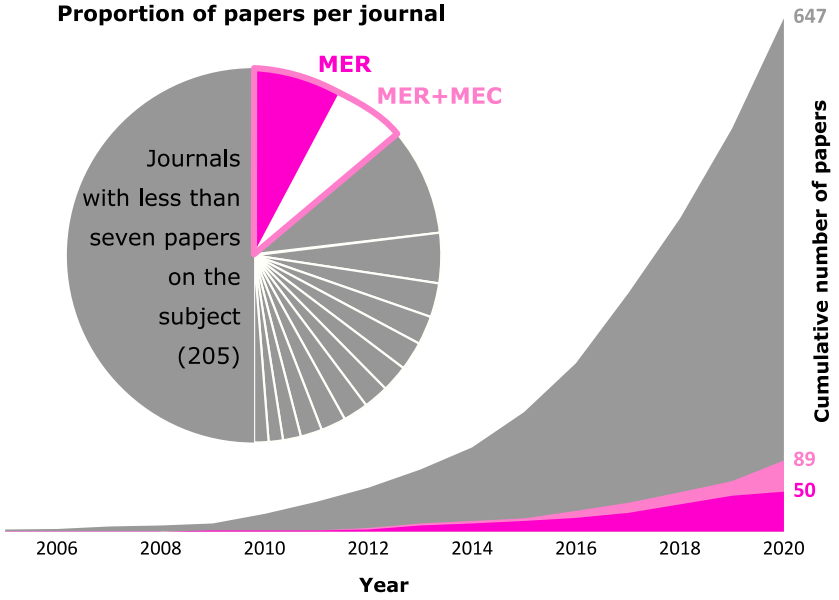

FIGURE 1 Proportion of papers on DNA based biomonitoring per journal, where *MolecularEcology Resources(MER)* and *Molecular Ecology(MEC)* together occupy the first position account for 13.7% of the papers, with *MER* accounting for 7.7% of them. Cumulative number of papers published per year on DNA based biomonitoring overall (grey) in *MER* and *MEC* together (light pink) and in *MER* only (dark pink). Numbers were obtained by searching using the search "([biodiversity AND monitoring] OR biomonitoring) AND (metabarcoding OR metagenomics OR qPCR OR ddPCR OR RT-PCR or barcoding) AND (DNA OR eDNA OR environmental DNA OR community DNA OR bulk DNA)" for the years 2006 to 2020 (both included) in Web of Knowledge (accessed in January 2021)

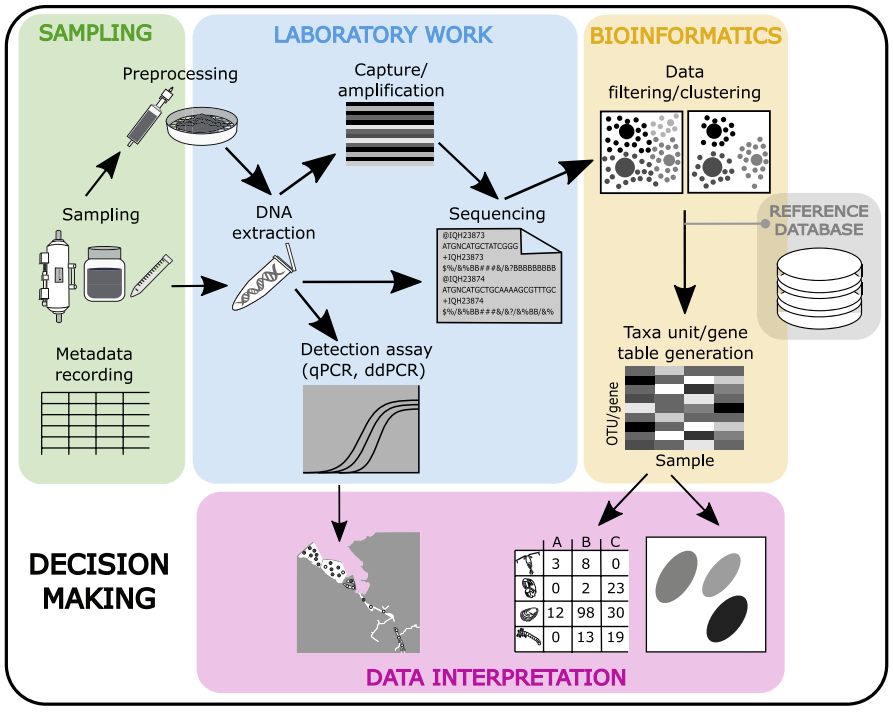

FIGURE 2 Steps involved in the process of using environmental DNA for biomonitoring. Samples can be directly used for DNA extraction or should be pre-processed (sieved, filtered), which can be done in the field or in the laboratory. Extracted DNA can be used for specific taxa detection and quantification using qPCR or ddPCR or sequenced either directly or after amplification or capture of targeted taxa. Bioinformatic analyses always include raw data cleaning and/or clustering steps, and interpretable taxa unit or gene tables can be built using or not a reference database. Analysis types for data interpretation are varied. The decision-making process should take into account all the uncertainties associated with all the steps within the process

appropriate original reference. Likewise, procedures used for pre-processing samples before DNA extraction (e.g., mesh size for water sample filtering or sediment sieving), and conditions (e.g., temperature) should be provided, as well as storage conditions of sample or preprocessed material. Sample collection for studies submitted to *Molecular Ecology Resources* are expected to comply with the Nagoya protocol (Marden et al., 2021).

2 | LABORATORY WORK: CONVERTING SAMPLES TO RAW DATA

DNA extraction should be performed in conditions that prevent external and cross-contamination among samples. As for sampling (see above), all procedures need to be provided, and also

references to the manufacturer and protocols for commercial kits, typically for DNA extraction. Once extracted, environmental DNA can be (i) interrogated for presence and quantification of a given taxon through a detection assay such as quantitative PCR (qPCR) or digital droplet PCR (ddPCR), (ii) enriched for a given taxonomic group before sequencing through PCR or capture (metabarcoding) or (iii) directly sequenced (metagenomics) (Figure 2). Primers and/or probes used for detection assays and metabarcoding should be carefully chosen to target the desired taxon in an unbiased way and specified in the protocol description by citing the original reference describing them if they were already published and providing their sequence. Amplification conditions, i.e., mix composition, including the exact reference of the polymerase used, and thermocycling conditions are also necessary. One should seek to provide all relevant details that would allow the experiment to be reproduced.

Each step of the laboratory work should include negative controls - from extraction to sequencing - to identify potential (cross-) contaminants, as well as technical replicates to provide information about the influence of stochasticity of DNA extraction and PCR. Likewise, including positive controls (e.g., the target species or a mock community made from various concentrations of DNA obtained from specimens of different species) enables verification that the laboratory work is not compromised, as well as the estimation of different parameters that could be used in downstream analyses (e.g., PCR/sequencing error rates, tag-jumps rates). An unambiguous and detailed description of all these procedures is especially important when using new and unpublished approaches and technologies, such as nanopore sequencing or capture. Additional considerations for metabarcoding analyses have been extensively discussed in Zinger et al. (2019). Raw metabarcoding/metagenomic data should be submitted to repositories such as SRA (https://www.ncbi.nlm.nih.gov/sra) or ENA (https://www.ebi.ac.uk/ena), and reference to accession numbers must be provided.

3 | BIOINFORMATICS: FROM RAW TO INTERPRETABLE DATA

For detection assays, such as qPCR, a threshold cycle (Ct value) indicating when a significantly fluorescent signal is detected above background should be informed. The Ct values are used for detection/nondetection of the target sequence and to determine the corresponding copy number. For qPCR it is important to report the limit of detection (LOD) or the Ct threshold indicative of detection versus a nondetection, and the limit of quantitation (LOQ) or the Ct value below which it is not possible to precisely quantify copy number (Forootan et al., 2017; Klymus et al., 2020). For dPCR or ddPCR estimates of copy number are a result of the assay and thus LOD and LOQ may not be required.

Metabarcoding and metagenomic projects involve converting raw high-throughput sequencing data such as that obtained from Illumina sequencers (e.g., MiSeq) into interpretable data tables (e.g., number of reads per taxonomic unit or genes per sampling point). A large suite of software packages can cover the whole or part of data processing, for example, mothur (Schloss et al., 2009), QIIME (Caporaso et al., 2010), OBITools (Boyer et al., 2016), SWARM (Mahé et al., 2014), UCHIME (Edgar et al., 2011), Meta-IDBA (Peng et al., 2011), Megahit (Li et al., 2015), and MetaBAT (Kang et al., 2015). Regardless of the software or of the analytical pipeline used, raw data processing from sequence reads to interpretable data tables should consider the experimental design and question to be addressed and should be developed with the purpose of distinguishing artefactual from true sequence variants. In that sense, steps to reduce chimeras, PCR/sequencing errors, tag-jumps or nuclear copies for mtDNA target sequences, should be applied. The data preprocessing and sequence clustering steps in amplicon sequence variants (ASV) or operational taxonomic units (OTU) should be described so that they can be easily replicated by the reader. Scripts used for data analyses must be provided to satisfy data accessibility requirements of the journal, ideally deposited into a permanent online repository with an assigned permanent DOI (e.g., FigShare, Zenodo), including a full documentation of software version (or stable release) used for data analysis. Likewise, when taxonomic assignment of the individual or clustered reads is performed, the method and parameter values used should be specified and the reference database provided as reference or website, including the version used, or the date when it was accessed.

4 | REFERENCE DATABASES: IMPROVING COMPLETENESS AND ACCURACY THROUGH CURATION

Accurate and complete reference databases are crucial for DNA based biomonitoring studies in general. Reference databases are needed for developing effective primers to amplify species or taxonomic groups of interest, as well as for achieving comprehensive taxonomic assignment for metabarcoding and metagenomic data. Databases required for primer design should include priming sites and generally barcode reference databases such as BOLD (https://www.boldsystems.org/), UNITE (https://unite.ut.ee/), SILVA (https://www.arb-silva.de/), PR2 (https://pr2-database.org/) have limitations unless barcodes/tags to be designed are within the region covered. Studies using in-house built databases or custom curated versions of public databases should specify the steps followed for sequence gathering and/or curation and provide a link to the resulting database.

5 | INTERPRETATION: FROM DATA TO BIOMONITORING

Data interpretation with biomonitoring purposes should consider the combination of knowns and unknowns implicit within the process, and decision-making agents should be informed of the associated

uncertainties (Gilbey et al., 2021). For example, the potential sampling, data generation and data analysis biases, which occur even when all possible sampling and analytical controls and data curation procedures have been implemented, would result in "known-unknowns"; that is, phenomena that we know occur, but which we do not know in which manner or magnitude affect each data set. Another example is the knowledge that some of the expected species are absent in reference databases or that the primers we are using do not amplify a given taxa; this would be the "unknown-knowns"; we do not know if that particular species or taxa is present in our sample, but we know we are missing this information. Thus, deriving biomonitoring conclusions from eDNA data requires being aware of these uncertainties and to quantify them as much as possible. For example, controls based on mock communities could be used to determine potential PCR condition or primer derived biases (Aylagas et al., 2016) and a prior evaluation of the completeness of reference databases could help assessing the reliability of eDNA derived biotic indices (Aylagas et al., 2014).

Naiara Rodríguez-Ezpeleta[1]
Lucie Zinger[2]
Andrew Kinziger[3]
Holly M. Bik[4]
Aurélie Bonin[5]
Eric Coissac[6]
Brent C. Emerson[7]
Carla Martins Lopes[8]
Tara A. Pelletier[9]
Pierre Taberlet[10,11]
Shawn Narum[12]

[1]*Marine Research, AZTI Basque Research and Technology Alliance, Sukarrieta, Bizkaia, Spain*
Email: nrodriguez@azti.es
[2]*Département de biologie, Institut de Biologie de l'ENS (IBENS, École normale supérieure, CNRS, INSERM, Université PSL, Paris, France*
[3]*Department of Fisheries Biology, Humboldt State University, Arcata, CA, USA*
[4]*Department of Marine Sciences and Institute of Bioinformatics, University of Georgia, Athens, Georgia, USA*
[5]*Department of Environmental Science and Policy, Università degli Studi di Milano, Milano, Italy*
[6]*Univ. Grenoble-Alpes, Univ. Savoie Mont Blanc, CNRS, LECA, Grenoble, France*
[7]*Island Ecology and Evolution Research Group, Institute of Natural Products and Agrobiology (IPNA-CSIC, La Laguna, Spain*
[8]*Departamento de Biodiversidade, Universidade Estadual Paulista (UNESP, Rio Claro, SP, Brazil*
[9]*Department of Biology, Radford University, Radford, VA, USA*
[10]*Univ. Grenoble-Alpes, Univ. Savoie Mont Blanc, CNRS, LECA, Grenoble, France*
[11]*The Arctic University Museum of Norway, UiT the Arctic University of Norway, Tromsø, Norway*
[12]*Hagerman Genetics Laboratory, Columbia River Inter-Tribal Fish Commission, University of Idaho, Hagerman, ID, USA*

ORCID

Naiara Rodríguez-Ezpeleta https://orcid.org/0000-0001-6735-6755
Lucie Zinger https://orcid.org/0000-0002-3400-5825
Andrew Kinziger https://orcid.org/0000-0002-8776-6230
Holly M. Bik https://orcid.org/0000-0002-4356-3837
Brent C. Emerson https://orcid.org/0000-0003-4067-9858
Carla Martins Lopes https://orcid.org/0000-0002-3277-1913

REFERENCES

Aylagas, E., Borja, Á., Irigoien, X., & Rodríguez-Ezpeleta, N. (2016). Benchmarking DNA metabarcoding for biodiversity-based monitoring and assessment. *Frontiers in Marine Science*, *3*, 1–12.

Aylagas, E., Borja, Á., & Rodríguez-Ezpeleta, N. (2014). Environmental status assessment using DNA metabarcoding: Towards a genetics based marine biotic index (gAMBI). *PLoS One*, *9*, e90529.

Baird, D. J., & Hajibabaei, M. (2012). Biomonitoring 2.0: A new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, *21*, 2039–2044.

Balvanera, P., Pfisterer, A. B., Buchmann, N., He, J.-S., Nakashizuka, T., Raffaelli, D., & Schmid, B. (2006). Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecology Letters*, *9*, 1146–1156.

Berry, O., Jarman, S., Bissett, A., Hope, M., Paeper, C., Bessey, C., Schwartz, M. K., Hale, J., & Bunce, M. (2020). *Making environmental DNA (eDNA) biodiversity records globally accessible*. Environmental DNA.

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., Yu, D. W., & de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, *29*, 358–367.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*, 176–182.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., & Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*, 335–336.

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S. et al (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, *26*, 5872–5895.

Doi, H., Fukaya, K., Oka, S.-I., Sato, K., Kondoh, M., & Miya, M. (2019). Evaluation of detection probabilities at the water-filtering and initial PCR steps in environmental DNA metabarcoding using a multispecies site occupancy model. *Scientific Reports*, *9*, 3581.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics (Oxford, England)*, *27*, 2194–2200.

Fišer, C., Robinson, C. T., & Malard, F. (2018). Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology*, *27*, 613–635.

Forootan, A., Sjöback, R., Björkman, J., Sjögreen, B., Linz, L., & Kubista, M. (2017). Methods to determine limit of detection and limit of quantification in quantitative real-time PCR (qPCR). *Biomolecular Detection and Quantification*, *12*, 1–6.

Gilbey, J., Carvalho, G., Castilho, R., Coscia, I., Coulson, M. W., Dahle, G., Derycke, S. et al (2021). Life in a drop: Sampling environmental DNA for marine fishery management and ecosystem monitoring. *Marine Policy*, *124*, 104331.

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

Kang, D. D., Froula, J., Egan, R., & Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*, *3*, e1165.

Kissling, W. D., Ahumada, J. A., Bowser, A., Fernandez, M., Fernández, N., García, E. A., Guralnick, R. P., Isaac, N. J. B., Kelling, S., Los, W., McRae, L., Mihoub, J.-B., Obst, M., Santamaria, M., Skidmore, A. K., Williams, K. J., Agosti, D., Amariles, D., Arvanitidis, C., & Hardisty, A. R. (2018). Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. *Biological Reviews*, *93*, 600–625.

Klymus, K. E., Merkes, C. M., Allison, M. J., Goldberg, C. S., Helbing, C. C., Hunter, M. E., Jackson, C. A., Lance, R. F., Mangan, A. M., Monroe, E. M., Piaggio, A. J., Stokdyk, J. P., Wilson, C. C., & Richter, C. A. (2020). Reporting the limits of detection and quantification for environmental DNA assays. *Environmental DNA*, *2*, 271–282.

Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*, 1674–1676.

Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2014). Swarm: Robust and fast clustering method for amplicon-based studies. *PeerJ*, *2*, e593.

Marden, E., Abbott, R. J., Austerlitz, F., Ortiz-Barrientos, D., Baucom, R. S., Bongaerts, P., Bonin, A. et al (2021). Sharing and reporting benefits from biodiversity research. *Molecular Ecology*, *30*, 1103–1107.

Mariani, S., Baillie, C., Colosimo, G., & Riesgo, A. (2019). Sponges as natural environmental DNA samplers. *Current Biology*, *29*, R401–R402.

Martel, C. M., Sutter, M., Dorazio, R. M., & Kinziger, A. P. (2020). Using environmental DNA and occupancy modelling to estimate range-wide metapopulation dynamics. *Molecular Ecology*, https://doi.org/10.1111/mec.15693

McClenaghan, B., Compson, Z. G., & Hajibabaei, M. (2020). Validating metabarcoding-based biodiversity assessments with multi-species occupancy models: A case study using coastal marine eDNA. *PLoS One*, *15*, e0224119.

Pawlowski, J., Apothéloz-Perret-Gentil, L., & Altermatt, F. (2020). Environmental DNA: What's behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring. *Molecular Ecology*, *29*, 4258–4264.

Pearman, J. K., Chust, G., Aylagas, E., Villarino, E., Watson, J. R., Chenuil, A., Borja, A., Cahill, A. E., Carugati, L., Danovaro, R., David, R., Irigoien, X., Mendibil, I., Moncheva, S., Rodríguez-Ezpeleta, N., Uyarra, M. C., & Carvalho, S. (2020). Pan-regional marine benthic cryptobiome biodiversity patterns revealed by metabarcoding Autonomous Reef Monitoring Structures. *Molecular Ecology*, *29*, 4882–4897.

Peng, Y., Leung, H. C., Yiu, S. M., & Chin, F. Y. (2011). Meta-IDBA: A de Novo assembler for metagenomic data. *Bioinformatics*, *27*, i94–101.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., & Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*, 7537.

Scholin, C. A, Brich, J., Jensen, S., Marin III, R., Massion, E., Pargett, D., Preston, C., Roman, B., & Ussler III, W. (2017). The quest to develop ecogenomic sensors: A 25-year history of the Environmental Sample Processor (ESP) as a case study. *Oceanography*, *30*, 110–113.

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

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology*, *21*, 17891793.

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*, e29715.

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.

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. et al (2020). Chapter Nine - Advances and prospects of environmental DNA in neotropical rainforests. In A. J. Dumbrell, E. C. Turner, & T. M. Fayle (Eds.), *Advances in ecological research* (pp. 331–373). Academic Press.