



**HAL**  
open science

## Opportunities and inherent limits of using environmental DNA for population genetics

Marjorie Couton, Frédérique Viard, Florian Altermatt

► **To cite this version:**

Marjorie Couton, Frédérique Viard, Florian Altermatt. Opportunities and inherent limits of using environmental DNA for population genetics. *Environmental DNA*, inPress, 10.1002/edn3.448 . hal-04179982

**HAL Id: hal-04179982**

**<https://hal.science/hal-04179982>**

Submitted on 10 Aug 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Opportunities and inherent limits of using environmental DNA for population genetics

Marjorie Couton<sup>1</sup>  | Frédérique Viard<sup>2</sup>  | Florian Altermatt<sup>1,3</sup> 

<sup>1</sup>Eawag: Swiss Federal Institute of Aquatic Science and Technology, Department of Aquatic Ecology, Dübendorf, Switzerland

<sup>2</sup>ISEM, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France

<sup>3</sup>Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zürich, Switzerland

## Correspondence

Marjorie Couton and Florian Altermatt, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Department of Aquatic Ecology, Überlandstrasse 133, Dübendorf CH-8600, Switzerland. Email: [marjorie.couton@eawag.ch](mailto:marjorie.couton@eawag.ch) and [florian.altermatt@ieu.uzh.ch](mailto:florian.altermatt@ieu.uzh.ch)

## Funding information

Agence Nationale de la Recherche, Grant/Award Number: ANR-16-IDEX-0006; Bundesamt für Umwelt; Fachverband für Wasser-, Gas- und Fernwärmeversorger; Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 31003A\_173074; TOTAL Foundation; Universität Zürich

## Abstract

Molecular techniques using DNA retrieved from community or environmental samples, in particular environmental DNA (eDNA), are becoming increasingly popular for detecting individual species, assessing biodiversity, and quantifying ecological indices. More recently, eDNA has also been proposed as a template for population genetics, and several studies have already tested the feasibility of this approach, mostly looking at vertebrate species. Their results along with general opportunities offered by these types of “community-based” samples, such as the possibility to target multiple species at the same time, have generated great enthusiasm and expectations for using eDNA in population genetics. However, not every aspect of population genetics can be addressed by eDNA-based data and some inherent limitations may challenge its conclusions. Here, we firstly review the state of current knowledge of DNA retrieved from environmental and community samples for population genetics. Then, focusing on eDNA, we summarize the opportunities but also detail four main limitations of its use for population-level inferences, namely, (1) the difficulty to retrieve a species-specific dataset, (2) the potential lack of correlation between observed and true allelic frequencies, (3) the loss of individual information in multi-locus genotyping and linkage between loci, and (4) the uncertainty about the individuals contributing to the sampled DNA pool (e.g., number, life-stage, or sex). Some of these limitations might be overcome with the development of new technologies or models that account for the specificities of eDNA. Others, however, are inherent, and their effect on the inferences must be thoroughly evaluated. The possibility of gaining insights into genetic diversity and population structure from DNA retrieved from community and environmental samples is appealing for scientists, conservation managers, and other practitioners. Yet, to avoid false expectations and incorrect inferences, it is imperative that these limitations are known and considered alongside the opportunities and advantages.

## KEYWORDS

conservation of natural resources, DNA barcoding, gene flow, genetic diversity, genetic markers, genetic models, phylogeography, population statistics

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. *Environmental DNA* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

The field of population genetics examines the distribution of the genetic diversity within species or species complexes, to investigate major eco-evolutionary processes such as migration (i.e., population connectivity), adaptation and speciation. Before the 1970s, population genetics was largely theoretical and mostly restricted to a few model organisms. Only few observations on phenotypes could provide data to validate the existing models (e.g., Charlesworth & Charlesworth, 2017). With the development of the study of allozymes in the 1970s, followed by DNA-based methods in the 1980s, the amount of data available to study genetic variation increased dramatically (Allendorf, 2017), and empirical studies rose at an unprecedented pace. The field of population genetics grew from oligo-marker approaches (e.g., a dozen of microsatellites) to multi-marker approaches (e.g., thousands of SNPs from Rad-sequencing and whole-genome sequencing), allowing to refine analyses and address new questions (Gagnaire, 2020). This set the scene to an emerging field named “population genomics” that appeared in the late 1990s for humans and early 2000s for non-model organisms (Allendorf, 2017; Luikart et al., 2003).

Building on the above-mentioned developments, molecular tools were also applied in the field of ecology, such as DNA barcoding for species identification (Hebert, Cywinska, et al., 2003), qPCR for species detection (Jerde et al., 2011), or DNA metabarcoding for describing communities (Cristescu, 2014). These new approaches have been increasingly applied to DNA retrieved from community sampling (i.e., collection of unknown specimens collected in the environment, for example bulkDNA; see Box 1) and, more particularly, to environmental DNA (eDNA) contained in environmental samples (e.g., from water or soil). eDNA includes free DNA in the medium, DNA encapsulated in organelles (e.g., mitochondria or chloroplasts) or in whole cells, either detached from an organism, or as microorganisms (Pawlowski et al., 2020; Taberlet, Coissac, Hajibabaei, & Rieseberg, 2012). Sampling directly from the environment is usually easier than catching organisms and is non-invasive. This renders eDNA very attractive for environments or organisms otherwise difficult to sample, such as many species in freshwater or marine habitats, or for threatened taxa of conservational concern. Moreover, eDNA is often less costly than traditional methods, especially when in need of examining a high number of samples or sites (Qu & Stewart, 2019; Smart et al., 2016). As part of a targeted approach to detect one or a few species, DNA retrieved from community samples or from the environment can be processed with quantitative or digital PCR methods, using specific probes or primers (e.g., Valsecchi et al., 2022). It can also serve as a matrix for analyzing a broader set of taxa, using metabarcoding of mitochondrial and/or nuclear genes (e.g., Couton et al., 2019; Olds et al., 2016).

DNA retrieved from environmental or community samples is already extensively used for a wide range of ecological applications (Deiner et al., 2017; Leray & Knowlton, 2016; Taberlet, Coissac, Hajibabaei, & Rieseberg, 2012; Taberlet, Coissac, Pompanon, et al., 2012). These new DNA templates (Lacoursière-Roussel &

Deiner, 2021) have been broadly applied in community ecology and functional ecology to answer questions regarding biodiversity patterns (e.g., Blackman et al., 2021; Lacoursière-Roussel et al., 2018; Zenker et al., 2020; Zinger et al., 2019), species' range dynamics (e.g., Hobbs et al., 2019; Valsecchi et al., 2022), spatio-temporal turn-over in species composition (e.g., Bálint et al., 2018; Bista et al., 2017; Chain et al., 2016), or trophic networks (e.g., Banerjee et al., 2022; Blackman et al., 2022; Carroll et al., 2019). Additionally, eDNA has been used in conservation biology, for instance for the monitoring of rare or endangered species (e.g., Burgoa et al., 2020; Niemiller et al., 2018), or for the detection of invasive species (e.g., Couton et al., 2022; Zaiko et al., 2018), as well as for the calculation of biotic indices (e.g., Brantschen et al., 2021; Pawlowski et al., 2018). The opportunity offered by the analyses of DNA retrieved from environmental or community samples to get insights on molecular diversity for a wide range of taxa also prompted the interest of researchers to explore its use at the intraspecific level (Adams et al., 2019; Barnes & Turner, 2016; Sigsgaard et al., 2020).

For a population geneticist, the prospect of using only one community or one environmental sample to get genetic information on many individuals is very appealing. The enthusiasm is even bigger if this one sample can provide data on many species, across several markers, and is non-destructive as with eDNA. The use of eDNA for population genetics in animals was first suggested by Barnes and Turner (2016), building on what was already done using non-invasive samples, such as feces and fur (Taberlet et al., 1999), and studies on human genetic variants from river water (Kapoor et al., 2014), or swabs from artificial surfaces across the New York subway (Afshinnekoo et al., 2015). The first attempts to study population genetics using community-based samples were performed in 2017, either using eDNA on whale sharks (Sigsgaard et al., 2017) or on the common carp (Uchii et al., 2017), or using bulk samples of arthropods (Pedro et al., 2017). The former study recovered successfully known mitochondrial haplotypes of whale sharks from seawater samples in similar frequencies as observed with tissue samples, the second was able to identify different seasonal patterns between a native and a non-native population of common carp using the detection of SNPs by probe capture, and the latter identified genetic structure between two populations of beetles living in coffee farms with different management practices. These pioneering works inspired other studies testing various approaches (e.g., SNP detection, microsatellite amplification, High-Throughput Sequencing (HTS) of amplicons) with different markers, either nuclear or mitochondrial, and on various taxa from macroinvertebrates to marine mammals (e.g., Andres et al., 2021; Elbrecht et al., 2018; Marshall & Stepien, 2019). Concomitantly, several papers discussed eDNA's opportunities and limitations for population level inferences (Adams et al., 2019; Goldberg & Parsley, 2021; Sigsgaard et al., 2020). It was suggested that future progress, notably technical, would overcome some of the current limitations. While we understand and share the legitimate optimism related to novel opportunities, we also

### BOX 1 The different types of samples that can be used in population genetics.

Besides tissue samples directly collected from the target organisms, a wide range of samples exists, from which molecular data for population genetics can be collected (cf Table below).

**List of the different types of DNA samples that can be used in population genetics studies with their definition and examples of publications where they are used.**

Name	Description	Publications
Tissue	DNA extracted from tissue samples of one individual	NA <sup>a</sup>
Non-invasive	DNA collected from cells left in the environment by a specific individual (e.g., feces, fur, saliva, skin cells)	Monge et al. (2020); Parsons et al. (2018)
bulkDNA	DNA extracted from a bulk of targeted organisms; targets can be either a given taxonomic group (e.g., insects with Malaise traps), life stage (e.g., planktonic larvae from plankton net), functional groups (e.g., benthic organisms sampled on passive sampling devices)	Couton et al. (2019); Elbrecht et al. (2018); Turon et al. (2020)
Diet DNA	DNA extract from the gut content of an organism	Bohmann et al. (2018); Sethi et al. (2019)
iDNA	DNA from vertebrates extracted from the gut content of an invertebrate parasite	Meekan et al. (2017); Nguyen et al. (2021)
ebDNA	Ethanol-based DNA: DNA extracted directly from the preservative ethanol in which the organism(s) were stored	Couton et al. (2021); Linard et al. (2016)
eDNA	DNA extracted from an environmental sample, including free DNA encapsulated in an organelle, in a whole cell or even as a full organism (e.g., protist, fertilized eggs)	Tsuji et al. (2020); Weitemier et al. (2021)

<sup>a</sup>No examples are cited here since tissue is the conventional type of sample used in population genetics and thousands of studies could be cited as relevant examples. These samples have distinct particularities, that make them suited to answer different questions, and their inherent limitations can challenge their use in some cases. Choosing the right type of sample for a particular study will thus depend on one's needs (e.g., in relation to sampling constraints or to the type of subsequent molecular analysis), and on the sample's limitations (cf. Figure 5). We also believe that the confusion between these different types of samples often referred to as eDNA (except for "Tissue") increases the complexity to define clear opportunities but also limitations for the use of eDNA in population genetics, and might create false expectations. For example, the lack of a link between an haplotype and a specific individual, one clear limitation of eDNA, is actually not an issue when sampling in the wake of whales as in Dugal et al., 2022 (Figure 6). We propose here to refer to this latter approach as non-invasive sampling, despite it being labeled eDNA in the original publication.

believe that some limitations of eDNA may not be overcome, and thus impact the spectrum of applications that can be achieved for population studies.

Here, we evaluate the possibilities offered by the analysis of DNA retrieved from environmental or community samples (i.e., samples containing multiple species and for which the composition is unknown), focusing on eDNA (here strictly defined as DNA obtained from environmental samples such as water, soil, or sediments; Box 1) for population genetics. Starting from the existing literature, we list current opportunities, but also highlight limitations to the use of such samples. Our perspective focuses on population inferences for multi-cellular eukaryotes, typical targets of conventional population genetics studies, but also discusses some aspects pertinent for other types of samples or taxa. We are convinced that, when developing new approaches, both the possibilities and fundamental limits must be part of the discussion for a forward-thinking implementation and a credible coverage of the research field. We propose here, for the first time, a fair assessment and a detailed discussion of both

opportunities but also limitations of eDNA in the context of population genetics illustrated with case studies. We hope our evaluation will build a solid and transparent basis for all future work on eDNA and population genetics.

## 2 | LITERATURE REVIEW

We performed a literature search on the use of eDNA and other community samples for population genetics analyses using the online database Web of Science (Clarivate Analytics) on April 24th 2023. We retrieved all studies with the following words in their title or keywords: ("environmental DNA" OR "eDNA" OR "environmental samples" OR "bulk" OR "ethanol" OR "preservative" OR "iDNA" OR "gut content" OR "non-invasive sampl\*" OR "metabarcoding") AND ("population" OR "haplotype" OR "haplotypic" OR "intraspecific" OR "infraspecific" OR "intra-specific" OR "infra-specific" OR "variants" OR "genotype" OR "genetic diversity" OR "genotyping" OR

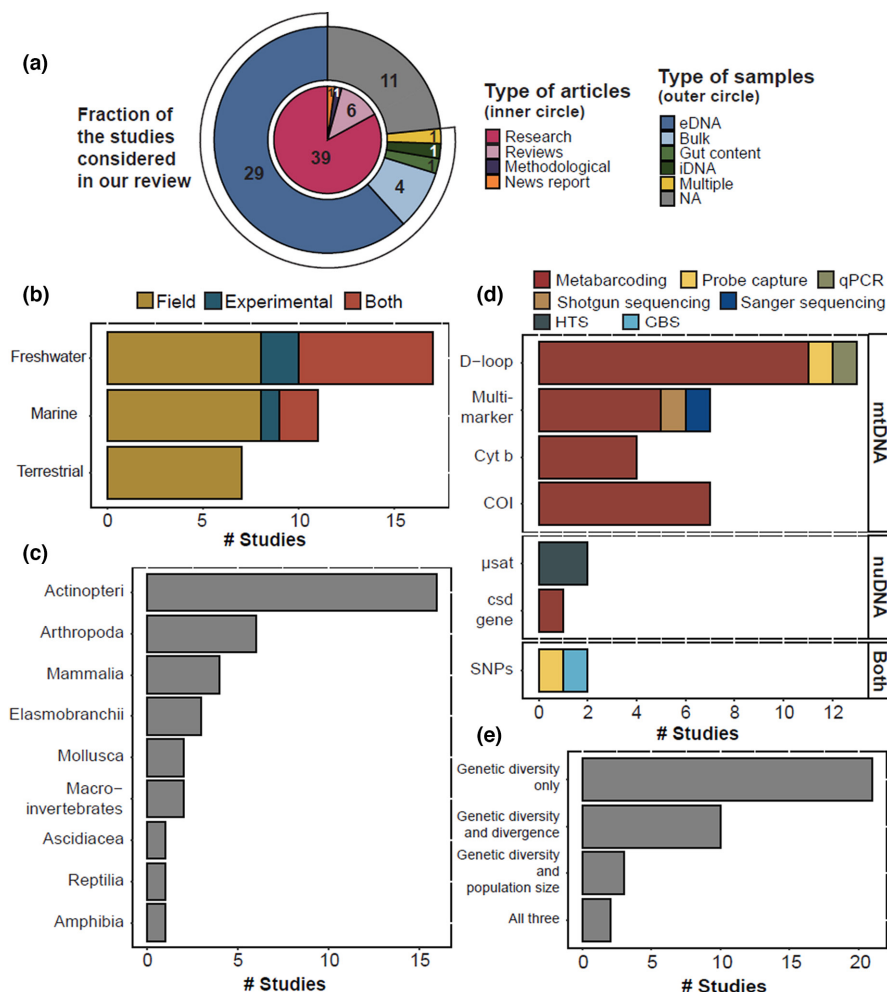
“introgression” OR “admixture” OR “population variation” OR “genetic structure” OR “phylogeography”). The same query was additionally implemented in the search tool of the journals *Environmental DNA* and *Metabarcoding and Metagenomics*, since their publications are not yet indexed in Web of Science. All retrieved publications were screened manually to evaluate their relevancy to our topic (i.e., publications targeting community or environmental samples for the study of multi-cellular organisms and using a molecular tool potentially retrieving information on several individuals at the same time).

The query returned 544 publications (Table S1). All studies retrieved were published between 1961 and 2023, and there was no publication with the required keywords prior to 1961. Among these publications, 47 publications considered the use of community samples for population genetics with a particular focus on multi-cellular eukaryotes (Figure 1a). They were all published since 2017 (Table S2). We restricted our synthesis to research articles and thus excluded six reviews, one methodological publication, and one news report. We also excluded from further analyses three research articles because they were based on computational modeling only. The remaining 36 publications use either environmental samples (29 publications), or other types of community samples (bulk samples (4), iDNA (1), gut content (1) or ethanol samples (1)). Twenty-six of the 36 publications considered (72%)

were published after the last review from Sigsgaard et al. (2020), illustrating the rapid development of the field.

From the 36 retained research articles, most target aquatic environments using water samples, either freshwater or marine (Figure 1b). Fish (Actinopteri and Elasmobranchi) are the main target group so far, with more than half of the publications focusing on 16 different species (Figure 1c). This is in line with the results of the review performed by Tsuji et al. (2019) on eDNA methods, in which 45% of publications target fish taxa. The second most targeted group are macroinvertebrates (Arthropoda, Ascidiacea, and Mollusca), mostly using bulk samples. Indeed, these organisms are often the focal point of eDNA metabarcoding (e.g., Keck et al., 2022). Additionally, four studies focus on a mammalian species, one on a reptile, and one on an amphibian.

Across all studies, several approaches and markers are used (Figure 1d). However, the majority of publications focuses solely on mitochondrial DNA (mtDNA). Since a high number of studies target fish, the mitochondrial control region (D-loop), known to have a high intraspecific resolution for these taxa (Ardura et al., 2013), is the preferred marker of our listed articles, followed by COI. The latter has been suggested as the marker of choice for (meta)barcoding studies (Hebert, Ratnasingham, & de Waard, 2003; Leray et al., 2013), but is also commonly used for intraspecific studies (e.g.,



**FIGURE 1** (a) Distribution of the 47 publications considering the use of community samples for population genetics according to their article type (inner circle) and the type of sample analyzed (outer circle). The number of related publications is given inside each section. The studies included in the literature review are then represented according to (b) the types of environments examined (the studies can be either field studies, experimental studies, or both), (c) the classes of taxa, (d) the markers and technical approaches (COI, cytochrome oxidase I (mitochondrial); Cyt. b, cytochrome b (mitochondrial); GBS, Genotyping By Sequencing; nuDNA, nuclear DNA; SNPs, Single Nucleotide Polymorphisms;  $\mu$ sat, microsatellites); and (e) the type of metrics computed.

phylogeography: Jenkins et al., 2018). Only five publications investigate nuclear markers, including microsatellites and sequencing the complementary sex determiner (*csd*) gene or single nucleotide polymorphisms (SNPs). Regarding the methods applied, the favored approach is the amplification of a DNA fragment followed by High Throughput Sequencing (HTS), as is done for metabarcoding. Eight studies, however, test other techniques such as shotgun sequencing or probe capture (Figure 1d).

Since the use of community or environmental samples for population level inferences is a recent endeavor, most studies (21; 58%) aim at evaluating the possibility to retrieve intraspecific data from various sample types, for different groups of organisms, and in various contexts. They mostly evaluate the adequacy of the results obtained, either by comparing them directly with data from individual tissue samples, or by confronting them with previous knowledge on the target taxa. All listed articles describe the genetic diversity of the studied populations, most of them only using haplotypic or allelic richness indices (Figure 1e). Only 15 studies (42%) provide other metrics to describe genetic divergence between populations, using for example a fixation index ( $F_{ST}$ ), and/or to identify the number of individuals either contributing to the sampled genetic pool, or reproducing in the total population.

Across the 36 publications encompassing different types of environmental and community samples, most (81%) are examining eDNA. In the following sections, we will focus on eDNA strictly speaking (i.e., DNA retrieved from environmental samples), unless otherwise specified.

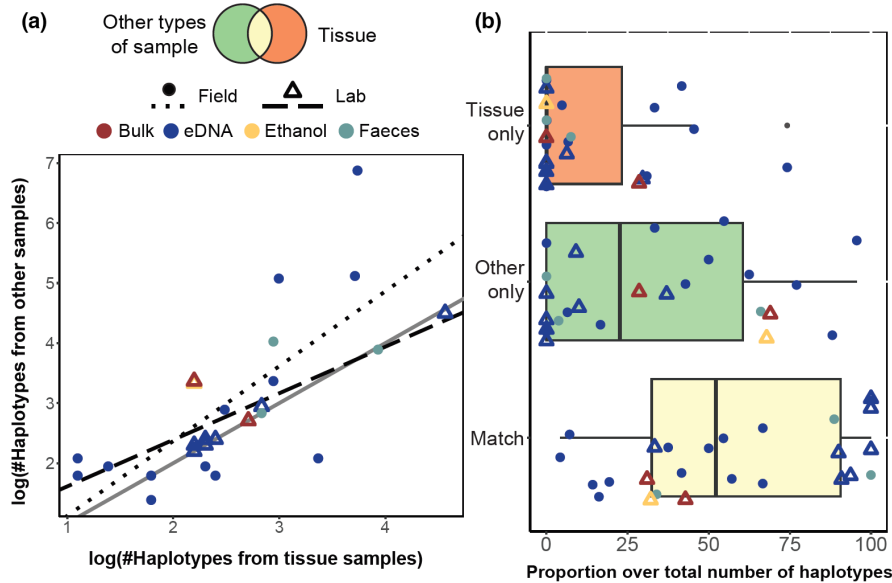
### 3 | THE VALUE OF eDNA FOR POPULATION GENETICS

The results obtained from the 36 publications reviewed (as described above) are promising, with values of metrics computed from DNA retrieved from community or environmental samples mostly corresponding to those obtained from tissue samples, when available. For example, Marshall and Stepien (2019) revealed a strong correlation between the number of unique sequences (i.e., amplicon sequence variants or ASVs) obtained with HTS from water samples and the number of expected haplotypes for Quagga and Zebra mussels in tank communities ( $0.77 < R^2 < 0.97$ ). They were also able to retrieve all expected haplotypes. Similarly, Andres et al. (2021) found a strong correlation between eDNA and tissue samples derived allelic frequencies from 26 microsatellite loci, both from a mesocosm experiment ( $r=0.95$ ) and from field samples ( $r=0.84$ ). Some concerns were raised, however, about both the overestimation of haplotypes/alleles richness due to technical errors (especially sequencing errors when using HTS-based approaches, for example, Couton et al., 2021; Tsuji et al., 2020), and underestimation due to detection failure of rare haplotypes/alleles (e.g., Adams et al., 2022). Both limitations were also observed in studies targeting other types of samples such as preservative ethanol (Couton et al., 2021) or bulk samples (Elbrecht et al., 2018).

To quantify this general observation, we retrieved the number of haplotypes identified with either community or environmental samples (i.e., named “community-based” in the following text and figures) or tissue methods from all publications reporting a direct comparison between these approaches (Table S3). For sake of comparison, we also included three studies comparing DNA extracted from tissue samples with DNA retrieved from samples collected in the environment but targeting a given individual (i.e., non-invasive DNA, see Box 1), such as DNA extracted from feces. One may indeed expect a better correlation with this type of individual-based sampling as compared to community-based sampling. In total, 17 studies reported 26 comparisons between environmental and tissue samples, all using an amplification of a mitochondrial marker followed by HTS, and all targeting aquatic organisms. In analogy to the main review, the vast majorities of comparisons are performed on fish DNA and target the D-loop marker. Across studies, the number of haplotypes retrieved with both methods is highly correlated (Pearson correlation on log transformed data:  $r=0.723$ ,  $p<0.001$ ), and this relationship holds for both laboratory-based experiments ( $r=0.784$ ,  $p=0.007$ ) and field experiments ( $r=0.749$ ,  $p=0.003$ ; Figure 2a). The slopes between the two linear regression for the two types of experiments are not significantly different (ANOVA comparing two linear models with fixed and random slopes;  $F_1=0.909$ ,  $p=0.351$ ; see Table S4 for more details), which is probably due to the low number of points included. However, the correlation between tissue and other types of samples seems stronger for lab experiments since most points are very close to the 1:1 line. The number of detected haplotypes shared by both methods is highly variable depending on the study, with only 4% for the least congruent comparison to up to 100% for some others (mean = 57%; Figure 2b). Almost half of the comparisons (11; 42%) failed to detect haplotypes with community-based methods, whereas 16 of them detected haplotypes exclusively found with community-based samples, sometimes representing up to 95% of the total number of haplotypes identified. Although the number of comparison is too scarce for statistical analyses with individual-based samples ( $N=3$ ), it seems that the overall pattern is similar. These promising results are encouraging but also advocate for more extensive testing on other organisms, markers and environments, and the need for fine-tuning of techniques and tools. Potential mechanisms explaining the differences between the two types of approaches are discussed in the next sections.

One advantage of eDNA, as well as from DNA retrieved from community samples, is the potential to provide intraspecific genetic insights across many species from the same sample. Comparative genetic diversity assessments across multiple taxa are particularly difficult with traditional sampling methods, because the time and cost required to process the necessary amount of individuals is proportional to the number of targeted species. The use of HTS-based eDNA analyses could thus allow testing existing theories, notably those linking community and population levels, for instance by examining species-genetic diversity correlation (SGDC; Vellend et al., 2014). At present, with traditional sampling methods, these questions can only be addressed through a limited number of





**FIGURE 2** Synthesis of 17 studies reporting a direct comparison between tissue samples and other types of samples (eDNA, bulk, feces or ethanol samples) for the study of intraspecific diversity. (a) Number of haplotypes identified with tissue samples as a function of the number of haplotypes detected with eDNA and other types of samples. Both variables are log-transformed. Dots represent comparisons performed in the field whereas triangles are for comparisons in controlled laboratory settings. The dotted and dashed lines represents the linear regression for field ( $y = 1.247x - 0.125$ ) and lab ( $y = 0.776x + 0.836$ ) studies respectively. The gray line indicates a 1:1 correlation. (b) Proportion of haplotypes identified in tissue samples only, in other types of samples only, or common across both samples types used in the target study.

species sampled at the exact same time, or using meta-analyses (e.g., Corbett-Detig et al., 2015; Marchesini et al., 2018). Environmental DNA can also be of great value when studying habitats containing undescribed species. Indeed, it is not uncommon in metabarcoding studies to identify DNA fragments of unique Operational Taxonomic Units (OTUs), but from unknown taxa, or taxa not contained in genetic databases, (e.g., Li et al., 2022; Wangenstein et al., 2018; Weigand et al., 2019). In this case, genetic diversity of these unknown taxa could still be investigated through the analysis of the different unique sequences (amplicon sequence variants or ASVs) clustering into the same OTU, which would be interpreted as haplotypes belonging to the same species. The value of this approach for multi-taxa comparison has been recently documented by Antich et al. (2022). By analyzing the genetic dissimilarity between localities for 437 marine eukaryotes MOTUs, among which only 127 could be assigned to a species name, these authors could examine, with a high resolution, the congruence between biogeographic and phylogeographic patterns at the Atlantic-Mediterranean transition zone.

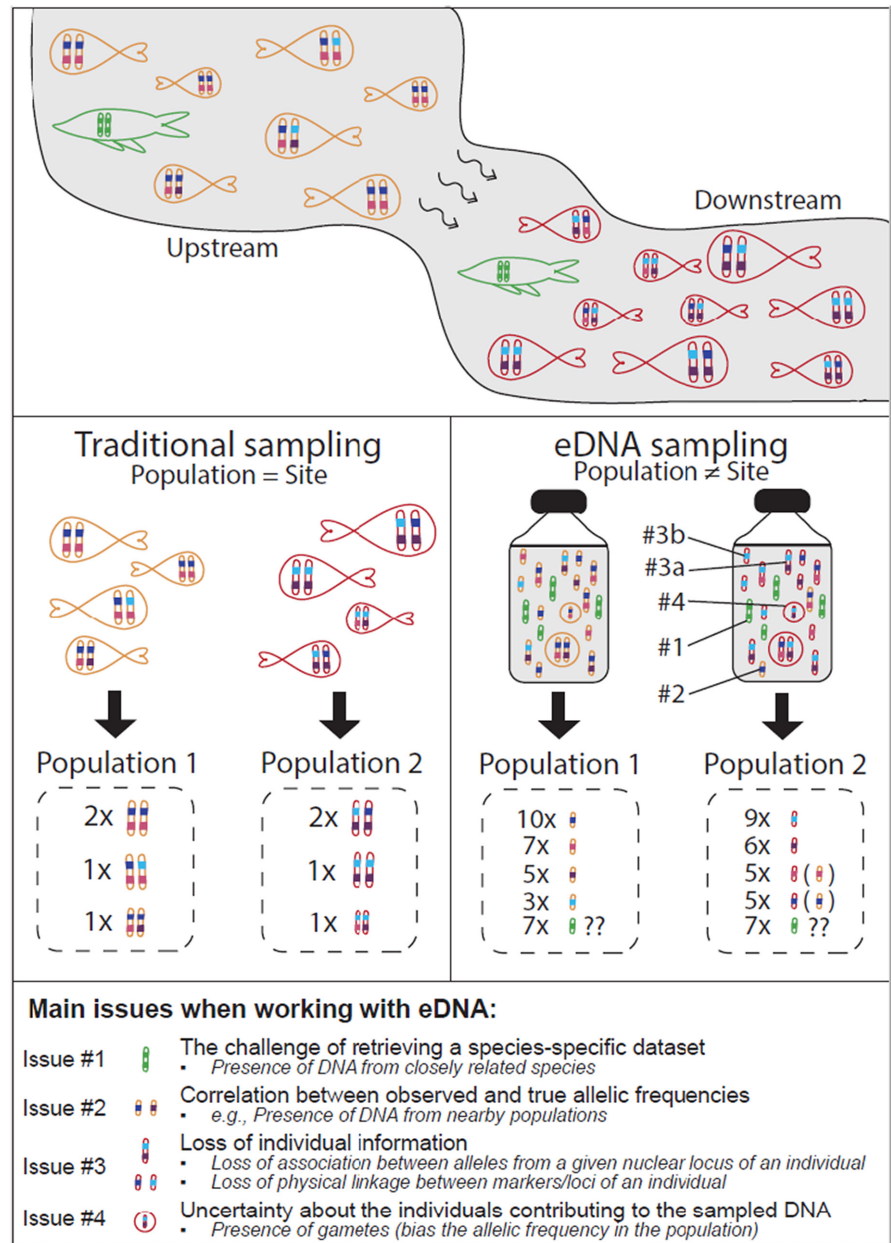
Driven by all these promising first results, the studies overall suggest that it may soon be possible to move from a proof of concept to actual use of eDNA for addressing evolutionary questions at an intraspecific level (e.g., Andres et al., 2021; Jensen et al., 2021; Macé et al., 2022). While we share some optimism, it is in our opinion crucial to also understand and respect possible—even inherent—limitations and their implications for using DNA retrieved from environmental and community samples in population genetics. In the next sections, we discuss four major limitations (Figure 3), some of them, but not all, having potential solutions (Table 1).

#### 4 | ISSUE #1: THE CHALLENGES OF RETRIEVING A SPECIES-SPECIFIC DATASET

The prerequisite to use information from DNA retrieved from environmental and community samples to answer population genetics questions is the collection of an appropriate dataset, for which data can be inarguably attributed to the target species. When using tissue samples, this crucial step can be challenged by morphological identification or by the presence of hybrids. The issue is of a different nature with environmental and community samples, as the extracted DNA is composed of a mixture from many different taxa across all major organismal groups. Ensuring that the processing protocol used retrieves all alleles from the target species, without accidentally picking up alleles from off-target species is therefore a crucial, yet non-trivial task (Figure 3, issue #1).

As listed in the above literature review, several approaches are applied to eDNA or other community-based DNAs to get intraspecific genetic data. When targeting short fragment polymorphisms, such as microsatellites or SNPs, similar alleles can be found in different species (e.g., Alther et al., 2021). For ensuring the sole amplification/capture of the target DNA, extra care and extensive testing are thus required when designing primers/probes. These primers must be comprehensive for all possible alleles of the target taxon, yet completely exclusive for alleles of any other species. In the particular case of microsatellites, for example, many studies report cases of cross-amplifications (e.g., Tardy et al., 2020; Wang et al., 2018), sometimes from different families or even different orders (Barbará et al., 2007). The design of primers thus requires testing against the

**FIGURE 3** Representation of the differences between the genetic information retrieved with either traditional individual-based or eDNA sampling in two fish populations of the same river. The main issues discussed/occurring are referred to as numbers, and explained in the bottom panels of the figure.



largest possible number of species from the same class as the target taxon, potentially living in the environment. This is a difficult or even impossible approach when the species composition of the study area is unknown. In addition, excessively increasing the specificity of probes or primers could lead to false negatives, with the non-amplification or capture of certain alleles. The use of microsatellites or SNPs with DNA retrieved from environmental or community samples should thus be restricted to focal species with no close relatives in the same habitat and for environments where the precise taxa composition is well documented. Only four publications in our review (Andres et al., 2021; Bovo et al., 2022; Jensen et al., 2021; Sethi et al., 2019) tested the detection of microsatellites or SNPs in eDNA samples, and although mentioning this shortcoming, none of them formally assessed its impact on the results. More research is thus needed to evaluate the feasibility of this approach on a broad scale.

Another approach in analyses of DNA retrieved from environmental or community samples, and the more commonly used so far, is amplicon sequencing (as for metabarcoding purposes). Data obtained are linked to the target species through taxonomic assignment by comparison with sequences deposited in reference database. Linking data to a particular taxon is thus highly dependent on the availability of reliable reference sequences, encompassing the whole range of genetic variability. The completeness and accuracy of reference databases is a well-known challenge for barcoding and metabarcoding studies, as described in Keck et al. (2023). Although some methods exist to curate reference databases and remove erroneous sequences, the extent of problematic or missing references is usually difficult to evaluate. Unlike the absence of references for a particular species, the absence of particularly divergent genetic variants of a target species



**TABLE 1** Limitations of eDNA when used for population genetics. The causes, consequences and potential solutions for the four categories of issues are detailed. The drawbacks associated to every solution are mentioned in parentheses.

Issue category	Cause	Consequences	Potential solution (and its drawbacks)
#1: Species-specific dataset	Amplification/capture of alleles from different species	Artificial inflation of genetic diversity The dataset becomes irrelevant	Design of specific primers/probes (potential loss of alleles) Extensive testing on co-occurring species (requires good knowledge on taxonomic composition)
	Assignment of an allele to the wrong species		Curation of reference databases Ensure completeness of reference databases
#2: Correlation between observed and true allelic frequencies	eDNA variations in time and space (shedding rates, degradation, transport, ...)	Over or under-representation of particular alleles Presence of false positives and false negatives	Increase number of replicates in time and space (increase time and cost) Continuous sampling Modeling (introduces uncertainty) Use presence-absence data (decrease power)
	Technical issues (amplification biases, sequencing errors, ...)		
#3: Loss of individual information	No link between alleles of an individual for diploid markers	No possibility to investigate heterozygosity and associated metrics	Use frequency-based indices (decrease power) Modeling as in pool-seq (introduces uncertainty) Use haploid markers (Not fit to address all questions)
	No link between loci of an individual	No possibility to investigate linkage disequilibrium or multi-marker approaches	Single-cell sequencing (only using the fraction of eDNA still encapsulated in a cell)
#4: Uncertainty about individuals contributing to the DNA pool	No phenotypic information (life-stage, age, sex, morphotype, ...)	Could violate assumption of including only individuals from the same generation	Adapt sampling to get only the desired phenotypes (requires good knowledge on the species life history)
	Unknown number of contributors	Could violate assumption of equal sampling effort between populations	None

is impossible to assess and might result in incomplete datasets. For instance, the well-studied invasive colonial ascidian *Botryllus schlosseri* (Pallas, 1766) is composed of five very distinct clades, characterized by COI sequence divergence ranging from 10.8% to 16.5%, (Bock et al., 2012). If references were lacking for one particular clade, and if using a standard threshold of 97% similarity clustering for COI sequences, no haplotypes belonging to this clade would be associated to the target species, thus creating false negatives.

Contrastingly, it is imperative to choose a marker suited to distinguish between the target and congeneric species, to avoid chimeric datasets and grouping haplotypes of two separate taxa. For example, if two species had a complete introgression of their mitochondrial genome, such as the marine mussel species *Mytilus trossulus* A. Gould, 1850 in the Baltic Sea, which has been introgressed by *M. edulis* Linnaeus, 1758 (Kijewski et al., 2006; Zbawicka et al., 2010), an HTS approach using only mitochondrial markers would not distinguish these two species, and may produce false positives (Couton et al., 2022). An additional source of false positives comes from the existence of erroneous sequences in public databases (Meiklejohn et al., 2019; Viard et al., 2019), which artificially inflate the genetic diversity of the target species. Retrieving intraspecific datasets from DNA collected from environmental or community samples must thus be thoroughly pondered knowing that its accuracy will largely depend on the amount and reliability of genetic information already

available for the species of interest. For example, using eDNA might be possible for studies restricted in spatial scale or in low-diversity environments, where it is achievable to collect an exhaustive baseline of the existing intraspecific diversity for the target species and closely related species beforehand. It is however important to note that it is impossible to maximize coverage of target alleles with minimizing inclusion of off-target alleles at the same time, and that the data will most often be the result of a trade-off between these two types of error.

## 5 | ISSUE #2: CORRELATION BETWEEN OBSERVED AND TRUE ALLELIC FREQUENCIES

The adequacy of data based on DNA retrieved from environmental or community samples for population genetics often relies on the major assumption that the observed (inferred) allele frequencies and the true allele frequencies are highly correlated in the target population. In traditional methods, most violations of this assumption are due to sampling efforts (i.e., when the subset of sampled individuals is not representative of the individuals present in the population). Yet, many more factors influence this relationship for community-based DNA analyses. First, the amount of DNA sampled from the environment is not necessarily representative of the current state

of the population (Cristescu & Hebert, 2018). It is notably impacted by different shedding rates between individuals (e.g., between sexes for organism with an important sexual dimorphism or between life stages; see Andruszkiewicz et al., 2021; Crane et al., 2021) and by DNA degradation and transport from one site to another (Figure 3, issue #2). Moreover, due to DNA degradation in the environment, the proportion of DNA fragments can favor individuals geographically close to the sampling point. These problems are difficult to quantify and might have a great influence on the results. For example, the transport of DNA between two marine populations living in connected water masses could lead to an overestimation of gene flow. To account for such limitations, it is important to have sufficient sampling replication, both spatially and in time. As a high number of replicates can be costly to process, however, spatial and temporal variability could also be integrated by continuous sampling, either using portable (Thomas et al., 2018) or autonomous passive sampling devices (Verdier et al., 2022). The issue of transport and degradation can also be addressed using mathematical models such as the one described by Carraro et al. (2020). Additionally, it is crucial to know the biology of the target species and tailor the sampling design to their lifestyle. For example, Jensen et al. (2022) showed that the amount of DNA retrieved from water samples varied greatly throughout the day for different marine fish species. Thus, depending if the sampling was performed during the day or during the night, different conclusions would be made on the genetic differentiation of these populations.

Further discrepancies between observed and true allelic frequencies are due to technical limitations. Different methods differ in biases regarding their ability to adequately evaluate the amount of DNA present in the target sample, and approaches based on quantitative or digital PCR might be better suited for this purpose than techniques based on HTS (Wood et al., 2019). With the latter, differences in read abundances and DNA amount can arise at all steps of the library preparation. For example, primer mismatches can favor the amplification of particular haplotypes, artificially inflating their abundance in the dataset, although this issue is less likely when targeting one species than when targeting multiple species. Even though Azarian et al. (2021) suggested to use only presence-absence data for population structure inferences, abundance data are better suited for computing most population genetics metrics and one could thus favor more quantitative techniques, such as digital PCR, when the type of variants studied is adequate. Moreover, amplification and sequencing can create a non-negligible amount of errors, which results in false positives (i.e., alleles or haplotypes that are not real) if not properly removed during the bioinformatics steps, as shown in our quantitative assessment of published studies (Figure 2b). Removing as much false positives as possible while at the same time retaining true alleles/haplotypes is a crucial step in metabarcoding approaches, and it has been tested at several occasions for intraspecific diversity studies (Couton et al., 2021; Macé et al., 2022; Turon et al., 2020). All publications have the same shared message, namely that it is not possible to remove all erroneous sequences while not removing some of the signal too. It is a

classical type I and type II error trade-off, and its optimization will affect downstream analyses. Since the false positives and false negatives are usually in low abundances, however, they should only have a minor effect on the calculation of frequency-based estimates, such as  $F_{ST}$ , which should be favored over metrics based on absolute number, such as allelic richness (Couton et al., 2021).

Despite all the potentialities for discrepancies between true and observed allele frequencies when using DNA retrieved from environmental or community samples, the experiments performed so far in tanks or even in the field are showing a rather good correlation between results from tissue and water samples (Figure 2a). Moreover, DNA retrieved from environmental or community samples could represent a better alternative than traditional sampling in the case of particularly abundant and diverse populations, for which getting a representative sample of individuals would be very time consuming and costly. It could also be particularly useful for particular fields where only presence-absence data are required such as phylogeography, as shown by Pedersen et al. (2021) or Tsuji et al. (2023). To date, the effect of quantification biases on population genetics metrics is a rather speculative field with little empirical testing. We thus strongly encourage future studies aiming at better understanding this topic, as such evidence is required for a reliable interpretation of the results.

## 6 | ISSUE #3: LOSS OF INDIVIDUAL INFORMATION

For many questions in population genetics, the distribution of different alleles for given markers within and between groups of individuals must be evaluated. For many metrics, however, this information alone is often insufficient, and must be completed by data on the distribution of these alleles within individuals, for example to compute key indices such as the observed heterozygosity, inbreeding coefficients, and linkage disequilibrium. The main limitation when working with DNA retrieved from environmental or community samples is thus the loss of physical linkage between loci of an individual and the loss of association between the alleles at a given locus (Figure 3, issues #3a and #3b). This is especially problematic when the different markers used are in fact in linkage disequilibrium, which would artificially inflate the differentiation between populations. This limitation is fundamental, and cannot be resolved a posteriori with bioinformatics tools. It is the main difference between community-based and what we here refer to as non-invasive samples (see Box 1). For this particular type of sampling, as used in Dugal et al. (2022), Farrell et al. (2022) or Parsons et al. (2018), the collected samples can be directly linked to an individual, even if this individual might not have been observed at the time of sampling. To avoid creating false expectations for the use of environmental samples and DNA, it is thus important to differentiate these two types of studies (i.e., community-based vs. individual-based) that can both use the same samples (e.g., water samples) but that display different opportunities and limitations.

Association of alleles at a given marker is not an issue with haploid markers, and our literature review showed that mtDNA markers are, to date, the favored approach for population genetics based on DNA retrieved from environmental samples. Since many biodiversity studies use metabarcoding of mitochondrial markers, it has been suggested as a basis for intraspecific inferences (Elbrecht et al., 2018). In most cases, mtDNA is uniparentally inherited and homoplasmic. If only one marker is used, there is no need to link haplotypes to individuals. Importantly, however, in some cases heteroplasmy of mtDNA exists as a result of paternal leakage (Mastrantonio et al., 2019), hybridization (Patten et al., 2015) or from doubly uniparental inheritance (DUI; Ghiselli et al., 2019). Even if this limitation also holds with tissue samples analyses, it is usually easier to detect (e.g., double-peaks in Sanger sequencing; multi-allelic SNPs in HTS). Current applications of mtDNA could be improved by using long-range sequencing to get physically intact large fragments, and to retain within-individual information. Some of mtDNA characteristics, however, such as (mostly) maternal inheritance or absence of recombination, regardless of the type of sample used, make it questionable for population genetics (Bazin et al., 2006; Galtier et al., 2009). Being haploid, some indices, such as  $F_{IS}$ , and associated processes (e.g., inbreeding), cannot be examined. In addition, mtDNA cannot be used alone to detect or investigate cases of hybridization and introgression (Ballard & Whitlock, 2004), and is prone to selective sweeps (Bazin et al., 2006; Patten et al., 2015). It is also not suited to investigate more complex questions, such as demographic history or admixture (Gagnaire, 2020). Although mtDNA is so far favored for eDNA-based studies, its use should be aligned with the focus of the study.

When using data from DNA retrieved from community or environmental samples, most metrics that can be calculated for population genetics are group-based (e.g., expected heterozygosity, but not observed heterozygosity; Figure 4), and thus much fewer questions can be addressed. For example, any subject requiring unequivocal data on a particular individual cannot be investigated, such as kinship, assigning a given individual to a population of origin or admixture as described by Pritchard et al. (2000). In fact, the latter is based on a model-based approach to infer the genetic ancestry (i.e., the proportion of ancestry) of a given individual to one or more of  $K$  genetically distinct sources. Individual information is thus required to compute these membership proportions. The question of how to extend the use of group-based data in population genetics, however, has already been successfully addressed for HTS techniques with the development of pool-seq methods (Box 2). In the case of admixture for example, Gautier et al. (2022) developed an approach using  $f$ -statistics (and associated  $F$  parameters). Their models, however, require the number of individuals contributing to the pool, which is not known from eDNA samples. Even if it is not possible to directly apply models developed for pool-seq data to community data because of differences in the sequencing techniques, and because the number of individuals in a pool is usually known in a pool-seq approach, similar kind of models could be developed for data based on DNA retrieved from community or environmental samples. These

models could consider the transport of DNA, its degradation, but also the different technical biases mentioned above. As with any model, however, their estimates will always have a probabilistic nature, and an inherent uncertainty, which leads to less precise inferences. Nonetheless, these drawbacks could be counteracted by the benefits of using community-based DNA in particular cases, such as the comparative analysis of many taxa at the same time.

When individual information from data obtained from DNA retrieved from community or environmental samples is still needed, single-cell sequencing could be used, as suggested by Sigsgaard et al. (2020) or Adams et al. (2019), and as is already performed for prokaryotes (Xu & Zhao, 2018). Thereby, cells are physically separated and the full genome of a cell is amplified and sequenced. This preserves the association between all alleles and loci/markers identified. Although well established for prokaryotes, this approach is still a prospect for multi-cellular eukaryotes (Minamoto, 2022), and its feasibility in the context of eDNA sampling needs to be thoroughly tested. For example, the correspondence between the number of cells in the environment and the number of individuals or an individual's biomass is not necessarily known (contrastingly to prokaryotes or unicellular eukaryotes, where one cell generally equals to one individual). Another potential limitation is the proportion of multi-cellular eukaryotic DNA still encapsulated in cells in an environmental sample. These aspects are not trivial and are still under investigation (Jo et al., 2022; Mauvisseau et al., 2022).

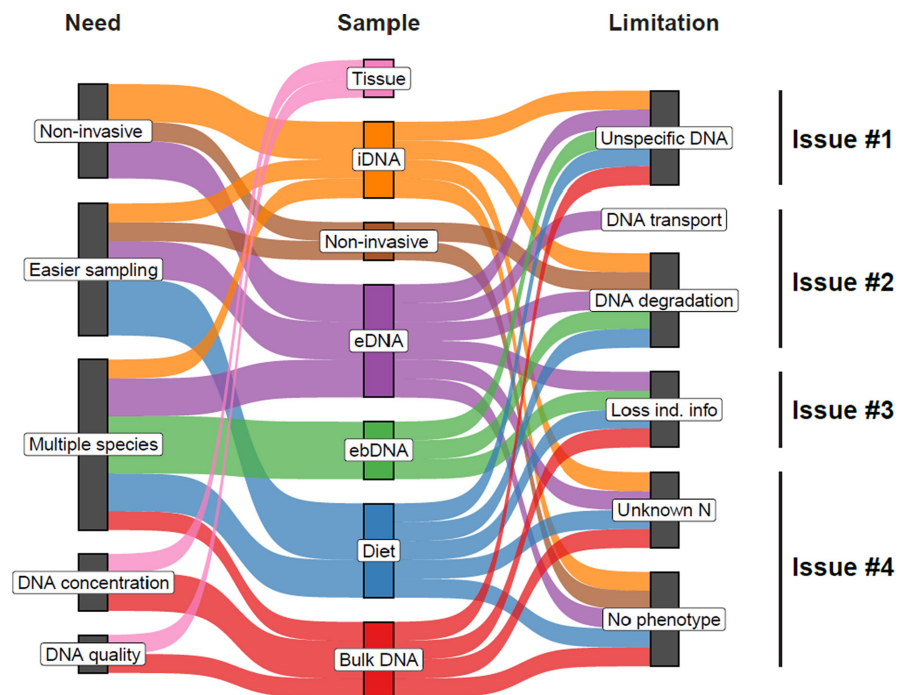
## 7 | ISSUE #4: UNCERTAINTY ABOUT THE INDIVIDUALS CONTRIBUTING TO THE SAMPLED DNA

One of the main strengths of eDNA, namely its ability to give information on organisms without physically seeing/sampling them, can also be one of its weaknesses. Indeed, if the sampling does not include processing actual individuals, it precludes the collection of phenotypic information, such as age, sex, morphotype, or other characteristics that might be useful to interpret the genetic patterns observed. Moreover, with environmental samples, it is impossible to know which life-stage is sampled, from gamete to adult (Figure 3, issue #4), and in which proportion. Unintentionally including DNA from juveniles or gametes into an analysis can alter the representativeness of the sample for the population (e.g., increase in the nuclear/mitochondrial DNA ratio; see Minamoto, 2022 and references therein). It can also severely affect the results as most models in population genetics assume that all specimens are from the same generation and can contribute to the next (i.e., the reproductive pool). Even more so, if these are in high proportion in the sample, which is expected for example for gametes in fishes, or juvenile larvae for many marine invertebrates. As an example, the genetic diversity of a population can be drastically underestimated if most of the sampled DNA comes from larvae from the same brood. Again, the sampling strategy and good knowledge of the target species' life-history are key. Samples should be taken,

**FIGURE 4** List of commonly used population genetics computations. For each computation, the applicability of information for different types of samples is evaluated. Green squares indicate that it is computable with the targeted sample type. The orange color for effective population size ( $N_e$ ) underlines that it is only computable based on particular estimates using nucleotide diversity ( $N_e = \pi / 2\mu$  with  $\mu$  being the mutation rate; see Sigsgaard et al. (2017)).

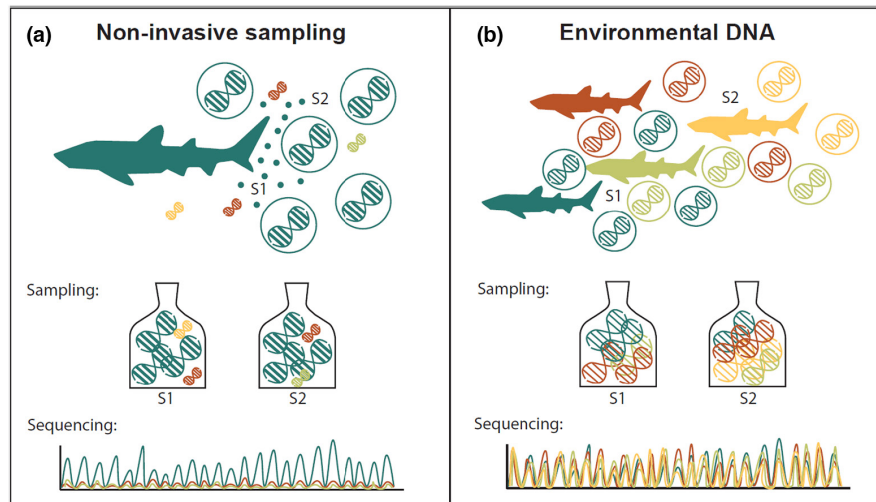
Population genetics computations	Code	eDNA samples	Pooled individuals	Individual Tissue
<b>1. Genetic diversity</b>				
Nucleotide diversity	$\pi$	Green	Green	Green
Allelic richness	$N_a$	Green	Green	Green
Minor allele frequency	MAF	Green	Green	Green
Number of private alleles	$N_p$	Green	Green	Green
Polymorphism information content	PIC	Green	Green	Green
Effective population size	$N_e$	Orange	Orange	Green
Expected Heterozygosity	$H_e$	Green	Green	Green
Observed Heterozygosity	Hobs	White (X)	White (X)	Green
Genotypic richness, diversity, evenness	e.g. $H'$ , $J'$	White (X)	White (X)	Green
Inbreeding coefficient	$F_{is}$	White (X)	White (X)	Green
Linkage disequilibrium	LD	White (X)	White (X)	Green
<b>2. Population structure</b>				
Genetic divergence	e.g., $F_{ST}$ , $\theta_{ST}$ , $G_{ST}$	Green	Green	Green
Isolation by distance	IBD	Green	Green	Green
Admixture	-	White (X)	Green	Green
Assign a given individual to a population of origin	-	White (X)	White (X)	Green
<b>3. Link to morphology - selection</b>				
Proportion of synonymous mutations	$\pi_s$	Green	Green	Green
Genome Wide Association Studies	-	White (X)	White (X)	Green
Genotype Environment Associations	-	White (X)	White (X)	Green

**FIGURE 5** Synthetic representation of the needs fulfilled by the different types of samples listed in the table in Box 1 and their inherent limitations. This should help for decision making, as both of these aspects should be carefully considered when choosing the type of sample to use during experimental design. The limitations mentioned in this figure are associated to the different issues discussed in this publication.



if possible, when only adults are present in the environment, and avoiding reproductive periods, in particular for aquatic species releasing their gametes and/or larvae into the environment (Couton et al., 2019; Lacoursière-Roussel et al., 2018). Sufficient replication can also reduce the effects of unwanted sampling of juveniles. Regarding phenotypic characteristics, retrieving them

separately with other approaches, such as camera traps, would not allow the association between genotypes and phenotypes. Only for very specific sub-aspects, however, it might be possible to infer phenotypic characteristics from eDNA (e.g., by employing dedicated markers such as sex-linked markers as suggested in Sigsgaard et al., 2020). Those could then be linked to individual



**FIGURE 6** Representation of the differences between (a) non-invasive sampling (as performed in Dugal et al., 2022) and (b) environmental DNA. In (a) the sampling is closely (in space and time) associated to a specific individual, for example released in the wake of a whale. In this case, the physical linkage can be established with a high certainty, and the target DNA dominates the signal. In (b) the physical linkage between DNA and any individual is lost, and possibly even between different marker regions. The DNA sampled at two different points (S1 and S2) is depicted as well as the potential result obtained after Sanger sequencing. This figure illustrates the possibility of collecting individual information with non-invasive sampling as opposed to environmental DNA.

### BOX 2 Developing models to use population-level data: the example of Pool-seq.

With the development of next generation sequencing techniques, the number of markers under investigation in population genetics studies has increased drastically. These new approaches, however, come with a higher processing cost per sample, and the requirement of a higher quantity and a better quality of DNA. Since these constraints cannot always be met, and to decrease labour and experimental costs associated with sequencing, researchers started pooling individuals and processing them in batches (Futschik & Schlötterer, 2010; Schlötterer et al., 2014). From the first attempt, this new approach was found to be effective (Futschik & Schlötterer, 2010) although it brought new challenges in a field previously oriented towards individual-based data. New models were developed to refine allelic frequencies estimates from pool-seq data, taking into account the number of individuals in the pool, the sequencing depth, the possibility of sequencing errors, and the different contributions of individuals in the pool (Anand et al., 2016; Gautier et al., 2013). The calculation of existing metrics and the detection of genetic patterns have also been modified to be more accurate with this type of data (e.g., selective sweeps in Boitard et al., 2013;  $F_{ST}$  in Hivert et al., 2018), and now allow to make complex inferences, such as demographic scenarios based on  $f$ -statistics (Gautier et al., 2022). This evolution in the field of population genetics could serve as an example for future developments in eDNA-based population studies.

genotypes/haplotypes providing the use of single-cell sequencing, as mentioned in the previous section. Another possibility would be to take advantage of eRNA in cases where specific transcripts could be associated to particular life-history stages or phenotypes of a given species (Yates et al., 2021). To our knowledge, such uses of environmental samples to go beyond species detection had not yet been implemented.

In classic population genetics derived from tissue samples, the strongest source of confidence and robustness of the results comes from the number of individuals sampled at each site. The information on the number of individuals sampled/contributing to an eDNA or a bulkDNA sample, however, is lacking, violating the assumption of equal sampling depths across populations in probably most cases. As an example, with tissue samples, low genetic diversity data coming from a population of five reproductive individuals will be interpreted very differently than the same diversity coming from a population of 500 reproductive individuals. Conversely, eDNA-based or bulkDNA-based data do not allow such a distinction. Several studies have evaluated the number of contributors from eDNA data by looking at the genetic diversity of their samples (e.g., Andres et al., 2021; Yoshitake et al., 2019). The numbers inferred, however, are dependent on the dataset itself and can thus not be used for evaluating the confidence of the results.

## 8 | CONCLUSION

The pros and cons of population genetics approaches based on tissue sampling are well known, and their discussion is already well established (e.g., Fumagalli, 2013; Meirns, 2015). Genetic data retrieved from community or environmental samples, such as eDNA, however, have different biases and their interpretation cannot be done with



the same priors as it is done for tissue samples. In this publication, we listed four categories of major limitations inherent to the use of DNA retrieved from community or environmental samples in population genetics, namely (1) the difficulty to get reliable species-specific data, (2) the potential lack of correlation between observed and true allelic frequencies, (3) the loss of individual information, and (4) the impossibility to know exactly what constitutes our sample. They are mutually non-exclusive, and one can face several of them at the same time. Some of these limitations might be solved with the development of future technologies or could be minimized by adapting the sampling strategy and bioinformatics processes. Others, such as the missing information on potential association between markers, actual population sizes/numbers of individuals sampled or DNA transport will probably never or hardly be resolved. With the rapid development of the research field, it is of the utmost importance that current studies carefully ponder their inferences against these potential biases. Based on our assessment, most publications to date on this topic are proof-of-concept studies where authors evaluate the feasibility of using eDNA for population genetics in relatively specific cases. From the small set of studies comparing tissue samples against community-based samples, it appears that the results from both approaches have a good correlation. However, there are discrepancies between methods and we still do not know how they can affect the indices computed in population genetics studies. It thus appears that DNA retrieved from community or environmental samples, and notably eDNA (which is becoming increasingly used) is not yet ready for a routine use for population genetics purposes, being now limited to a set of mostly group-based metrics. It should also be employed preferably in environments and to study organisms for which a good knowledge is already available. Thus, while it can be a valuable complement to traditional methods based on individual tissue samples for particular species or habitats, and for comparative studies of multiple taxa, it is important to be aware of its current limitations for an appropriate use of this promising type of data.

#### AUTHOR CONTRIBUTIONS

All authors conceived the manuscript and contributed to writing. MC performed the literature search.

#### ACKNOWLEDGMENTS

The authors thank the Swiss National Science Foundation (grant nr. 31003A\_173074), the Swiss Federal Office for the Environment (FOEN/BAFU), the Fachverband für Gas, Wärme und Wasser (SVGW/SSIGE), as well as the University of Zurich Research Priority Programme in Global Change and Biodiversity (URPP GCB) for funding (grants to FA). This work also benefited from results and ideas developed in the Aquanis2.0 (TOTAL Foundation) and MarEEE (i-site MUSE; French National Research Agency under the "Investissements d'Avenir" programme ANR-16-IDEX-0006) projects allocated to FV. This is publication ISEM 2023-132.

#### FUNDING INFORMATION

Swiss National Science Foundation (grant nr. 31003A\_173074); Swiss Federal Office for the Environment (FOEN/BAFU); Fachverband

für Gas, Wärme und Wasser (SVGW/SSIGE); University of Zurich Research Priority Programme in Global Change and Biodiversity (URPP GCB); Aquanis2.0 (TOTAL Foundation); MarEEE (i-site MUSE; French National Research Agency under the "Investissements d'Avenir" programme ANR-16-IDEX-0006).

#### CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

#### DATA AVAILABILITY STATEMENT

No data were produced for this publication.

#### ORCID

Marjorie Couton  <https://orcid.org/0000-0001-9880-8646>

Frédérique Viard  <https://orcid.org/0000-0001-5603-9527>

Florian Altermatt  <https://orcid.org/0000-0002-4831-6958>

#### REFERENCES

- Adams, C. I., Hepburn, C., Jeunen, G.-J., Cross, H., Taylor, H. R., Gemmell, N. J., Bunce, M., & Knapp, M. (2022). Environmental DNA reflects common haplotypic variation. *Environmental DNA*, 1–14.
- Adams, C. I., Knapp, M., Gemmell, N. J., Jeunen, G.-J., Bunce, M., Lamare, M. D., & Taylor, H. R. (2019). Beyond biodiversity: Can environmental DNA (eDNA) cut it as a population genetics tool? *Genes*, 10(3), 192.
- Afshinnekoo, E., Meydan, C., Chowdhury, S., Jaroudi, D., Boyer, C., Bernstein, N., Maritz, J. M., Reeves, D., Gandara, J., Chhangawala, S., Ahsanuddin, S., Simmons, A., Nessel, T., Sundaresh, B., Pereira, E., Jorgensen, E., Kolokotronis, S.-O., Kirchberger, N., Garcia, I., ... Mason, C. E. (2015). Geospatial resolution of human and bacterial diversity with city-scale metagenomics. *Cell Systems*, 1(1), 72–87.
- Allendorf, F. W. (2017). Genetics and the conservation of natural populations: Allozymes to genomes. *Molecular Ecology*, 26(2), 420–430.
- Alther, R., Fronhofer, E. A., & Altermatt, F. (2021). Dispersal behaviour and riverine network connectivity shape the genetic diversity of freshwater amphipod metapopulations. *Molecular Ecology*, 30(24), 6551–6565.
- Anand, S., Mangano, E., Barizzone, N., Bordoni, R., Sorosina, M., Clarelli, F., Corrado, L., Martinelli, B. F., D'Alfonso, S., & de Bellis, G. (2016). Next generation sequencing of pooled samples: Guideline for variants' filtering. *Scientific Reports*, 6(1), 33735.
- Andres, K. J., Sethi, S. A., Lodge, D. M., & Andrés, J. (2021). Nuclear eDNA estimates population allele frequencies and abundance in experimental mesocosms and field samples. *Molecular Ecology*, 30(3), 685–697.
- Andruszkiewicz, E. A., Zhang, W. G., Lavery, A. C., & Govindarajan, A. F. (2021). Environmental DNA shedding and decay rates from diverse animal forms and thermal regimes. *Environmental DNA*, 3(2), 492–514.
- Antich, A., Palacín, C., Zarcerro, J., Wangenstein, O. S., & Turon, X. (2022). Metabarcoding reveals high-resolution biogeographical and metaphylogeographical patterns through marine barriers. *Journal of Biogeography*, 50(3), 515–527.
- Ardura, A., Planes, S., & Garcia-Vazquez, E. (2013). Applications of DNA barcoding to fish landings: Authentication and diversity assessment. *ZooKeys*, 365, 49–65.
- Azarian, C., Foster, S., Devloo-Delva, F., & Feutry, P. (2021). Population differentiation from environmental DNA: Investigating the potential of haplotype presence/absence-based analysis of molecular variance. *Environmental DNA*, 3(3), 541–552.

- Bálint, M., Pfenninger, M., Grossart, H.-P., Taberlet, P., Vellend, M., Leibold, M. A., Englund, G., & Bowler, D. (2018). Environmental DNA time series in ecology. *Trends in Ecology & Evolution*, 33(12), 945–957.
- Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, 13(4), 729–744.
- Banerjee, P., Stewart, K. A., Antognazza, C. M., Bunholi, I. V., Deiner, K., Barnes, M. A., Saha, S., Verdier, H., Doi, H., Maity, J. P., Chan, M. W. Y., & Chen, C. Y. (2022). Plant–animal interactions in the era of environmental DNA (eDNA)—A review. *Environmental DNA*, 4(5), 987–999.
- Barbará, T., Palma-Silva, C., Paggi, G. M., Bered, F., Fay, M. F., & Lexer, C. (2007). Cross-species transfer of nuclear microsatellite markers: Potential and limitations. *Molecular Ecology*, 16(18), 3759–3767.
- Barnes, M. A., & Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, 17(1), 1–17.
- Bazin, E., Glémin, S., & Galtier, N. (2006). Population size does not influence mitochondrial genetic diversity in animals. *Science*, 312(5773), 570–572.
- Bista, I., Carvalho, G. R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., Christmas, M., & Creer, S. (2017). Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nature Communications*, 8(1), 14087.
- Blackman, R. C., Ho, H.-C., Walser, J.-C., & Altermatt, F. (2022). Spatio-temporal patterns of multi-trophic biodiversity and food-web characteristics uncovered across a river catchment using environmental DNA. *Communications Biology*, 5(1), 259.
- Blackman, R. C., Osathanunkul, M., Brantschen, J., di Muri, C., Harper, L. R., Mächler, E., Hänfling, B., & Altermatt, F. (2021). Mapping biodiversity hotspots of fish communities in subtropical streams through environmental DNA. *Scientific Reports*, 11(1), 10375.
- Bock, D. G., MacIsaac, H. J., & Cristescu, M. E. (2012). Multilocus genetic analyses differentiate between widespread and spatially restricted cryptic species in a model ascidian. *Proceedings of the Royal Society B: Biological Sciences*, 279(1737), 2377–2385.
- Bohmann, K., Gopalakrishnan, S., Nielsen, M., Nielsen, L. S. B., Jones, G., Streicker, D. G., & Gilbert, M. T. P. (2018). Using DNA metabarcoding for simultaneous inference of common vampire bat diet and population structure. *Molecular Ecology Resources*, 18(5), 1050–1063.
- Boitard, S., Kofler, R., Françoise, P., Robelin, D., Schlötterer, C., & Futschik, A. (2013). Pool-hmm: A python program for estimating the allele frequency spectrum and detecting selective sweeps from next generation sequencing of pooled samples. *Molecular Ecology Resources*, 13(2), 337–340.
- Bovo, S., Utzeri, V. J., Ribani, A., Taurisano, V., Schiavo, G., & Fontanesi, L. (2022). A genotyping by sequencing approach can disclose *Apis mellifera* population genomic information contained in honey environmental DNA. *Scientific Reports*, 12, 19541.
- Brantschen, J., Blackman, R. C., Walser, J.-C., & Altermatt, F. (2021). Environmental DNA gives comparable results to morphology-based indices of macroinvertebrates in a large-scale ecological assessment. *PLoS One*, 16(9), e0257510.
- Burgoa, C. J., Deconinck, D., Márquez, I., Peón, T. P., Garcia-Vazquez, E., & Machado-Schiaffino, G. (2020). New eDNA based tool applied to the specific detection and monitoring of the endangered European eel. *Biological Conservation*, 250, 108750.
- Carraro, L., Mächler, E., Wüthrich, R., & Altermatt, F. (2020). Environmental DNA allows upscaling spatial patterns of biodiversity in freshwater ecosystems. *Nature Communications*, 11(1), 3585.
- Carroll, E. L., Gallego, R., Sewell, M. A., Zeldis, J., Ranjard, L., Ross, H. A., Tooman, L. K., O'Rourke, R., Newcomb, R. D., & Constantine, R. (2019). Multi-locus DNA metabarcoding of zooplankton communities and scat reveal trophic interactions of a generalist predator. *Scientific Reports*, 9(1), 281.
- Chain, F. J. J., Brown, E. A., Macisaac, H. J., & Cristescu, M. E. (2016). Metabarcoding reveals strong spatial structure and temporal turnover of zooplankton communities among marine and freshwater ports. *Diversity and Distributions*, 22(5), 493–504.
- Charlesworth, B., & Charlesworth, D. (2017). Population genetics from 1966 to 2016. *Heredity*, 118(1), 2–9.
- Corbett-Detig, R. B., Hartl, D. L., & Sackton, T. B. (2015). Natural selection constrains neutral diversity across a wide range of species. *PLoS Biology*, 13(4), e1002112.
- Couton, M., Baud, A., Daguin-Thiébaud, C., Corre, E., Comtet, T., & Viard, F. (2021). High-throughput sequencing on preservative ethanol is effective at jointly examining intraspecific and taxonomic diversity, although bioinformatics pipelines do not perform equally. *Ecology and Evolution*, 11, 5533–5546.
- Couton, M., Comtet, T., Le Cam, S., Corre, E., & Viard, F. (2019). Metabarcoding on planktonic larval stages: An efficient approach for detecting and investigating life cycle dynamics of benthic aliens. *Management of Biological Invasions*, 10(4), 657–689.
- Couton, M., Lévêque, L., Daguin-Thiébaud, C., Comtet, T., & Viard, F. (2022). Water eDNA metabarcoding is effective in detecting non-native species in marinas, but detection errors still hinder its use for passive monitoring. *Biofouling*, 38(4), 367–383.
- Crane, L. C., Goldstein, J. S., Thomas, D. W., Rexroth, K. S., & Watts, A. W. (2021). Effects of life stage on eDNA detection of the invasive European green crab (*Carcinus maenas*) in estuarine systems. *Ecological Indicators*, 124, 107412.
- Cristescu, M. E. (2014). From barcoding single individuals to metabarcoding biological communities: Towards an integrative approach to the study of global biodiversity. *Trends in Ecology & Evolution*, 29(10), 566–571.
- Cristescu, M. E., & Hebert, P. D. N. (2018). Uses and misuses of environmental DNA in biodiversity science and conservation. *Annual Review of Ecology, Evolution, and Systematics*, 49(1), 209–230.
- Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., de Vere, N., Pfrender, M. E., & Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26(21), 5872–5895.
- Dugal, L., Thomas, L., Jensen, M. R., Sigsgaard, E. E., Simpson, T., Jarman, S., Thomsen, P. F., & Meekan, M. (2022). Individual haplotyping of whale sharks from seawater environmental DNA. *Molecular Ecology Resources*, 22(1), 56–65.
- Elbrecht, V., Vamos, E. E., Steinke, D., & Leese, F. (2018). Estimating intraspecific genetic diversity from community DNA metabarcoding data. *PeerJ*, 6, e4644.
- Farrell, J. A., Whitmore, L., Mashkour, N., Rollinson Ramia, D. R., Thomas, R. S., Eastman, C. B., Burkhalter, B., Yetsko, K., Mott, C., Wood, L., Zirkelbach, B., Meers, L., Kleinsasser, P., Stock, S., Libert, E., Herren, R., Eastman, S., Crowder, W., Boverly, C., ... Duffy, D. J. (2022). Detection and population genomics of sea turtle species via non-invasive environmental DNA analysis of nesting beach sand tracks and oceanic water. *Molecular Ecology Resources*, 22, 2471–2493.
- Fumagalli, M. (2013). Assessing the effect of sequencing depth and sample size in population genetics inferences. *PLoS One*, 8(11), e79667.
- Futschik, A., & Schlötterer, C. (2010). The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics*, 186(1), 207–218.
- Gagnaire, P.-A. (2020). Comparative genomics approach to evolutionary process connectivity. *Evolutionary Applications*, 13(6), 1320–1334.
- Galtier, N., Nabholz, B., Glémin, S., & Hurst, G. D. D. (2009). Mitochondrial DNA as a marker of molecular diversity: A reappraisal. *Molecular Ecology*, 18(22), 4541–4550.
- Gautier, M., Foucaud, J., Gharbi, K., Cézard, T., Galan, M., Loiseau, A., Thomson, M., Pudlo, P., Kerdelhué, C., & Estoup, A. (2013). Estimation of population allele frequencies from next-generation

- sequencing data: Pool-versus individual-based genotyping. *Molecular Ecology*, 22(14), 3766–3779.
- Gautier, M., Vitalis, R., Flori, L., & Estoup, A. (2022). *f*-statistics estimation and admixture graph construction with Pool-seq or allele count data using the R package poolstat. *Molecular Ecology Resources*, 22(4), 1394–1416.
- Ghiselli, F., Maurizii, M. G., Reunov, A., Ariño-Bassols, H., Cifaldi, C., Pecci, A., Alexandrova, Y., Bettini, S., Passamonti, M., Franceschini, V., & Milani, L. (2019). Natural heteroplasmy and mitochondrial inheritance in bivalve molluscs. *Integrative and Comparative Biology*, 59(4), 1016–1032.
- Goldberg, C. S., & Parsley, M. B. (2021). Environmental population genomics: Challenges and opportunities. In P. A. Hohenlohe & O. P. Rajora (Eds.), *Population genomics: Wildlife* (pp. 101–113). Springer International Publishing.
- Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, 270(1512), 313–321.
- Hebert, P. D. N., Ratnasingham, S., & de Waard, J. R. (2003). Barcoding animal life: Cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences*, 270(suppl\_1), S96–S99.
- Hivert, V., Leblois, R., Petit, E. J., Gautier, M., & Vitalis, R. (2018). Measuring genetic differentiation from pool-seq data. *Genetics*, 210(1), 315–330.
- Hobbs, J., Round, J. M., Allison, M. J., & Helbing, C. C. (2019). Expansion of the known distribution of the coastal tailed frog, *Ascaphus truei*, in British Columbia, Canada, using robust eDNA detection methods. *PLoS One*, 14(3), e0213849.
- Jenkins, T. L., Castilho, R., & Stevens, J. R. (2018). Meta-analysis of Northeast Atlantic marine taxa shows contrasting phylogeographic patterns following post-LGM expansions. *PeerJ*, 6, e5684.
- 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., Møller, P. R., & Thomsen, P. F. (2022). Short-term temporal variation of coastal marine eDNA. *Environmental DNA*, 4(4), 747–762.
- Jensen, M. R., Sigsgaard, E. E., Liu, S., Manica, A., Bach, S. S., Hansen, M. M., Møller, P. R., & Thomsen, P. F. (2021). Genome-scale target capture of mitochondrial and nuclear environmental DNA from water samples. *Molecular Ecology Resources*, 21(3), 690–702.
- Jerde, C. L., Mahon, A. R., Chadderton, W. L., & Lodge, D. M. (2011). “Sight-unseen” detection of rare aquatic species using environmental DNA. *Conservation Letters*, 4(2), 150–157.
- Jo, T., Takao, K., & Minamoto, T. (2022). Linking the state of environmental DNA to its application for biomonitoring and stock assessment: Targeting mitochondrial/nuclear genes, and different DNA fragment lengths and particle sizes. *Environmental DNA*, 4(2), 271–283.
- Kapoor, V., DeBry, R. W., Boccelli, D. L., & Wendell, D. (2014). Sequencing human mitochondrial hypervariable region II as a molecular fingerprint for environmental waters. *Environmental Science & Technology*, 48(18), 10648–10655.
- Keck, F., Blackman, R. C., Bossart, R., Brantschen, J., Couton, M., Hürlemann, S., Kirschner, D., Locher, N., Zhang, H., & Altermatt, F. (2022). Meta-analysis shows both congruence and complementarity of DNA and eDNA metabarcoding to traditional methods for biological community assessment. *Molecular Ecology*, 31(6), 1820–1835.
- Keck, F., Couton, M., & Altermatt, F. (2023). Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. *Molecular Ecology Resources*, 23, 742–755. <https://doi.org/10.1111/1755-0998.13746>
- Kijewski, T. K., Zbawicka, M., Väinölä, R., & Wenne, R. (2006). Introgression and mitochondrial DNA heteroplasmy in the Baltic populations of mussels *Mytilus trossulus* and *M. edulis*. *Marine Biology*, 149(6), 1371–1385.
- Lacoursière-Roussel, A., & Deiner, K. (2021). Environmental DNA is not the tool by itself. *Journal of Fish Biology*, 98(2), 383–386.
- Lacoursière-Roussel, A., Howland, K., Normandeau, E., Grey, E. K., Archambault, P., Deiner, K., Lodge, D. M., Hernandez, C., Leduc, N., & Bernatchez, L. (2018). eDNA metabarcoding as a new surveillance approach for coastal Arctic biodiversity. *Ecology and Evolution*, 8(16), 7763–7777.
- Leray, M., & Knowlton, N. (2016). Censusing marine eukaryotic diversity in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1702), 20150331.
- Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Boehm, J. T., & Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, 10(1), 34.
- Li, F., Zhang, Y., Altermatt, F., Zhang, X., Cai, Y., & Yang, Z. (2022). Gap analysis for DNA-based biomonitoring of aquatic ecosystems in China. *Ecological Indicators*, 137, 108732.
- Linard, B., Arribas, P., Andújar, C., Crampton-Platt, A., & Vogler, A. P. (2016). Lessons from genome skimming of arthropod-preserved ethanol. *Molecular Ecology Resources*, 16(6), 1365–1377.
- Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics*, 4(12), 981–994.
- Macé, B., Hocdé, R., Marques, V., Guerin, P.-E., Valentini, A., Arnal, V., Pellissier, L., & Manel, S. (2022). Evaluating bioinformatics pipelines for population-level inference using environmental DNA. *Environmental DNA*, 4(3), 674–686.
- Marchesini, A., Vernesi, C., Battisti, A., & Ficetola, G. F. (2018). Deciphering the drivers of negative species–genetic diversity correlation in alpine amphibians. *Molecular Ecology*, 27(23), 4916–4930.
- Marshall, N. T., & Stepien, C. A. (2019). Invasion genetics from eDNA and thousands of larvae: A targeted metabarcoding assay that distinguishes species and population variation of zebra and quagga mussels. *Ecology and Evolution*, 9(6), 3515–3538.
- Mastrantonio, V., Urbanelli, S., & Porretta, D. (2019). Ancient hybridization and mtDNA introgression behind current paternal leakage and heteroplasmy in hybrid zones. *Scientific Reports*, 9(1), 19177.
- Mauvisseau, Q., Harper, L. R., Sander, M., Hanner, R. H., Kleyer, H., & Deiner, K. (2022). The multiple states of environmental DNA and what is known about their persistence in aquatic environments. *Environmental Science & Technology*, 56(9), 5322–5333.
- Meekan, M., Austin, C. M., Tan, M. H., Wei, N.-W. V., Miller, A., Pierce, S. J., Rowat, D., Stevens, G., Davies, T. K., Ponzo, A., & Gan, H. M. (2017). iDNA at sea: Recovery of whale shark (*Rhincodon typus*) mitochondrial DNA sequences from the whale shark copepod (*Pandarus rhincodoniscus*) confirms global population structure. *Frontiers in Marine Science*, 4, 420.
- Meiklejohn, K., Damaso, N., & Robertson, J. (2019). Assessment of BOLD and GenBank – Their accuracy and reliability for the identification of biological materials. *PLoS One*, 14, e0217084.
- Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology*, 24(13), 3223–3231.
- Minamoto, T. (2022). Environmental DNA analysis for macro-organisms: Species distribution and more. *DNA Research*, 29(3), dsac018.
- Monge, O., Dumas, D., & Baus, I. (2020). Environmental DNA from avian residual saliva in fruits and its potential uses in population genetics. *Conservation Genetics Resources*, 12(1), 131–139.
- Nguyen, T. V., Tilker, A., Nguyen, A., Hörig, L., Axtner, J., Schmidt, A., Le, M., Nguyen, A. H. Q., Rawson, B. M., Wilting, A., & Fickel, J. (2021). Using terrestrial leeches to assess the genetic diversity of an elusive species: The Annamite striped rabbit *Nesolagus timminsi*. *Environmental DNA*, 3(4), 780–791.
- Niemiller, M. L., Porter, M. L., Keany, J., Gilbert, H., Fong, D. W., Culver, D. C., Hobson, C. S., Kendall, K. D., Davis, M. A., & Taylor, S. J.



- (2018). Evaluation of eDNA for groundwater invertebrate detection and monitoring: A case study with endangered *Stygobromus* (Amphipoda: Crangonyctidae). *Conservation Genetics Resources*, 10(2), 247–257.
- Olds, B. P., Jerde, C. L., Renshaw, M. A., Li, Y., Evans, N. T., Turner, C. R., Deiner, K., Mahon, A. R., Brueseke, M. A., Shirey, P. D., Pfrender, M. E., Lodge, D. M., & Lambert, G. A. (2016). Estimating species richness using environmental DNA. *Ecology and Evolution*, 6(12), 4214–4226.
- Parsons, K. M., Everett, M., Dahlheim, M., & Park, L. (2018). Water, water everywhere: Environmental DNA can unlock population structure in elusive marine species. *Royal Society Open Science*, 5(8), 180537.
- Patten, M. M., Carioscia, S. A., & Linnen, C. R. (2015). Biased introgression of mitochondrial and nuclear genes: A comparison of diploid and haplodiploid systems. *Molecular Ecology*, 24(20), 5200–5210.
- 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(22), 4258–4264.
- Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apothéloz-Perret-Gentil, L., Beja, P., Boggero, A., Borja, A., Bouchez, A., Cordier, T., Domaizon, I., Feio, M. J., Filipe, A. F., Fornaroli, R., Graf, W., Herder, J., van der Hoorn, B., Iwan, J. J., Sagova-Mareckova, M., Moritz, C., ... Kahlert, M. (2018). The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. *Science of the Total Environment*, 637, 1295–1310.
- Pedersen, M. W., De Sanctis, B., Saremi, N. F., Sikora, M., Puckett, E. E., Gu, Z., Moon, K. L., Kapp, J. D., Vinner, L., Vardanyan, Z., Ardelean, C. F., Arroyo-Cabrales, J., Cahill, J. A., Heintzman, P. D., Zazula, G., MacPhee, R. D. E., Shapiro, B., Durbin, R., & Willerslev, E. (2021). Environmental genomics of late Pleistocene black bears and giant short-faced bears. *Current Biology*, 31(12), 2728–2736.
- Pedro, P. M., Piper, R., Bazilli, N. P., Cullen, L., Jr., Dropa, M., Lorenao, R., Matté, M.-H., Rech, T. C., Rufato, M. O., Jr., Silva, M., & Turati, D. T. (2017). Metabarcoding analyses enable differentiation of both interspecific assemblages and intraspecific divergence in habitats with differing management practices. *Environmental Entomology*, 46(6), 1381–1389.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959.
- Qu, C., & Stewart, K. A. (2019). Evaluating monitoring options for conservation: Comparing traditional and environmental DNA tools for a critically endangered mammal. *The Science of Nature*, 106(3), 9.
- Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals – Mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, 15(11), 749–763.
- Sethi, S. A., Larson, W., Turnquist, K., & Isermann, D. (2019). Estimating the number of contributors to DNA mixtures provides a novel tool for ecology. *Methods in Ecology and Evolution*, 10(1), 109–119.
- Sigsgaard, E. E., Jensen, M. R., Winkelmann, I. E., Møller, P. R., Hansen, M. M., & Thomsen, P. F. (2020). Population-level inferences from environmental DNA—Current status and future perspectives. *Evolutionary Applications*, 13(2), 245–262.
- Sigsgaard, E. E., Nielsen, I. B., Bach, S. S., Lorenzen, E. D., Robinson, D. P., Knudsen, S. W., Pedersen, M. W., Jaidah, M. A., Orlando, L., Willerslev, E., Møller, P. R., & Thomsen, P. F. (2017). Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution*, 1, 4.
- Smart, A. S., Weeks, A. R., van Rooyen, A. R., Moore, A., McCarthy, M. A., & Tingley, R. (2016). Assessing the cost-efficiency of environmental DNA sampling. *Methods in Ecology and Evolution*, 7(11), 1291–1298.
- Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology*, 21(8), 1789–1793.
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, 21(8), 2045–2050.
- Taberlet, P., Waits, L. P., & Luikart, G. (1999). Noninvasive genetic sampling: Look before you leap. *Trends in Ecology & Evolution*, 14(8), 323–327.
- Tardy, C., Planes, S., Jung, J.-L., Ody, D., & Boissin, E. (2020). Characterization of 25 new microsatellite markers for the fin whale (*Balaenoptera physalus*) and cross-species amplification in other cetaceans. *Molecular Biology Reports*, 47(9), 6983–6996.
- Thomas, A. C., Howard, J., Nguyen, P. L., Seimon, T. A., & Goldberg, C. S. (2018). eDNA sampler: A fully integrated environmental DNA sampling system. *Methods in Ecology and Evolution*, 9(6), 1379–1385.
- Tsuji, S., Miya, M., Ushio, M., Sato, H., Minamoto, T., & Yamanaka, H. (2020). Evaluating intraspecific genetic diversity using environmental DNA and denoising approach: A case study using tank water. *Environmental DNA*, 2(1), 42–52.
- Tsuji, S., Shibata, N., Inui, R., Nakao, R., Akamatsu, Y., & Watanabe, K. (2023). Environmental DNA phylogeography: Successful reconstruction of phylogeographic patterns of multiple fish species from cups of water. *Molecular Ecology Resources*, 23(5), 1050–1065.
- Tsuji, S., Takahara, T., Doi, H., Shibata, N., & Yamanaka, H. (2019). The detection of aquatic macroorganisms using environmental DNA analysis—A review of methods for collection, extraction, and detection. *Environmental DNA*, 1(2), 99–108.
- Turon, X., Antich, A., Palacín, C., Præbel, K., & Wangensteen, O. S. (2020). From metabarcoding to metaphylogeography: Separating the wheat from the chaff. *Ecological Applications*, 30(2), e202036.
- Uchii, K., Doi, H., Yamanaka, H., & Minamoto, T. (2017). Distinct seasonal migration patterns of Japanese native and non-native genotypes of common carp estimated by environmental DNA. *Ecology and Evolution*, 7(20), 8515–8522.
- Valsecchi, E., Coppola, E., Pires, R., Parmegiani, A., Casiraghi, M., Galli, P., & Bruno, A. (2022). A species-specific qPCR assay provides novel insight into range expansion of the Mediterranean monk seal (*Monachus monachus*) by means of eDNA analysis. *Biodiversity and Conservation*, 31, 1175–1196.
- Vellend, M., Lajoie, G., Bourret, A., Múrria, C., Kembel, S. W., & Garant, D. (2014). Drawing ecological inferences from coincident patterns of population- and community-level biodiversity. *Molecular Ecology*, 23(12), 2890–2901.
- Verdier, H., Konecny-Dupre, L., Marquette, C., Reveron, H., Tadier, S., Grémillard, L., Barthès, A., Datry, T., Bouchez, A., & Lefebvre, T. (2022). Passive sampling of environmental DNA in aquatic environments using 3D-printed hydroxyapatite samplers. *Molecular Ecology Resources*, 22(6), 2158–2170.
- Viard, F., Roby, C., Turon, X., Bouchemousse, S., & Bishop, J. D. D. (2019). Cryptic diversity and database errors challenge non-indigenous species surveys: An illustration with *Botrylloides* spp. in the English Channel and Mediterranean Sea. *Frontiers in Marine Science*, 6, 615.
- Wang, H.-Y., Wei, B.-Y., & Xiao, H.-X. (2018). Development of 15 microsatellite markers in *Acer triflorum* (Aceraceae) and cross-amplification in congeneric species. *Applications in Plant Sciences*, 6(7), e01166.
- Wangensteen, O. S., Palacín, C., Guardiola, M., & Turon, X. (2018). DNA metabarcoding of littoral hard-bottom communities: High diversity and database gaps revealed by two molecular markers. *PeerJ*, 6, e4705.
- Weigand, H., Beerhmann, A. J., Čiampor, F., Costa, F. O., Csabai, Z., Duarte, S., Geiger, M. F., Grabowski, M., Rimet, F., Rulik, B., Strand, M., Szucsich, N., Weigand, A. M., Willassen, E., Wyler, S. A., Bouchez, A., Borja, A., Čiamporová-Zatovičová, Z., Ferreira, S., ... Ekrem, T. (2019). DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. *Science of the Total Environment*, 678, 499–524.
- Weitemier, K., Penaluna, B. E., Hauck, L. L., Longway, L. J., Garcia, T., & Cronn, R. (2021). Estimating the genetic diversity of Pacific salmon

- and trout using multigene eDNA metabarcoding. *Molecular Ecology*, 30(20), 4970–4990.
- Wood, S. A., Pochon, X., Laroche, O., von Ammon, U., Adamson, J., & Zaiko, A. (2019). A comparison of droplet digital polymerase chain reaction (PCR), quantitative PCR and metabarcoding for species-specific detection in environmental DNA. *Molecular Ecology Resources*, 19, 1407–1419.
- Xu, Y., & Zhao, F. (2018). Single-cell metagenomics: Challenges and applications. *Protein & Cell*, 9(5), 501–510.
- Yates, M. C., Derry, A. M., & Cristescu, M. E. (2021). Environmental RNA: A Revolution in Ecological Resolution? *Trends in Ecology & Evolution*, 36(7), 601–609.
- Yoshitake, K., Yoshinaga, T., Tanaka, C., Mizusawa, N., Reza, M. S., Tsujimoto, A., Kobayashi, T., & Watabe, S. (2019). HaCeD-seq: A novel method for reliable and easy estimation about the fish population using haplotype count from eDNA. *Marine Biotechnology*, 21(6), 813–820.
- Zaiko, A., Pochon, X., Garcia-Vazquez, E., Olenin, S., & Wood, S. A. (2018). Advantages and limitations of environmental DNA/RNA tools for marine biosecurity: Management and surveillance of non-indigenous species. *Frontiers in Marine Science*, 5, 322.
- Zbawicka, M., Burzyński, A., Skibinski, D., & Wenne, R. (2010). Scottish *Mytilus trossulus* mussels retain ancestral mitochondrial DNA: Complete sequences of male and female mtDNA genomes. *Gene*, 456(1), 45–53.
- Zenker, M. M., Specht, A., & Fonseca, V. G. (2020). Assessing insect biodiversity with automatic light traps in Brazil: Pearls and pitfalls of metabarcoding samples in preservative ethanol. *Ecology and Evolution*, 10(5), 2352–2366.
- Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., De Barba, M., Gaucher, P., Gielly, L., Giguët-Covex, C., Iribar, A., Réjou-Méchain, M., Rayé, G., Rioux, D., Schilling, V., Tymen, B., Viers, J., Zouiten, C., Thuiller, W., Coissac, E., & Chave, J. (2019). Body size determines soil community assembly in a tropical forest. *Molecular Ecology*, 28(3), 528–543.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Couton, M., Viard, F., & Altermatt, F. (2023). Opportunities and inherent limits of using environmental DNA for population genetics. *Environmental DNA*, 00, 1–17. <https://doi.org/10.1002/edn3.448>