

Ichthyoplankton DNA metabarcoding: challenges and perspectives

Daniel Carvalho¹

¹Pontifícia Universidade Católica de Minas Gerais

November 5, 2021

Abstract

DNA metabarcoding has been widely used to access and monitor species. However, several challenges remain open for its mainstream application in ecological studies, particularly when dealing with a quantitative approach. In a from the Cover article in this issue of *Molecular Ecology*, Cédric et al. (2021) report species-level ichthyoplankton dynamics for 97 fish species from two Amazon river basins using a clever quantitative metabarcoding approach employing a probe capture method. They clearly show that most species spawned during the rainy season when the floods started, but interestingly, species from the same genus reproduced in distinct periods (i.e., inverse phenology). Opportunistically, Cédric et al. (2021) reported that during an intense hydrological anomaly, several species had a sharp reduction in spawning activity, demonstrating a quick response to environmental cues. This is an interesting result since the speed at which fish species can react to environmental changes, during the spawning period, is largely unknown. Thus, this study brings remarkable insights into basic life history information that is imperative for proposing strategies that could lead to a realistic framework for sustainable fisheries management practices and conservation, fundamental for an under-studied and threatened realm, such as the Amazon River basin.

Article type: News and Views Perspectives

Ichthyoplankton DNA metabarcoding: challenges and perspectives

Daniel Cardoso Carvalho

Conservation Genetics Lab, Postgraduate Program in Vertebrate Biology, Pontifical Catholic University of Minas Gerais (PUC Minas), Belo Horizonte, Brazil. e-mail: carvalho.lgc@gmail.com

ORCID

<https://orcid.org/0000-0003-4517-5069>

Funding information

CNPq Fellowship Award number 306155/2018-4.

Keywords

Environmental genomics, Biomonitoring 2.0, DNA barcode, High-throughput DNA sequencing, Neotropical fish,

DNA metabarcoding has been widely used to access and monitor species. However, several challenges remain open for its mainstream application in ecological studies, particularly when dealing with a quantitative approach. In a from the Cover article in this issue of *Molecular Ecology*, Cédric et al. (2021) report species-level ichthyoplankton dynamics for 97 fish species from two Amazon river basins using a clever quantitative metabarcoding approach employing a probe capture method. They clearly show that most species spawned during the rainy season when the floods started, but interestingly, species from the same genus reproduced in distinct

periods (i.e., inverse phenology). Opportunistically, Cédric et al. (2021) reported that during an intense hydrological anomaly, several species had a sharp reduction in spawning activity, demonstrating a quick response to environmental cues. This is an interesting result since the speed at which fish species can react to environmental changes, during the spawning period, is largely unknown. Thus, this study brings remarkable insights into basic life history information that is imperative for proposing strategies that could lead to a realistic framework for sustainable fisheries management practices and conservation, fundamental for an understudied and threatened realm, such as the Amazon River basin .

The use of novel molecular tools to access and monitor biodiversity has been revolutionized by the introduction of high-throughput sequencing (HTS) platforms that allow the analysis not only of individual specimens but of whole communities (Leese et al. 2018). The use of HTS coupled with DNA barcoding enables the identification of multiple species from environmental bulk samples, commonly termed DNA metabarcoding. Such DNA based next generation biomonitoring, or “Biomonitoring 2.0” (Baird and Hajibabaei, 2012), has a tremendous potential to advance the traditional protocols applied to assess and monitor the environment (Leese et al. 2018), ensuring a rapid and cost-efficient biodiversity assessment. This is particularly important when dealing with a large number of organisms, as is common with ichthyoplankton samples.

Despite the large potential of DNA metabarcoding for analyzing bulk samples, technical issues that may hinder full mainstream Biomonitoring 2.0 practices are still under debate, particularly regarding quantitative analysis (Lamb et al. 2018, Deagle et al. 2018). The potential quantitative value of DNA metabarcoding has been questioned, mainly due to the difficulty of obtaining sound and unbiased data. However, it is controversial whether or not the proportions of reads generated from DNA metabarcoding studies correspond to the real proportions of species in a community (Lamb et al. 2018). Constraints affecting the quantitative performance of metabarcoding are due to bias introduced during the four main stages of its workflow: (1) sample collection and processing, (2) choice of target gene, (3) HTS DNA sequencing and (4) bioinformatics (Figure 1). Numerous approaches have been implemented to mitigate bias during each stage in order to avoid constraints that hinder accurate quali-quantitative DNA metabarcoding analysis (Thomas et al. 2016, Deagle et al. 2018; Ratcliffe et al. 2021).

Mariac et al. (2021) addresses well the bias that may arise from ichthyoplankton sampling and the unequal body size of fish larvae that may lead to differences in cell and DNA quantity from each species/specimen per bulk sample. Since, when dealing with ichthyoplankton, it is possible to estimate the number of larvae per second drifting through the river using the total larval flow (TLF), so the authors applied TLF as a correction factor for species abundance. Also, the authors were able to size fractionated specimens and extract DNA from individuals of similar size to minimize bias due to different initial amounts of DNA from each larva. Similarly, Ratcliffe et al. (2021) used a standardized amount of tissue for each species to improve quantification accuracy. But instead, they used PCR and conserved primer binding sites to amplify the mitochondrial 12S gene and avoid bias due to differential amplification rates. Further adjustment of read counts (i.e. relative read abundance estimates - RRA) implemented in their bioinformatics pipeline was capable of generating more accurate DNA metabarcoding quantitative estimations. Thomas et al. (2016), also using a PCR DNA metabarcoding approach, dealt with read abundance biases by applying the relative correction factor (RCF) obtained from HTS sequencing of 50/50 mixtures of target and control species. By applying the RCF, they were able to correct the majority of species-specific biases from control and field samples and improve the relationship between RRA and mass percentage of each taxon.

Most DNA metabarcoding bias may be previously diagnosed and adjusted using mock communities composed of known species composition. The analyses of such mock samples are paramount to validate DNA metabarcoding accuracy to detect and quantify species, and therefore, are recommended to be conducted prior to analyzing bulk samples of unknown species composition obtained from field samples (Duke & Burton, 2020). Following that, before analyzing ichthyoplankton sampled from the environment, Mariac et al. (2018) previously demonstrated that metabarcoding by capture using a single COI probe was able to identify and quantify fish species in ichthyoplankton swarms from the Amazon realm using mock communities.

Nonetheless, powerful ecological inferences from metabarcoding studies are strongly dependent on reference DNA databases that may allow the identification of reads to the lowest taxonomic level possible. While the COI gene database is one of the most complete for molecular assignment of fish species, it is still incomplete for several groups, particularly in high biodiversity realms such as the Amazon. Moreover, the use of multiple markers could increase sensitivity of species detection due to mismatches between probes/primers (Duke & Burton, 2020). With the decrease of DNA sequencing costs, a possible good cost/benefit solution would be "genome skimming" of voucher tissue samples and assembling whole mitogenomes for the development of a broader reference database. Such a mitogenome database would allow the analysis of multiple markers and primers development for metabarcoding communities from specific river basins (Milan et al. 2020). However, sequencing several mitogenomes per species to generate intraspecific diversity, important for detecting highly conserved primer sites, might severely increase costs. Despite this, examples of Neotropical fish species with multiple mitogenomes are already available (Santos et al. 2021).

In perspective, DNA metabarcoding (PCR free or not) may allow accurate estimation of ecological information, since there is a great amount of evidence that adjustments can overcome known constraints (Figure 1). Thus, with the continuous refinements of DNA metabarcoding methodology, I foresee its mainstream use to monitor and assess biodiversity allowing rapid and relatively inexpensive processing of complex bulk and environmental samples for several taxa and highly diverse realms.

Figure Caption

Figure 1. DNA metabarcoding constraints and proposed solutions.

References

- Baird DJ, Hajibabaei M (2012) Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology* ,**21** , 2039-2044.
- Duke EM, Burton RS (2020) Efficacy of metabarcoding for identification of fish eggs evaluated with mock communities. *Ecology and evolution* ,**10** (7), 3463-3476.
- Deagle BE, Thomas AC, McInnes JC *et al.* (2018) Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? *Molecular Ecology* , **28** (2), 391-406.
- Lamb PD, Hunter E, Pinnegar JK, *et al.* (2019) How quantitative is metabarcoding: A meta-analytical approach. *Molecular Ecology* , **28** (2), 420-430.
- Leese F, Bouchez A, Abarenkov K, *et al.* (2018) Chapter Two – Why We Need Sustainable Networks Bridging Countries, Disciplines, Cultures and Generations for Aquatic Biomonitoring 2.0: A Perspective Derived From the DNAqua-Net COST Action, in: Bohan DA, Dumbrell AJ, Woodward G, Jackson M (Eds.), *Advances in Ecological Research, Next Generation Biomonitoring: Part 1* . Academic Press, pp. 63–99.
- Mariac C, Renno JF, Carmen Garcia-Davila *et al.* (2021) Species-level ichthyoplankton dynamics for 97 fishes in two major river basins of the Amazon using quantitative metabarcoding. *Molecular ecology*, 30, XXX-XXX.
- Mariac C, Vigouroux Y, Duponchelle F *et al.* (2018) Metabarcoding by capture using a single COI probe (MCSP) to identify and quantify fish species in ichthyoplankton swarms. *PLoS One* , 13(9), e0202976.
- Milan DT, Mendes IS, Damasceno JS *et al.* (2020) New 12S metabarcoding primers for enhanced Neotropical freshwater fish biodiversity assessment. *Scientific reports* , **10** , 17966.
- Ratcliffe FC, Uren Webster TM, Rodriguez-Barreto D *et al.* (2021) Quantitative assessment of fish larvae community composition in spawning areas using metabarcoding of bulk samples. *Ecological Applications* ,**31** (3), e02284.

Santos RP, Melo BF, Yazbeck GM *et al.* (2021). Diversification of *Prochilodus* in the eastern Brazilian Shield: Evidence from complete mitochondrial genomes (Teleostei, Prochilodontidae). *Journal of Zoological Systematics and Evolutionary Research* , **59** (5), 1053-1063.

Thomas AC, Deagle BE, Eveson JP, Harsch CH, & Trites AW (2016) Quantitative DNA metabarcoding: improved estimates of species proportional biomass using correction factors derived from control material. *Molecular ecology resources* , **16** (3), 714-726.

