Received: 26 January 2023 | Revised: 21 April 2023 | Accepted: 15 May 2023

DOI: 10.1111/mec.17031

INVITED REVIEW

MOLECULAR ECOLOGY WILEY

Detecting and analysing intraspecific genetic variation with eDNA: From population genetics to species abundance

Kara J. Andres[1,2,3] | David M. Lodge[1,4] | Suresh A. Sethi[5] | Jose Andrés[1]

[1]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York, USA

[2]Department of Biology, Washington University in St. Louis, St. Louis, Missouri, USA

[3]Living Earth Collaborative, Washington University in St. Louis, St. Louis, Missouri, USA

[4]Cornell Atkinson Center for Sustainability, Cornell University, Ithaca, New York, USA

[5]Fisheries, Aquatic Science and Technology Laboratory, Alaska Pacific University, Anchorage, Alaska, USA

Correspondence
Kara J. Andres, Living Earth Collaborative, Washington University in St. Louis, St. Louis, MO, USA.
Email: akara@wustl.edu

Funding information
Cooperative Institute for Great Lakes Research, Grant/Award Number: NA17OAR4320152; Cornell Atkinson Center for Sustainability; U.S. Department of Defense, Grant/Award Number: RC19-1004

Handling Editor: Pierre Taberlet

Abstract

Advancements in environmental DNA (eDNA) approaches have allowed for rapid and efficient species detections in diverse environments. Although most eDNA research is focused on leveraging genetic diversity to identify taxa, some recent studies have explored the potential for these approaches to detect within-species genetic variation, allowing for population genetic assessments and abundance estimates from environmental samples. However, we currently lack a framework outlining the key considerations specific to generating, analysing and applying eDNA data for these two purposes. Here, we discuss how various genetic markers differ with regard to genetic information and detectability in environmental samples and how analysis of eDNA samples differs from common tissue-based analyses. We then outline how it may be possible to obtain species absolute abundance estimates from eDNA by detecting intraspecific genetic variation in mixtures of DNA under multiple scenarios. We also identify the major causes contributing to allele detection and frequency errors in eDNA data, discuss their consequences for population-level analyses and outline bioinformatic approaches to detect and remove erroneous sequences. This review summarizes the key advances required to harness the full potential of eDNA-based intraspecific genetic variation to inform population-level questions in ecology, evolutionary biology and conservation management.

KEYWORDS
abundance, biodiversity, DNA mixtures, environmental DNA, metabarcoding

1 | INTRODUCTION

The use of molecular surveillance approaches such as environmental DNA (eDNA) to detect species has become more prevalent in recent years due to improved sensitivity and efficiency compared with conventional sampling methods (Fediajevaite et al., 2021). Through the analysis of genetic material present in environmental samples, eDNA approaches allow for the detection and quantification of single species (Ficetola et al., 2008; Jerde et al., 2011; Ruppert et al., 2019) or entire species assemblages (Deiner, Bik, et al., 2017; Valentini et al., 2016). The latter approach, called eDNA metabarcoding, involves the use of universal PCR primers aimed at amplifying a particular genetic region of multiple species simultaneously followed by high-throughput sequencing and taxonomic assignment of DNA sequences to species. Because eDNA metabarcoding allows for cost-efficient, rapid and non-invasive assessments of species richness, it is increasingly recognized as a useful alternative or complement to standard field sampling approaches for use in short- and long-term biomonitoring programs, biodiversity assessments and conservation (Lodge, 2022).

While eDNA metabarcoding approaches are well established for detecting organisms at or above the taxonomic level of species, genetic variation within the level of species is often disregarded. However, in recent years studies have emerged exploring the detection and quantification of genetic variation within species (i.e. intraspecific genetic variation) using eDNA approaches (see reviews in Adams et al., 2019; Sigsgaard et al., 2020; Yao et al., 2022). To date, analysis of intraspecific genetic variation from eDNA samples has been restricted to the detection and analysis of haplotypes in short regions of the mitochondrial genome similar to those typically targeted in eDNA metabarcoding (Parsons et al., 2018; Sigsgaard et al., 2016, 2020; Weitemier et al., 2021). By targeting longer and more variable mitochondrial markers or nuclear genetic markers, eDNA approaches have the potential to uncover much more detailed intraspecific genetic information that may be leveraged to conduct population genetic analyses and produce estimates of population abundance. These approaches present an appealing alternative to tissue-based population genetic research and species abundance estimations that require the physical capture of several individuals of the target species. However, analysis of population-level genetic variation from eDNA samples has several key limitations that require careful consideration. While recent work has demonstrated the potential to detect intraspecific genetic variation in eDNA samples, a synthesis of the key considerations specific to generating, analysing and applying such data for the purposes of characterizing population genetic diversity and assessing species abundance is needed.

In this paper, we discuss the challenges and opportunities to study intraspecific genetic variation using eDNA. Specifically, we discuss how genetic marker types (i.e. mitochondrial vs. nuclear genes) may differ with regard to detectability and genetic information content. We also outline the types of population genetics analyses that may be possible with data derived from mitochondrial eDNA and nuclear eDNA, where population-level genetic information is obtained but individual-level genotypes are unknown. Then, we explore the potential limits for using intraspecific genetic variation to estimate species-specific abundance by simulating datasets of different genetic markers under different conditions. Finally, we review the major causes of errors in eDNA data, discuss their consequences for population-level analyses and outline bioinformatic approaches to detect and remove erroneous sequences. By highlighting recent developments and future opportunities to study intraspecific genetic variation with eDNA, we provide a synthesis that lays a foundation for researchers interested in using eDNA to gain insights into the population characteristics and abundance of species. While the scope of this review is primarily limited to the study of contemporary macro-organisms using eDNA collected from water samples, many of the concepts within are applicable to metabarcoding of bulk samples, faecal samples or eDNA collected from substrates such as soil, snow and air.

2 | CHOICE OF MOLECULAR GENETIC MARKER

As in conventional specimen tissue-based population genetics research, the choice of molecular genetic marker targeted in eDNA research should be made by considering both practical criteria and the biological questions at hand (Anne, 2006). To deduce patterns of intraspecific genetic variation from eDNA, the targeted genetic region must contain sufficient levels of genetic variation at frequencies that can be detected using eDNA approaches. Thus, key considerations in the selection of genetic targets for eDNA research include the target amplicon length as well as the type and number of molecular genetic markers. For instance, longer amplicons are more likely to contain sequence variation, yet longer DNA fragments degrade more rapidly and are less abundant in environmental samples (Bylemans et al., 2018; Wei et al., 2018). The optimal amplicon length for intraspecific eDNA research must, therefore, strike a balance between a higher amount of detectable genetic variation and higher degradation rates. In the sections below, we focus on the choice of genetic marker type, another key decision that will impact the ability to detect intraspecific genetic variants from eDNA.

2.1 | Genetic marker type

Nuclear and mitochondrial genetic material may vary widely in their respective concentrations within environmental samples, which in turn can impact genetic marker detectability during sample processing and analysis. This is because a typical eukaryotic cell contains hundreds to thousands of mitochondrial genomes versus a single nuclear genome (Cole, 2016), resulting in much higher expected mitochondrial DNA (mtDNA) concentrations in eDNA samples compared with nuclear DNA (nuDNA). To date, research on the ratio of mtDNA to nuDNA concentrations in eDNA samples has focused on multi-copy nuclear ribosomal RNA (rRNA) genes such as 18S and ITS1, which are repeated nuclear genes that may have hundreds or thousands of copies per cell (Long & Dawid, 1980). Comparable copy numbers between mtDNA and nuclear rRNA markers are frequently reported, with nuclear rRNA genes sometimes exhibiting even higher detectability than mtDNA genes in eDNA samples (Dysthe et al., 2018; Gantz et al., 2018; Minamoto et al., 2017; Moushomi et al., 2019; Piggott, 2016). However, the ratio of mitochondrial to nuclear eDNA concentrations in environmental samples may change depending on a variety of factors including organism age, size, activity, tissue type and environmental conditions that influence the production and degradation rates of different marker types (Bylemans et al., 2017, 2018; Furtwängler et al., 2018; Jo et al., 2019, 2020). While most research on eDNA production, transport and degradation is focused on mtDNA markers, the dynamics of eDNA may change depending on the target gene, and the impact of such processes on eDNA detection and quantification for different marker types is not well understood.

In addition to their relative concentrations in environmental samples, mtDNA and nuDNA markers differ in their evolutionary properties (e.g. mutation rate, mode of inheritance and degree of recombination), influencing the type and amount of genetic information they contain. In general, mitochondrial genes are maternally inherited, do not recombine and are physically linked together. Because of this, single short mtDNA markers may not contain enough intraspecific genetic variation to conduct detailed population genetic analyses, and this limited information content cannot be overcome by targeting multiple mtDNA gene regions because they do not represent independent loci. Caution is warranted when drawing population inferences from mtDNA alone (Ballard & Whitlock, 2004), and many questions regarding intraspecific genetic variation may require the addition of nuclear genetic markers. Ultimately, the molecular genetic marker selected to detect intraspecific genetic variation from eDNA will depend on the goals of the study, including the number of targeted taxa and the amount of genetic variation required to address the research questions (Figure 1).

2.2 | Mitochondrial eDNA markers

Mitochondrial genes are the marker of choice for DNA-based taxonomic identification, with the COI gene representing the most popular marker in animal DNA barcoding as the Barcode of Life (Hebert et al., 2003; Ratnasingham & Hebert, 2007). Other mitochondrial gene regions commonly targeted for DNA barcoding of macro-organisms include 12S, 16S, cytochrome B (cytb) and ND2. DNA barcodes are specifically designed to maximize interspecific diversity while minimizing intraspecific diversity, and in most cases, intraspecific sequence divergence in these gene regions is not expected to reach levels required for detailed population genetic analyses. However, diagnostic sequence variants may be identified within mtDNA barcoding regions that can allow for the detection of specific haplotypes of interest. For instance, COI markers have been used for the diagnostic identification of specific zebra and quagga mussel haplotypes (Marshall & Stepien, 2019), and a diagnostic cytb marker has been developed to distinguish among several closely related invasive carp species as well as among several silver carp haplotypes (Stepien et al., 2019). Diagnostic SNPs in the cytb gene have also been used to differentiate between black and white morphs of the salamander *Proteus anguinus* (Gorički et al., 2017). Therefore, while mtDNA barcoding genes may not contain high levels of genetic variation, they may be useful for the identification of *a priori* established genetic variants.

Highly polymorphic regions of the mitogenome may be well suited for population-level inferences from eDNA due to high levels of genetic information contained within a relatively short segment of DNA. For instance, the mitochondrial control region (i.e. d-loop) exhibits a higher mutation rate than other mtDNA markers, with a higher ratio of haplotypes to individuals making it a better candidate for intraspecific diversity analyses. Similar to barcoding genes, diagnostic SNPs within the d-loop have been used to differentiate native and non-native carp populations (Uchii et al., 2016, 2017) and to identify killer whale ecotypes (Baker et al., 2018) in environmental samples. However, more detailed population genetic information may also be uncovered. For instance, the relative read abundance of d-loop haplotypes can be estimated from eDNA samples in proportions similar to frequencies in the focal population (Parsons et al., 2018; Sigsgaard et al., 2016; Tsuji, Shibata, et al., 2020).

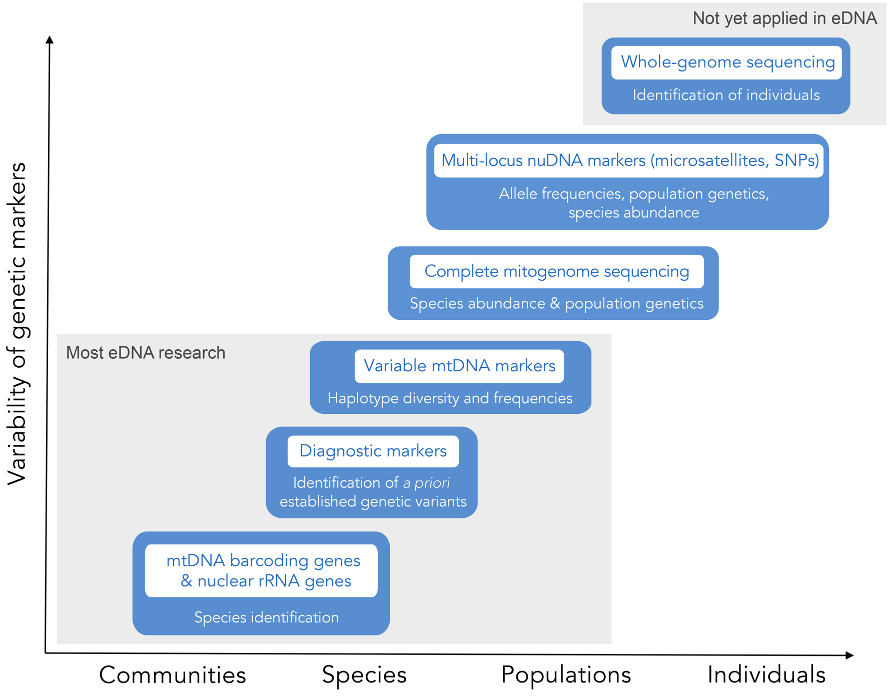
Not yet applied in eDNA
Whole-genome sequencing
Identification of individuals
Multi-locus nuDNA markers (microsatellites, SNPs)
Allele frequencies, population genetics, species abundance
Complete mitogenome sequencing
Species abundance & population genetics
Most eDNA research
Variable mtDNA markers
Haplotype diversity and frequencies
Diagnostic markers
Identification of *a priori* established genetic variants
mtDNA barcoding genes & nuclear rRNA genes
Species identification
Variability of genetic markers
Communities
Species
Populations
Individuals

FIGURE 1 Molecular genetic markers used in eDNA research situated along axes of taxonomic resolution and the genetic variation contained in the targeted genetic markers. The vast majority of eDNA research to date is aimed at characterizing communities or species using mitochondrial markers. While a few studies have investigated within-species genetic variation using mitochondrial markers, more detailed patterns of population genetic diversity may be detected using longer and more variable gene regions.

2.3 | Nuclear eDNA markers

Nuclear genetic markers provide novel opportunities to explore a wider range of ecological and evolutionary hypotheses with eDNA, but few eDNA studies have employed the amplification of nuclear markers. Because multiple unlinked loci have higher information content and, therefore, greater power to detect genetic patterns, nuDNA markers may be a better choice than mtDNA for some population-level genetic analyses from eDNA.

Nuclear rRNA genes such as 18S and ITS1 are present in multiple copies throughout the nuclear genome and have higher detectability in eDNA samples than single-copy nuDNA and, in some cases, mtDNA (Minamoto et al., 2017). Although nuclear ribosomal genes may be useful as a barcode gene for differentiating closely related species (Toju et al., 2012), these genetic regions may only provide limited insight into intraspecific genetic variation (Booton et al., 1999). However, as with mtDNA barcoding genes, diagnostic sequence variants may be identified in these genes, making them useful for the detection and quantification of a priori sequence variants in eDNA.

Very few studies have documented the ability to amplify single-copy nuDNA markers (e.g. microsatellites and SNPs) from eDNA samples (but see Andres et al., 2021; Jensen et al., 2021; Olson et al., 2012). Andres et al. (2021) reported the first detection and quantification of multi-allelic microsatellite alleles from eDNA samples using sequencing data in both laboratory and field experiments and found that allele frequencies estimated from sequencing eDNA samples with a panel of 28 microsatellite loci closely resembled allele frequencies from genotyped issues. While the per-locus information content of bi-allelic SNPs is lower than that of multi-allelic microsatellites, SNPs are abundant throughout the nuclear genome, and dozens or hundreds of SNP loci may be targeted at once. To facilitate the amplification of multiple microsatellite or SNP loci, PCR assays may be multiplexed, where several loci are co-amplified in a single reaction (De Barba et al., 2017). However, primers in multiplex PCR may interact and lead to inhibition of some loci, a problem that may worsen when starting concentrations of template DNA are variable as is expected to be the case in eDNA samples. Careful optimization of multiplex PCR is therefore recommended (Elnifro et al., 2000).

2.4 | Genome-wide genetic assessments

While most current eDNA approaches rely on PCR amplification of small sections of target genes, PCR-free approaches such as environmental shotgun sequencing may also be possible, allowing for the sequencing of genes spanning the entire mitochondrial or nuclear genome. However, the indiscriminate nature of this method makes it highly inefficient for characterizing macro-organism diversity, with non-target taxa dominating shotgun sequence reads (Stat et al., 2017). Therefore, unless highly targeted sample collection methods are used (e.g. collecting samples in footprints as in Farrell et al., 2022; Székely et al., 2021), environmental shotgun sequencing is unlikely to provide enough reads from the target species to characterize population-level diversity. Alternatively, eDNA samples may be enriched for specific genetic regions or full genomes of the target species using synthetic DNA probes designed from reference sequences (Dowle et al., 2016; Jensen et al., 2021; Wilcox et al., 2018). Although this approach may be challenging when starting concentrations of target species DNA is very low (Pinfield et al., 2019), recent work has demonstrated promise in target capture for sequencing full mitogenomes and hundreds of nuclear loci from aquatic eDNA samples (Jensen et al., 2021). Thus, while PCR-free and target enrichment methods for genome-wide genetic assessments may be possible, the efficiency and reliability of such approaches for revealing intraspecific genetic variation from eDNA warrant further exploration.

In addition to sequencing short fragments, it may be possible to recover sequence data from entire mitochondrial genomes from eDNA samples (Deiner, Renshaw, et al., 2017). However, although estimating full-mitogenome variation from eDNA is possible, no study has yet explored the potential to estimate population genetic parameters from the resulting sequence data. Nonetheless, this area of research remains promising, as the sequencing of full mitogenomes indicates the persistence of intact organelles or cells, illuminating the possibility that entire cells could be isolated from environmental samples and sequenced individually. Sequencing of single cells would allow for the generation of multi-locus genotypes or whole genomes to be linked to the same individual, with the ability to genetically distinguish individuals from one another within a mixed eDNA sample.

3 | ESTIMATING POPULATION GENETIC DIVERSITY USING eDNA

The detection and quantification of intraspecific genetic variants derived from eDNA sequencing data allow for population genetic analyses to be conducted on natural populations without capturing individuals of the target species. However, an important consideration for all analyses of genetic variation from eDNA is that genetic variation is aggregated at the population level, representing a mixture of genotypes from an unknown number of genetic contributors (but see Section 4 below; Dugal et al., 2021), with no ability to assign multi-locus genotypes to individuals. Most existing population genetic models, statistical software and analytical frameworks are based on the availability of individual-level genotypes, which are not readily available from eDNA sequence data. Therefore, even if existing programs to estimate population genetic parameters can handle the intraspecific genetic datasets produced through the analysis of eDNA (e.g. haplotype frequencies), the assumptions and limitations of each analysis should be thoroughly investigated to ensure the robustness of the resulting population genetic inferences.

3.1 | Genetic diversity from mtDNA

The ability to recover several mtDNA haplotypes from eDNA allows for analyses to be conducted on the presence of different haplotypes or haplotype relative abundance (e.g. frequency). With data on the detection of haplotypes, parameters such as haplotype diversity, nucleotide diversity and segregating sites may be estimated for each sampled population. Rarefaction curves for the detection of haplotypes as a function of sample size may also be useful for predicting the total amount of genetic diversity at a site as well as the sampling effort that would be required to completely sample the diversity (Székely et al., 2021).

Haplotype frequencies from eDNA data may be inferred directly from eDNA sequence read frequencies, under the assumption that sequence abundance accurately reflects the abundance of the haplotype in the population. However, due to stochastic sampling, variable contributions of DNA from different individuals and biases introduced during PCR or sequencing, this assumption may not hold true in many cases. More conservative alternatives to directly quantifying haplotype frequencies include characterizing ranks of haplotype frequencies (Turon et al., 2020) or the presence/absence of haplotypes across replicated samples (Azarian et al., 2021). Regardless of how haplotype frequencies are estimated, the data may be used to construct haplotype networks to illustrate the haplotype differences among sites or populations. Analysis of molecular variance (AMOVA) may also be conducted to determine the amount of genetic variation associated with each level of hierarchical organization of sampled sites. Other population genetic parameters, such as Fst and Tajima's D, may be calculated from population-level haplotype frequencies (Shum & Palumbi, 2021; Weitemier et al., 2021). However, such metrics are sensitive to haplotype richness and frequencies, and an investigation of how they are impacted by the specific approach used to characterize haplotype frequencies from eDNA samples is needed.

An exciting avenue of research capitalizes on the understanding that eDNA samples contain a mixture of DNA from many species, making possible the evaluation of mtDNA diversity from multiple species simultaneously using eDNA metabarcoding approaches (Shum & Palumbi, 2021; Stat et al., 2017; Turon et al., 2020; Weitemier et al., 2021). Although identifying mtDNA haplotypes for entire communities may present additional challenges over single-species approaches, it may soon be possible to monitor genetic diversity for hundreds of species across the tree of life using eDNA sequence data.

3.2 | Genetic diversity from nuDNA

As with mtDNA haplotypes, the estimation of nuDNA allele frequencies from eDNA sequences may be conducted using quantitative (read frequencies) or semi-quantitative (read frequency rank; presence/absence of alleles in replicates) approaches. Once allele frequencies are estimated, several population genetic characteristics may be quantified. For instance, allelic richness, expected heterozygosity and the number of private alleles can be estimated and compared among populations, and population genetic structure may be determined using AMOVA or principal component analysis of allele frequencies. Genetic distances among populations may also be calculated using the allele frequency distance (AFD), a metric that is highly correlated with common genetic differentiation metrics such as Fst but does not require individual genotypes (Berner, 2019).

If several independent SNPs can be amplified from eDNA samples, population-level genetic analysis of SNPs may be possible through the use of analytical frameworks developed for pooled sequencing (Pool-seq) approaches (Gautier et al., 2013; Hivert et al., 2018; Kofler et al., 2011). Pool-seq approaches involve sequencing the pooled DNA from many individuals, a cost- and time-effective alternative to individual genotyping. Analytical frameworks have been developed to estimate allele frequencies and population genetic parameters such as Fst from Pool-seq data, and several pipelines for the analysis of pooled sequencing data are already available (Kofler et al., 2011). Most Pool-seq analytical frameworks are developed only for bi-allelic loci, and while useful for the analysis of SNPs, these frameworks are not applicable to multi-allelic markers such as microsatellites.

Although a single SNP may not contain high information content, several SNPs may be present within a short gene region that can be amplified and sequenced to generate multi-allelic markers. These ‘microhaplotype’ markers contain more information content per locus, allowing for more powerful genetic analyses to be conducted using fewer loci (Baetscher et al., 2018; Kidd et al., 2014). However, such multi-allelic microhaplotype markers may not be applicable in most Pool-seq analysis pipelines, many of which assume independence among loci, an assumption that is violated when SNPs are located in close proximity (Slatkin, 2008). In this sense, analysis of microhaplotype loci may be more similar to that of microsatellite loci.

While analysis of highly variable mitochondrial and nuclear genetic markers is not yet common in eDNA research, eDNA approaches tailored for this purpose offer an opportunity to study detailed population genetic patterns, potentially with higher sensitivity at a lower cost. However, validation experiments and special considerations will be necessary to fully understand the potential and limitations of population genetic analysis of eDNA samples. In a later section, we summarize the importance of considering false positive and false negative detections and the resulting impact on the precision of estimates of population genetic parameters.

4 | ESTIMATING ABUNDANCE USING eDNA

4.1 | eDNA concentration as a metric of abundance

In addition to the molecular detection of genetic diversity, eDNA approaches may be useful for estimating organismal abundance, a metric that is central to species monitoring, management and conservation (Yoccoz et al., 2001). To date, most efforts to assess species abundance with eDNA have focused on correlating ambient species-specific eDNA concentration or metabarcoding read counts with numerical abundance, with several studies reporting a positive correlation between the two metrics (Baldigo et al., 2017; Doi et al., 2017; Lacoursière-Roussel et al., 2016; Pilliod et al., 2013; Schmelzle & Kinziger, 2016; Shelton et al., 2022). However, this relationship may be confounded by variation in eDNA concentration due to biotic and abiotic factors influencing DNA production and loss rates as well as biases introduced through the use of different eDNA sampling, processing and amplification approaches (Beng & Corlett, 2020; Shelton et al., 2023).

The amount of eDNA captured in a sample is the collective result of several complex processes influencing eDNA transport, decay, settling and resuspension, all of which may occur at different rates under different environmental conditions including temperature, pH, UV exposure, microbial activity and substrate type (Barnes & Turner, 2015; Carraro et al., 2018; Shogren et al., 2017; Strickler et al., 2015). The production rate of eDNA can also vary among source individuals as a function of their size, behaviour or metabolism, further obscuring the relationship between eDNA concentration and organism abundance (Dunn et al., 2017; Klymus et al., 2015; Maruyama et al., 2014). Although studies have attempted to account for some of these factors (e.g. through allometric scaling; Stoeckle et al., 2021; Yates et al., 2022), these efforts largely remain taxa- and ecosystem-specific, and the extent to which they are generalizable to other taxonomic groups or ecosystems is not known. Without accounting for these sources of variation, eDNA concentration and/or metabarcoding read counts may not accurately reflect the abundance of the focal species.

The type and size of the molecular marker targeted may also impact variation in eDNA copy numbers and distort the link between DNA concentration and species abundance. For instance, while the number of nuDNA sequences per cell does not vary among source tissue types, mtDNA copy number per cell can be highly variable within and among individuals (Long & Dawid, 1980; Robin & Wong, 1988). As discussed above, the decay rate of eDNA depends on the target fragment length and source (i.e. mtDNA or nuDNA), which can impact the detection and yield of target species DNA (Jo et al., 2020). Measurements of eDNA concentration may also be sensitive to eDNA sampling and processing approaches including volume filtered, filter type, preservation buffer, filter membrane type, extraction method, quantification approach (i.e. quantitative PCR, digital droplet PCR) and primer specificity (Goldberg et al., 2016). Although there have been attempts to improve the methodological and reporting standards in eDNA research, the lack of standardized protocols and analytical frameworks for evaluating the quantity of eDNA prevents comparisons of eDNA-based species abundance estimates among studies and ecosystems (Loeza-Quintana et al., 2020).

While several studies have improved our understanding of eDNA dynamics under different environmental conditions and protocols, it is often not feasible to account for all of the potential sources of variation in eDNA concentrations, particularly in complex natural environments. A meta-analysis of the studies reporting correlations between the concentration of species-specific eDNA particles and the density or biomass of a species found that on average, eDNA particle concentration accounted for only 57% of the observed variation in species abundance in natural systems (Yates et al., 2019). Understanding and accounting for the sources of variation in eDNA concentration will be a requisite step in obtaining robust estimates of species abundance based on DNA copy number or metabarcoding read counts. However, in the section below, we propose an alternative to eDNA concentration-based approaches through the use of alternative analytical frameworks that incorporate information about genetic diversity in eDNA samples to estimate species abundance.

4.2 | Genetic variation as a metric of abundance

Rather than focusing on the concentration of target species eDNA, the amount of genetic variation contained in an eDNA sample may be used to assess species abundance by estimating the absolute number of individuals contributing DNA to the mixture. In the simplest sense, the minimum number of individuals detected in an eDNA sample is equivalent to the number of individuals required to produce the observed set of genetic variants. That is, the number of genetic contributors to a DNA mixture must be at least the number of detected genetic variants at a locus divided by the ploidy of the genetic marker (Carreon-Martinez et al., 2014). This approach, termed the maximum allele count, is useful for providing a lower bound on the number of genetic contributors required to explain the observed set of haplotypes or alleles. If haplotype diversity in the target gene region is near 1 (i.e. nearly every individual in the population exhibits a unique haplotype), species abundance may be accurately estimated by simply counting the number of haplotypes identified from eDNA (Yoshitake et al., 2019, 2021). However, this does not account for instances where multiple individuals share the same haplotype, and thus, tends to underestimate the number of contributors in mixed DNA samples, a bias that worsens as the number of individuals increases (Haned et al., 2011; Paoletti et al., 2005). This problem is particularly severe when targeting genetic markers that exhibit a highly skewed frequency spectrum, as it is more likely that sampled individuals will contain redundant genotypes.

The problem of redundant alleles may be addressed through the use of likelihood-based DNA mixture models (Egeland et al., 2003; Haned et al., 2011; Sethi et al., 2019), where the observed alleles in a mixture and the allele frequencies of the focal population are used to calculate the likelihood that a putative number of individuals (x) produced an observed set of n alleles, $A = \{a_1, \ldots, a_n\}$, given the associated population allele frequencies, $p = \{p_1, \ldots, p_n\}$, of the observed alleles. This model accounts for all possible combinations of alleles that may arise in a DNA mixture, including the copies of allele a_i truly present in the mixture (g_i) while addressing 'masked' copies of the allele (d_i). At a single locus (j) of ploidy $k = 1$ or 2 (for mitochondrial or nuclear markers, respectively), this likelihood can be calculated as:

$$L_j(x|A,p) = \sum_{d_{1=0}}^{d} \sum_{d_2=0}^{d-d_1} \cdots \sum_{d_{n-1}}^{d-d_1-\ldots-d_{n-2}} \left[\left(\frac{(kx)\,!}{\prod_{i=1}^{n} g_i\,!} \right) \prod_{i=1}^{n} p_i^{g_i} \right] \quad (1)$$

where the total number of masked alleles in a mixture is $d = kx - n$ and for any specific observed allele $d_i \in \{0, \ldots, kx - n\}$, and where the true total number of allele copies in the mixture is $\sum_{i=1}^{n} g_i = kx$, with $g_i = 1 + d_i$ for any specific observed allele. This likelihood can be calculated and multiplied across any number of unlinked loci, with the estimated number of individuals determined by the maximum likelihood estimate across any given number of putative contributors. Although this approach originated in the forensic sciences, Sethi et al. (2019) extended the applications of DNA mixture models into ecological frameworks, demonstrating the performance of the model using simulated mixtures of different nuclear markers and assessing predation rates by estimating the number of prey items through the analysis of predator stomach contents. Andres et al. (2021) subsequently demonstrated potential for estimating abundance using DNA mixture models from microsatellite alleles detected in eDNA samples from experimental mesocosms.

Estimating abundance using genetic diversity minimizes the challenge of variable DNA production rates within and among individuals by estimating the number of distinct individuals contributing DNA to an environmental sample. Because it does not rely on DNA copy number, this approach is not strongly influenced by differences in eDNA production due to organism body size, behaviour, metabolism or genetic marker type (although eDNA production rates may influence detection probability; see Section 4.4 below). While Sethi et al. (2019) and Andres et al. (2021) successfully used DNA mixture models to estimate the number of genetic contributors when the total number of individuals is small (≤10), the limitations of this approach have not been thoroughly explored in highly complex mixtures of greater numbers of individuals. In the section below, we explore the ability of DNA mixture models to resolve mixtures of large numbers of individuals to better understand the applicability of this approach in natural systems, opening the door for future avenues of research and application.

4.3 | Simulating DNA mixtures with real genotypes

To advance our understanding of the performance of DNA mixture models in eDNA applications, we simulated mixtures of DNA made up of up to 100 putative contributors to reflect local numbers that may be experienced in ecological studies on abundant taxa. In addition to testing the performance of DNA mixture models under large numbers of putative contributors, we explored model performance using a suite of genetic marker types including nuclear markers (SNPs and microsatellites) as well as a novel application of the model to haploid mtDNA markers. To simulate DNA mixtures, we sampled genotypes from a published dataset that contains microsatellite and SNP marker data for 1129 linesnout gobies (*Elacatinus lori*) collected from 35 sites across the Belize Barrier Reef (for detailed collection, laboratory and sequencing methods, see D'Aloia et al., 2020). We also used mitogenome sequence data (~500 bp contigs) that were collected for each of these individuals (unpublished data; see Appendix S1 for methods). Our goal was to generate artificial DNA mixtures using combinations of real genotypes for different molecular marker types, numbers of loci and numbers of individuals used to calculate population allele frequencies.

Due to the high reported levels of genetic structuring in this region, we selected only 127 individuals from four sites comprising one of the major subpopulations detected in D'Aloia et al. (2020). Furthermore, the DNA mixture abundance model is combinatorial in nature and estimating the number of contributors can become prohibitively computationally intensive with large numbers of alleles at loci. We, therefore, excluded highly multi-allelic microsatellite loci (i.e. >20 alleles) from the dataset to reduce computation times in analysing simulated mixtures. To simulate the behaviour of the DNA mixture model using a large and highly variable mtDNA fragment, we assembled eight mitochondrial sequence contigs to generate haplotype data for a 4 kB mtDNA fragment, where any sequence variations within the 4 kB region were designated as a unique mitochondrial haplotype. The final dataset included individual genotypes for 44 microsatellite loci, 256 nuclear SNP loci and mitochondrial haplotype information for 127 individuals.

The DNA mixture model requires an observed list of alleles (A) from a DNA mixture sample and their associated population allele frequencies (p) as inputs in estimating sample abundance. We generated DNA mixture genotypes by randomly sampling n individuals from each molecular marker dataset, with n ranging from 2–100 individuals (i.e. in silico mixtures made up of 2–100 contributors). Genotypes from sampled individuals were combined to generate a list of observed alleles (A) as a simulated DNA mixture. Accuracy of the DNA mixture model may depend on the presence of rare alleles, whereby rare alleles are generally strongly informative in characterizing mixture abundance (Sethi et al., 2019). By their nature, however, characterizing the frequencies of rare alleles from samples of specimens from a population may be challenging, and many individuals may need to be genotyped to detect the rarest alleles. We, therefore, explored the impact of assessing allele frequencies from relatively small (25) or large (100) reference sets of individuals.

The number of loci used in the mixture may also influence model accuracy (Sethi et al., 2019). We estimated the number of individuals in each DNA mixture using different numbers of nuclear loci to represent a relatively small vs. large marker panel (10 vs. 44 loci for microsatellites; 64 vs. 256 loci for SNPs). Because mitochondrial DNA haplotypes were generated by concatenating multiple contigs, we ran the DNA mixture model using mitochondrial haplotypes from a single mitochondrial contig (480 bp) in addition to the full 4 kB fragment. In sum, the DNA mixture model was used to estimate the number of contributors in simulated mixtures ranging in size from 2 to 100 individuals for each of three marker types (microsatellites, SNPs and mitochondrial), with population allele frequencies specified using different numbers of individuals (25 or 100), using either a small or large panel/fragment size for each marker type (Table 1). Each condition was simulated 100 times.

A challenge with mixtures made up of large numbers of contributors is that loci may 'saturate' insofar that all possible alleles manifest in a sample. In such cases when all possible alleles are observed at a locus, the likelihood that a putative number of contributors produces an observed set of alleles will monotonically increase, and, when all alleles are observed across all loci, a maximum likelihood estimate cannot be reached (Egeland et al., 2003). Furthermore, because the likelihood is taken as a product across all loci, the maximum likelihood may be upwardly biased when some loci exhibit a monotonically increasing likelihood (i.e. all alleles are observed), even if other loci exhibit a single maximum likelihood (Figure S1). To address these maximum likelihood estimation challenges for large mixtures, we removed any microsatellite loci for which all alleles from the reference set were observed when estimating the number of contributors to a sample DNA mixture. While this filtering step works well for multiallelic loci such as microsatellites, filtering out saturated loci was not possible for SNPs, as both alleles for most loci are frequently observed even in small mixtures of DNA. This was also not possible for mitochondrial haplotype data, as only a single locus was used.

With these simulations, we show that the number of genetic contributors can be estimated in complex DNA mixtures of up to 100 individuals, although the accuracy and precision of the estimate varied depending on the marker type, size of the marker panel and frequency of the rarest alleles. Across all three genetic markers, the accuracy of the estimated number of genetic contributors to DNA mixtures is greatest when allele frequencies are estimated using 100 individuals compared with 25 individuals (Table 1; Figure 2), corresponding to higher resolution allele frequency estimates that can accommodate informative rare alleles (see inset allele frequency distributions). When fewer rare alleles are present in the population and, therefore, in the DNA mixtures, the model systematically underestimates the number of contributors to mixtures, a problem that worsens as True *N* increases. The size of the marker panel does not have a strong impact on the average bias of the contributor estimation, although the precision of the estimate is greater when more loci or longer mtDNA fragments are used (top panel in Figure 2).

All marker types could resolve mixtures of up to 25 contributors, which may be sufficient for many ecological applications. However, to approach larger mixtures, the choice of genetic marker type influences the performance of the DNA mixture model. In this dataset, microsatellite markers show the greatest promise for accurately estimating the number of individuals in very large DNA mixtures, with average estimates within two individuals when True *N* = 100 (Table 1; top left panel in Figure 2). With allelic richness of up to 20 alleles per locus, both 44-locus and 10-locus marker panels could estimate up to 100 individuals in a DNA mixture, although the number of saturated (i.e. 'failed') loci steadily increases as True *N* increases. While large panels of SNPs can accurately resolve mixtures of up to 25 individuals, the impact of monotonically increasing loci (i.e. loci in which both alleles are observed in the mixture) is apparent in larger mixtures, resulting in an upward bias in the estimated number of individuals (middle panel). The single mitochondrial fragment of 4 kB used in these simulations contains enough variation to provide accurate estimates of up to 40 individuals, but mixtures larger than this exhibit high numbers of haplotypes and the DNA mixture approach becomes prohibitively computationally intensive (top right panel). A smaller mitochondrial fragment can also resolve DNA mixtures of up to 40 individuals, but beyond this

TABLE 1 Mean bias (estimated *N* − true *N*, ±1 s.d.) in the estimated number of contributors in 100 simulated DNA mixtures of True *N* = 10, 40 and 100 individuals.

Marker type	No. of individuals for allele freq.	No. of loci[a]	No. of alleles[b]	Mean bias *N* = 10	Mean bias *N* = 40	Mean bias *N* = 100
Microsatellite	100	44	559	0.1 ± 0.7	2.5 ± 3.5	1.8 ± 8.4
	25	44	382	−1.5 ± 0.7	−17.1 ± 1.8	−65.1 ± 3.9
	100	10	130	0.0 ± 1.3	2.3 ± 6.3	−0.7 ± 15.4
	25	10	93	−1.6 ± 1.2	−13.8 ± 4.8	−57.8 ± 6.8
SNP	100	256	465	0.3 ± 1.6	8.0 ± 8.5	36.2 ± 11.5
	25	256	422	−1.6 ± 1.2	−7.5 ± 5.5	−28.8 ± 12.0
	100	64	117	0.8 ± 2.9	13.3 ± 19.3	15.4 ± 22.4
	25	64	113	1.7 ± 3.3	10.6 ± 12.6	−37.0 ± 0.0
Mitochondrial	100	8	39	−0.4 ± 1.6	1.2 ± 8.6	—
	25	8	20	−4.3 ± 1.9	−19.9 ± 4.7	−47.0 ± 11.7
	100	1	24	−0.3 ± 3.3	1.5 ± 12.4	−68.0 ± 29.8
	25	1	9	−2.3 ± 3.2	−18.8 ± 9.5	−62.1 ± 5.8

Note: Mixtures were constructed using *Elacatinus lori* genotypes for microsatellites, SNPs and mitochondrial haplotypes. The number of individuals in each mixture was estimated using relatively large or small marker panels and population allele frequencies were estimated using either 100 or 25 individuals.

[a] Numbers of loci for mitochondrial data represent the numbers of contigs used to generate mitochondrial haplotypes rather than independent loci.

[b] Alleles for SNP data exclude fixed alleles.

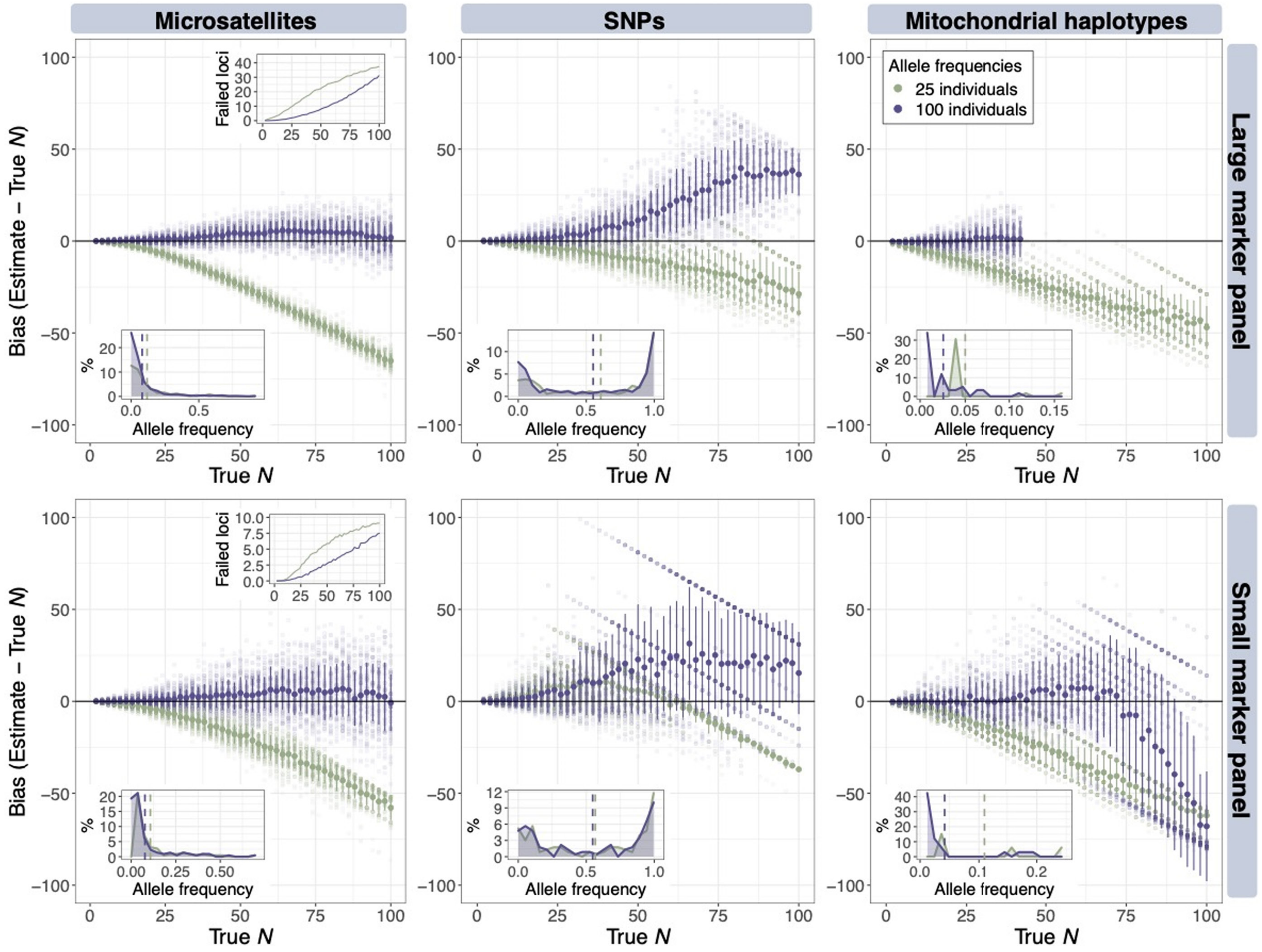

FIGURE 2 Bias in the estimated number of contributors from mixtures of DNA containing genotypes from 2 to 100 sampled individuals (True N) using relatively large (top row) and small (bottom row) marker panels. Transparent circles show individual DNA mixture simulation results, and solid points and lines show the mean bias ±1 SD across 100 simulated mixtures for each True N. Mixtures were generated for multi-allelic microsatellite markers (left column), bi-allelic SNP markers (middle column) and single-locus mitochondrial haplotypes (right column). For each marker type, population allele frequencies were calculated using either 25 (purple) or 100 (green) individuals from the full dataset. Inset figures illustrate the distribution of allele frequencies calculated from 25 or 100 individuals, with mean allele frequencies denoted with vertical dashed lines. For microsatellite markers, the inset figure in the top right displays the number of loci for which all alleles were observed as True N increases. These saturated loci were removed from the maximum likelihood calculation.

point, most haplotypes are present in all mixtures, and the precision of the estimate is greatly reduced (lower right panel).

4.4 | Using DNA mixture models in future eDNA research

Efforts to correlate eDNA concentration with species abundance have proven to be moderately successful in providing an index of relative species abundance, which may be useful for characterizing the distribution of organisms based on relative abundances across large spatial or temporal scales (Shelton et al., 2022). However, an alternative approach to estimating species abundance using the amount of intraspecific genetic diversity detected in eDNA samples may account for variation in eDNA production rate within and among individuals, possibly leading to more accurate estimates of absolute abundance at local scales. We find that DNA mixture models can estimate the number of individuals contributing to mixtures of DNA of up to 100 individuals, provided enough genetic diversity and rare alleles are present in the mixtures. As reported by Sethi et al. (2019), the presence of rare alleles in a DNA mixture was more important than the number of loci for accurately estimating the number of individuals. Microsatellites or mitochondrial markers are the best marker choices if the expected number of individuals in a mixture of DNA is high because the per-locus information content is much higher than that of bi-allelic SNPs, but all marker types are able to accommodate smaller mixtures of up to 25 individuals.

Although the simulations described here are important for understanding the limitations and biases in the DNA mixture model, additional research and considerations may be required when using this approach to estimate abundance in eDNA samples. Importantly, our simulated mixtures combined genotypes of sampled individuals with perfect detection of all genetic variants in a mixture. However, imperfect detection of alleles is expected in eDNA samples, particularly when targeting nuDNA markers, and the non-detection of alleles may result in the downward bias of abundance estimates (Sethi et al., 2019). Maximizing the detection of rare alleles may be possible through the improvement of eDNA collection and laboratory approaches including collecting larger volumes of water, target enrichment or the development of extraction, amplification and sequencing protocols designed to maximize the recovery of all genetic variation in eDNA samples. Statistical frameworks designed to account for imperfect species detections may also be implemented to account for the imperfect detection of alleles, with models able to evaluate eDNA capture probability and detection probability under different environmental conditions (Burian et al., 2021). Another consideration is that the DNA mixture model presented here requires independence among loci and alleles within and among contributors, and we tested the model using simulated mixtures from genetic datasets from a population with a very large effective population size (Ne). The limitations of the DNA mixture model with regard to small Ne, deviations from Hardy–Weinberg, and linkage disequilibrium should be explored before the model is applied to estimate species abundance in populations exhibiting those conditions. We also tested the DNA mixture model on nuclear and mitochondrial markers for diploid organisms, and the formulation and behaviour of the model for polyploid organisms is an avenue of research that has not yet been explored.

When using genetic variation in mitochondrial markers to estimate species abundance, it is important to examine possible heteroplasmy, where an individual exhibits more than one mtDNA variant. In instances where the minor alleles of heteroplasmic mitochondrial markers are detectable in eDNA, the number of contributors in mixed DNA samples may be overestimated (Nakanishi et al., 2020). Similarly, any false alleles introduced through PCR error, sequencing error, non-target amplification or sample contamination will artificially inflate the estimated number of individuals in a mixture. Because rare alleles provide strong evidence about the presence of many individuals in a mixture, the ability to discern between true low-frequency alleles and spurious sequences becomes increasingly important as the true number of contributors increases. Other frameworks developed in criminal forensics research have been able to incorporate information about the magnitude of observed alleles in a mixture (e.g. read counts) as well as model errors such as PCR stutter to improve the interpretation of DNA mixtures (Paoletti et al., 2011; Swaminathan et al., 2015). Such approaches may be also useful for improving the detection of rare alleles in eDNA mixtures.

The results of the simulated DNA mixtures presented here demonstrate that accurate estimates of the number of individuals contributing to eDNA samples requires a relatively large panel of highly multi-allelic markers. The detection of rare (low-frequency) alleles is important for providing information about the presence of individuals in a mixture of DNA. Thus, if large numbers of individuals are expected in a mixture, population allele frequencies should be estimated using as many individuals as is practically possible to increase the number of rare alleles that may be detected and to reduce the number of saturated loci that must be dropped from the calculation. To resolve DNA mixtures of more than 100 individuals, genetic markers with a greater number of alleles, and, therefore, a greater number of rare alleles, will be required. However, applications of DNA mixture models for very large numbers of individuals may be limited as the calculation becomes impractical when large numbers of haplotypes are observed in mixtures of DNA. We recommend that anyone wishing to employ DNA mixture models to estimate abundance from eDNA use a simulation approach as described here to understand the biases and limitations associated with the specific marker panel employed.

5 | UNDERSTANDING AND ACCOUNTING FOR ERRORS IN eDNA DATA

Regardless of the length or type of marker selected, a requisite step in any eDNA study is the detection of spurious DNA sequences to distinguish authentic DNA sequences from errors. This step becomes even more important in the analysis of intraspecific genetic variation from eDNA, as the detection and quantification of exact sequence variants (e.g. haplotypes or alleles) is the primary goal. At the same time, eDNA samples often contain DNA in low quantities and quality, and the resulting sequences may be more prone to errors than those from high-quality DNA. If not removed, such erroneous sequence variants may have substantial implications for downstream population genetic analyses and abundance estimates from eDNA. Therefore, understanding the consequences of sequencing errors and testing approaches to remove errors while retaining true low-frequency genetic variants presents a priority area for statistical and bioinformatics applications in eDNA research.

5.1 | Causes and consequences of errors in eDNA sequence data

Addressing erroneous sequence variants in eDNA data first requires an understanding of how such errors arise. Furlan et al. (2020) provide a comprehensive overview of errors in metabarcoding workflows and classify them as either contamination errors or misidentification errors that can be generated during sample collection, laboratory processing, sequencing or bioinformatics processes. Because errors in eDNA data can arise from multiple complex and potentially interacting origins, the nature and severity of errors may differ depending on the specific study conditions. Interpreting the detection of intraspecific genetic variation in eDNA metabarcoding data requires particular consideration of the causes and consequences of such errors in different contexts.

Contamination errors and misidentification errors (i.e. erroneous DNA sequences or incorrect sequence assignment) generally lead to false positive detections and an overestimation of intraspecific genetic diversity in a sample, with implications for downstream analyses of genetic diversity, differentiation and species abundance estimates. On the other hand, rare genetic variants that are present at very low frequencies in eDNA could be missed during field sampling or laboratory procedures, leading to false negatives and an underestimation of genetic diversity. Stringent protocols to filter out erroneous sequences can also result in the removal of true DNA fragments, particularly if the abundance of sequences from rare genetic variants does not exceed the abundance of erroneous sequences. While the sensitivity and efficiency of different assays should be investigated to determine bioinformatic thresholds, a trade-off between decreasing the rate of false positives and increasing the rate of false negatives may persist. Several approaches to detect and remove PCR and sequencing errors have been developed (Schnell et al., 2015), but a better understanding of these specific processes may be needed to increase confidence in delineating genetic variation within species from eDNA.

5.2 | Identification and removal of erroneous sequences

Removing contamination and misidentification errors in eDNA data may be accomplished through the use of bioinformatic thresholds and algorithms designed to identify and eliminate sequences that are likely to be erroneous. The selection of specific bioinformatic parameters may vary by study and should be made with the study goals, data type and implications of errors in mind.

Erroneous sequences will likely be present at low frequencies compared with most authentic sequences in eDNA samples and setting minimum abundance thresholds to remove low-frequency sequence variants may be a straightforward approach to filtering out errors. Several studies use the estimated error rate of a particular gene region or sequencing platform as a threshold that any given sequence variant must exceed to be retained in the dataset (Sigsgaard et al., 2016; Stat et al., 2017). If using multiple loci with variable numbers of alleles per locus, a variable threshold based on per-locus allelic richness may be required to obtain more accurate allele frequency estimates, as erroneous sequences may consume a greater percentage of reads when the genetic complexity of a sample is low (Andres et al., 2021). Other threshold-based approaches may include retaining only the alleles that are observed in multiple field or laboratory replicates, under the assumption that the probability of random errors is low and observing any allele in multiple replicates becomes exceedingly rare. However, for the same reason, the probability of detecting low-frequency genetic variants in multiple replicates is low and such thresholds may exclude true genetic variants.

While threshold-based exclusion of sequences can effectively remove low-frequency errors, several pipelines have been developed to correct sequencing errors and determine real biological sequences, also referred to as amplicon sequence variants (ASVs). Such denoising approaches, implemented in algorithms such as DADA2 (Callahan et al., 2016), UNOISE2/3 (Edgar, 2016) and Deblur (Amir et al., 2017), allow for more refined detection and removal of spurious sequences. Analyses of bioinformatic pipelines have shown DADA2 to be highly effective in removing sequencing errors while retaining true ASVs (Macé et al., 2022; Tsuji, Maruyama, et al., 2020; Tsuji, Miya, et al., 2020), although UNOISE3 may be preferable for coding genes such as COI (Antich et al., 2021). Other approaches to determine erroneous sequences include an examination of entropy changes in the different codon positions, an approach that may be used if variation in coding sequences is being assessed (Turon et al., 2020). For nuclear genetic markers to become more useful in eDNA research, bioinformatic strategies to identify and remove errors specific to nuDNA (e.g. microsatellite stutter) may be needed.

To reduce the possibility of misidentification errors such as the amplification of non-target species, the application of a sequence similarity threshold or alignment algorithm to known (subject) sequences may be desirable (e.g. Smith-Waterman; BLAST; Altschul et al., 1990; Smith & Waterman, 1981). However, to maximize confidence that the variation detected in eDNA is biologically real, it may be beneficial to restrict the list of alleles to only those known from high-quality tissue-based genotyped samples (as in Parsons et al., 2018; Shum & Palumbi, 2021). Such a conservative approach will remove nearly all errors due to PCR/sequencing error or non-target amplification but may result in an underestimate of genetic diversity, as any genetic variants that have not been previously identified will not be retained. This approach is, therefore, best suited for studies where large numbers of individual tissues have been genotyped and rare genetic variants have been specified.

5.3 | Minimizing the effects of errors in eDNA analyses

Some of the challenges with distinguishing true genetic variation from background noise may be addressed by minimizing erroneous sequences and increasing the quality and quantity of target DNA captured in environmental samples. For instance, novel filtration approaches allowing for the collection of large sample volumes (e.g. up to 3000 L of water) have exhibited improved detection of rare species (Sepulveda et al., 2019), and similar methods will likely improve the detection of rare genetic variants from eDNA. Sampling at the locations or times of the year where focal species are known to occur may also be employed to increase the concentration of target species' DNA in a sample, as in the case of sampling in the wake of marine mammals (Baker et al., 2018; Parsons et al., 2018) or during known feeding or spawning aggregations (Sigsgaard et al., 2016). Further improvements to eDNA extraction, amplification and sequencing efficiency may also be able to overcome some of the technical limitations of detecting intraspecific genetic variation in eDNA, and the potential for such technological improvements should be explored.

Even if all measures are taken to prevent, identify and remove spurious DNA sequences, it may remain difficult to distinguish between erroneous sequences and true low-frequency alleles in the population using eDNA approaches. In such cases, sequencing mock communities containing known levels of intraspecific genetic variation may allow for the estimation of the frequency of sequence artefacts and non-detection of rare genetic variants (Miller et al., 2002; Taberlet et al., 1996). The impact of false positives, false negatives and allele frequency misspecifications on the estimation of population genetic parameters from eDNA data must also be investigated. For instance, several studies have examined the consequences of sequencing and genotyping errors on biological inferences from tissue-based sequence data, revealing the bias in genetic parameters induced under different numbers and types of errors (Burian et al., 2021; Hivert et al., 2018; Pompanon et al., 2005). Similar approaches will be useful for understanding how different errors may lead to biased estimates of population genetic parameters from eDNA sequence data which can include genetic information from multiple individuals.

6 | CONCLUSION

The field of eDNA has made substantial progress in recent years, yet environmental samples contain much more genetic information than is currently analysed in most eDNA research. Recent studies have identified intraspecific genetic variation from short mitochondrial markers in eDNA sequence data, yet obtaining higher-resolution genetic information may be feasible by targeting full mitogenomes or nuclear genetic markers. These variable genetic markers may open up possibilities to conduct detailed population genetic analyses and estimate species absolute abundance from eDNA, with the potential to inform questions in ecology, evolutionary biology and conservation management. Although the analysis of intraspecific genetic variation in eDNA samples may introduce additional challenges over traditional tissue-based population genetics approaches, further research on the properties of different genetic markers and the development of bioinformatic and analytical pipelines may allow for robust community-wide population genetic analyses from eDNA.

ACKNOWLEDGEMENTS

This project was supported by funding awarded to the Cooperative Institute for Great Lakes Research (CIGLR) through the National Oceanic and Atmospheric Administration (NOAA) Cooperative Agreement with the University of Michigan (NA17OAR4320152; CIGLR contribution #1215). Additional funding was provided by the Department of Defence (DoD) Strategic Environmental Research and Development Program (SERDP) (grant number RC19-1004) and the Cornell Atkinson Center for Sustainability. We thank S.M. Bogdanowicz, P. Buston and C. D'Aloia for providing the genetic dataset used for DNA mixture simulations.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

This manuscript applies a novel formulation of the haplotype likelihood and removal of saturated loci in the DNA mixture model. All datasets and code are publicly available at https://github.com/karaandres/eDNA_contributor_estimations.

ORCID

Kara J. Andres https://orcid.org/0000-0003-4822-7047
Suresh A. Sethi https://orcid.org/0000-0002-0053-1827

REFERENCES

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.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Zech Xu, Z., Kightley, E. P., Thompson, L. R., Hyde, E. R., Gonzalez, A., & Knight, R. (2017). Deblur rapidly resolves single-nucleotide community sequence patterns. *MSystems*, *2*(2), e00191-16.

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.

Anne, C. (2006). Choosing the right molecular genetic markers for studying biodiversity: From molecular evolution to practical aspects. *Genetica*, *127*(1), 101–120.

Antich, A., Palacin, C., Wangensteen, O. S., & Turon, X. (2021). To denoise or to cluster, that is not the question: Optimizing pipelines for COI metabarcoding and metaphylogeography. *BMC Bioinformatics*, *22*, 1–24.

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.

Baetscher, D. S., Clemento, A. J., Ng, T. C., Anderson, E. C., & Garza, J. C. (2018). Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. *Molecular Ecology Resources*, *18*(2), 296–305.

Baker, C. S., Steel, D., Nieukirk, S., & Klinck, H. (2018). Environmental DNA (eDNA) from the wake of the whales: Droplet digital PCR for detection and species identification. *Frontiers in Marine Science*, *5*, 133.

Baldigo, B. P., Sporn, L. A., George, S. D., & Ball, J. A. (2017). Efficacy of environmental DNA to detect and quantify brook trout populations in headwater streams of the Adirondack Mountains, New York. *Transactions of the American Fisheries Society*, *146*(1), 99–111.

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*(4), 729–744.

Barnes, M. A., & Turner, C. R. (2015). The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, *17*(1), 1–17. https://doi.org/10.1007/s10592-015-0775-4

Beng, K. C., & Corlett, R. T. (2020). Applications of environmental DNA (eDNA) in ecology and conservation: Opportunities, challenges and prospects. *Biodiversity and Conservation*, *29*(7), 2089–2121.

Berner, D. (2019). Allele frequency difference AFD–an intuitive alternative to FST for quantifying genetic population differentiation. *Genes*, *10*(4), 308.

Booton, G. C., Kaufman, L., Chandler, M., Oguto-Ohwayo, R., Duan, W., & Fuerst, P. A. (1999). Evolution of the ribosomal RNA internal transcribed spacer one (ITS-1) in cichlid fishes of the Lake

Victoria region. *Molecular Phylogenetics and Evolution*, *11*(2), 273–282.

Burian, A., Mauvisseau, Q., Bulling, M., Domisch, S., Qian, S., & Sweet, M. (2021). Improving the reliability of eDNA data interpretation. *Molecular Ecology Resources*, *21*(5), 1422–1433.

Bylemans, J., Furlan, E. M., Gleeson, D. M., Hardy, C. M., & Duncan, R. P. (2018). Does size matter? An experimental evaluation of the relative abundance and decay rates of aquatic environmental DNA. *Environmental Science & Technology*, *52*(11), 6408–6416.

Bylemans, J., Furlan, E. M., Hardy, C. M., McGuffie, P., Lintermans, M., & Gleeson, D. M. (2017). An environmental DNA-based method for monitoring spawning activity: A case study, using the endangered Macquarie perch (Macquaria australasica). *Methods in Ecology and Evolution*, *8*(5), 646–655.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583.

Carraro, L., Hartikainen, H., Jokela, J., Bertuzzo, E., & Rinaldo, A. (2018). Estimating species distribution and abundance in river networks using environmental DNA. *Proceedings of the National Academy of Sciences*, *115*(46), 11724–11729.

Carreon-Martinez, L. B., Wellband, K. W., Johnson, T. B., Ludsin, S. A., & Heath, D. D. (2014). Novel molecular approach demonstrates that turbid river plumes reduce predation mortality on larval fish. *Molecular Ecology*, *23*(21), 5366–5377.

Cole, L. W. (2016). The evolution of per-cell organelle number. *Frontiers in Cell and Developmental Biology*, *4*, 85.

D'Aloia, C. C., Andrés, J. A., Bogdanowicz, S. M., McCune, A. R., Harrison, R. G., & Buston, P. M. (2020). Unraveling hierarchical genetic structure in a marine metapopulation: A comparison of three high-throughput genotyping approaches. *Molecular Ecology*, *29*(12), 2189–2203.

De Barba, M., Miquel, C., Lobréaux, S., Quenette, P. Y., Swenson, J. E., & Taberlet, P. (2017). High-throughput microsatellite genotyping in ecology: Improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Molecular Ecology Resources*, *17*(3), 492–507.

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.

Deiner, K., Renshaw, M. A., Li, Y., Olds, B. P., Lodge, D. M., & Pfrender, M. E. (2017). Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA. *Methods in Ecology and Evolution*, *8*(12), 1888–1898.

Doi, H., Inui, R., Akamatsu, Y., Kanno, K., Yamanaka, H., Takahara, T., & Minamoto, T. (2017). Environmental DNA analysis for estimating the abundance and biomass of stream fish. *Freshwater Biology*, *62*(1), 30–39.

Dowle, E. J., Pochon, X., Banks, J. C., Shearer, K., & Wood, S. A. (2016). Targeted gene enrichment and high-throughput sequencing for environmental biomonitoring: A case study using freshwater macroinvertebrates. *Molecular Ecology Resources*, *16*(5), 1240–1254.

Dugal, L., Thomas, L., Reinholdt Jensen, M., Sigsgaard, E. E., Simpson, T., Jarman, S., Thomsen, P. F., & Meekan, M. (2021). Individual haplotyping of whale sharks from seawater environmental DNA. *Molecular Ecology Resources*, *22*(1), 56–65.

Dunn, N., Priestley, V., Herraiz, A., Arnold, R., & Savolainen, V. (2017). Behavior and season affect crayfish detection and density inference using environmental DNA. *Ecology and Evolution*, *7*(19), 7777–7785.

Dysthe, J. C., Franklin, T. W., McKelvey, K. S., Young, M. K., & Schwartz, M. K. (2018). An improved environmental DNA assay for bull trout (Salvelinus confluentus) based on the ribosomal internal transcribed spacer I. *PLoS One*, *13*(11), e0206851.

Edgar, R. C. (2016). UNOISE2: Improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv*, 081257.

Egeland, T., Dalen, I., & Mostad, P. F. (2003). Estimating the number of contributors to a DNA profile. *International Journal of Legal Medicine*, *117*(5), 271–275.

Elnifro, E. M., Ashshi, A. M., Cooper, R. J., & Klapper, P. E. (2000). Multiplex PCR: Optimization and application in diagnostic virology. *Clinical Microbiology Reviews*, *13*(4), 559–570.

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., Bovery, C., … Duffy, D. J. (2022). Detection and population genomics of sea turtle species via noninvasive environmental DNA analysis of nesting beach sand tracks and oceanic water. *Molecular Ecology Resources*, *22*(7), 2471–2493.

Fediajevaite, J., Priestley, V., Arnold, R., & Savolainen, V. (2021). Meta-analysis shows that environmental DNA outperforms traditional surveys, but warrants better reporting standards. *Ecology and Evolution*, *11*(9), 4803–4815.

Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology Letters*, *4*(4), 423–425. https://doi.org/10.1098/rsbl.2008.0118

Furlan, E. M., Davis, J., & Duncan, R. P. (2020). Identifying error and accurately interpreting environmental DNA metabarcoding results: A case study to detect vertebrates at arid zone waterholes. *Molecular Ecology Resources*, *20*(5), 1259–1276.

Furtwängler, A., Reiter, E., Neumann, G. U., Siebke, I., Steuri, N., Hafner, A., Lösch, S., Anthes, N., Schuenemann, V. J., & Krause, J. (2018). Ratio of mitochondrial to nuclear DNA affects contamination estimates in ancient DNA analysis. *Scientific Reports*, *8*(1), 14075.

Gantz, C. A., Renshaw, M. A., Erickson, D., Lodge, D. M., & Egan, S. P. (2018). Environmental DNA detection of aquatic invasive plants in lab mesocosm and natural field conditions. *Biological Invasions*, *20*(9), 2535–2552.

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.

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., Spear, S. F., McKee, A., Oyler-McCance, S. J., Cornman, R. S., Laramie, M. B., Mahon, A. R., Lance, R. F., Pilliod, D. S., Strickler, K. M., Waits, L. P., Fremier, A. K., Takahara, T., Herder, J. E., & Taberlet, P. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, *7*(11), 1299–1307.

Gorički, Š., Stanković, D., Snoj, A., Kuntner, M., Jeffery, W. R., Trontelj, P., Pavićević, M., Grizelj, Z., Năpăruş-Aljančič, M., & Aljančič, G. (2017). Environmental DNA in subterranean biology: Range extension and taxonomic implications for Proteus. *Scientific Reports*, *7*(1), 1–11.

Haned, H., Pene, L., Lobry, J. R., Dufour, A. B., & Pontier, D. (2011). Estimating the number of contributors to forensic DNA mixtures: Does maximum likelihood perform better than maximum allele count? *Journal of Forensic Sciences*, *56*(1), 23–28.

Hebert, P. D., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1512), 313–321.

Hivert, V., Leblois, R., Petit, E. J., Gautier, M., & Vitalis, R. (2018). Measuring genetic differentiation from Pool-seq data. *Genetics*, *210*(1), 315–330.

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. https://doi.org/10.1111/j.1755-263X.2010.00158.x

Jo, T., Arimoto, M., Murakami, H., Masuda, R., & Minamoto, T. (2019). Particle size distribution of environmental DNA from the nuclei of marine fish. *Environmental Science & Technology*, *53*(16), 9947–9956.

Jo, T., Arimoto, M., Murakami, H., Masuda, R., & Minamoto, T. (2020). Estimating shedding and decay rates of environmental nuclear DNA with relation to water temperature and biomass. *Environmental DNA*, *2*(2), 140–151.

Kidd, K. K., Pakstis, A. J., Speed, W. C., Lagacé, R., Chang, J., Wootton, S., Haigh, E., & Kidd, J. R. (2014). Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics. *Forensic Science International: Genetics*, *12*, 215–224.

Klymus, K. E., Richter, C. A., Chapman, D. C., & Paukert, C. (2015). Quantification of eDNA shedding rates from invasive bighead carp *Hypophthalmichthys nobilis* and silver carp *Hypophthalmichthys molitrix*. *Biological Conservation*, *183*, 77–84.

Kofler, R., Pandey, R. V., & Schlotterer, C. (2011). PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-seq). *Bioinformatics*, *27*(24), 3435–3436. https://doi.org/10.1093/bioinformatics/btr589

Lacoursière-Roussel, A., Côté, G., Leclerc, V., & Bernatchez, L. (2016). Quantifying relative fish abundance with eDNA: A promising tool for fisheries management. *Journal of Applied Ecology*, *53*(4), 1148–1157.

Lodge, D. M. (2022). Policy action needed to unlock eDNA potential. *Frontiers in Ecology and the Environment*, *20*(8), 448–449. https://doi.org/10.1002/fee.2563

Loeza-Quintana, T., Abbott, C. L., Heath, D. D., Bernatchez, L., & Hanner, R. H. (2020). Pathway to increase standards and competency of eDNA surveys (PISCeS)—Advancing collaboration and standardization efforts in the field of eDNA. *Environmental DNA*, *2*(3), 255–260.

Long, E. O., & Dawid, I. B. (1980). Repeated genes in eukaryotes. *Annual Review of Biochemistry*, *49*(1), 727–764.

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.

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.

Maruyama, A., Nakamura, K., Yamanaka, H., Kondoh, M., & Minamoto, T. (2014). The release rate of environmental DNA from juvenile and adult fish. *PLoS One*, *9*(12), e114639.

Miller, C. R., Joyce, P., & Waits, L. P. (2002). Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics*, *160*(1), 357–366.

Minamoto, T., Uchii, K., Takahara, T., Kitayoshi, T., Tsuji, S., Yamanaka, H., & Doi, H. (2017). Nuclear internal transcribed spacer-1 as a sensitive genetic marker for environmental DNA studies in common carp Cyprinus carpio. *Molecular Ecology Resources*, *17*(2), 324–333.

Moushomi, R., Wilgar, G., Carvalho, G., Creer, S., & Seymour, M. (2019). Environmental DNA size sorting and degradation experiment indicates the state of Daphnia magna mitochondrial and nuclear eDNA is subcellular. *Scientific Reports*, *9*(1), 1–9.

Nakanishi, H., Fujii, K., Nakahara, H., Mizuno, N., Sekiguchi, K., Yoneyama, K., Hara, M., Takada, A., & Saito, K. (2020). Estimation of the number of contributors to mixed samples of DNA by mitochondrial DNA analyses using massively parallel sequencing. *International Journal of Legal Medicine*, *134*, 101–109.

Olson, Z. H., Briggler, J. T., & Williams, R. N. (2012). An eDNA approach to detect eastern hellbenders (*Cryptobranchus a. alleganiensis*) using samples of water. *Wildlife Research*, *39*(7), 629. https://doi.org/10.1071/wr12114

Paoletti, D. R., Doom, T. E., Krane, C. M., Raymer, M. L., & Krane, D. E. (2005). Empirical analysis of the STR profiles resulting from conceptual mixtures. *Journal of Forensic Sciences*, *50*(6), 1–6.

Paoletti, D. R., Krane, D. E., Doom, T. E., & Raymer, M. (2011). Inferring the number of contributors to mixed DNA profiles. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *9*(1), 113–122.

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.

Piggott, M. P. (2016). Evaluating the effects of laboratory protocols on eDNA detection probability for an endangered freshwater fish. *Ecology and Evolution*, *6*(9), 2739–2750.

Pilliod, D. S., Goldberg, C. S., Arkle, R. S., & Waits, L. P. (2013). Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences*, *70*(8), 1123–1130.

Pinfield, R., Dillane, E., Runge, A. K. W., Evans, A., Mirimin, L., Niemann, J., Reed, T. E., Reid, D. G., Rogan, E., Samarra, F. I. P., Sigsgaard, E. E., & Foote, A. D. (2019). False-negative detections from environmental DNA collected in the presence of large numbers of killer whales (*Orcinus orca*). *Environmental DNA*, *1*(4), 316–328.

Pompanon, F., Bonin, A., Bellemain, E., & Taberlet, P. (2005). Genotyping errors: Causes, consequences and solutions. *Nature Reviews Genetics*, *6*(11), 847–859.

Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The barcode of life data system (http://www.Barcodinglife.Org). *Molecular Ecology Notes*, *7*(3), 355–364.

Robin, E. D., & Wong, R. (1988). Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *Journal of Cellular Physiology*, *136*(3), 507–513.

Ruppert, K. M., Kline, R. J., & Rahman, M. S. (2019). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation*, *17*, e00547.

Schmelzle, M. C., & Kinziger, A. P. (2016). Using occupancy modelling to compare environmental DNA to traditional field methods for regional-scale monitoring of an endangered aquatic species. *Molecular Ecology Resources*, *16*(4), 895–908.

Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated–reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, *15*(6), 1289–1303.

Sepulveda, A. J., Schabacker, J., Smith, S., Al-Chokhachy, R., Luikart, G., & Amish, S. J. (2019). Improved detection of rare, endangered and invasive trout in using a new large-volume sampling method for eDNA capture. *Environmental DNA*, *1*(3), 227–237.

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. https://doi.org/10.1111/2041-210x.13079

Shelton, A. O., Gold, Z. J., Jensen, A. J., D'Agnese, E., Andruszkiewicz Allan, E., Van Cise, A., Gallego, R., Ramón-Laca, A., Garber-Yonts, M., Parsons, K., & Kelly, R. P. (2023). Toward quantitative metabarcoding. *Ecology*, *104*(2), e3906.

Shelton, A. O., Ramón-Laca, A., Wells, A., Clemons, J., Chu, D., Feist, B. E., Kelly, R. P., Parker-Stetter, S. L., Thomas, R., Nichols, K. M., & Park, L. (2022). Environmental DNA provides quantitative estimates of Pacific hake abundance and distribution in the open ocean. *Proceedings of the Royal Society B*, *289*(1971), 20212613.

Shogren, A. J., Tank, J. L., Andruszkiewicz, E., Olds, B., Mahon, A. R., Jerde, C. L., & Bolster, D. (2017). Controls on eDNA movement in streams: Transport, retention, and resuspension. *Scientific Reports*, *7*(1), 5065. https://doi.org/10.1038/s41598-017-05223-1

Shum, P., & Palumbi, S. R. (2021). Testing small-scale ecological gradients and intraspecific differentiation for hundreds of kelp forest species using haplotypes from metabarcoding. *Molecular Ecology*, *30*(13), 3355–3373.

Sigsgaard, E. E., Jensen, M. R., Winkelmann, I. E., Moller, 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. https://doi.org/10.1111/eva.12882

Sigsgaard, E. E., Nielsen, I. B., Bach, S. S., Lorenzen, E. D., Robinson, D. P., Knudsen, S. W., Pedersen, M. W., Al Jaidah, M., Orlando, L., Willerslev, E., Møller, P. R., & Thomsen, P. F. (2016). Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution*, *1*(1), 4.

Slatkin, M. (2008). Linkage disequilibrium—Understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, *9*(6), 477–485.

Smith, T. F., & Waterman, M. S. (1981). Identification of common molecular subsequences. *Journal of Molecular Biology*, *147*(1), 195–197.

Stat, M., Huggett, M. J., Bernasconi, R., DiBattista, J. D., Berry, T. E., Newman, S. J., Harvey, E. S., & Bunce, M. (2017). Ecosystem biomonitoring with eDNA: Metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports*, *7*(1), 12240.

Stepien, C. A., Snyder, M. R., & Elz, A. E. (2019). Invasion genetics of the silver carp *Hypophthalmichthys molitrix* across North America: Differentiation of fronts, introgression, and eDNA metabarcode detection. *PLoS One*, *14*(3), e0203012.

Stoeckle, M. Y., Adolf, J., Charlop-Powers, Z., Dunton, K. J., Hinks, G., & VanMorter, S. M. (2021). Trawl and eDNA assessment of marine fish diversity, seasonality, and relative abundance in coastal New Jersey, USA. *ICES Journal of Marine Science*, *78*(1), 293–304.

Strickler, K. M., Fremier, A. K., & Goldberg, C. S. (2015). Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. *Biological Conservation*, *183*, 85–92.

Swaminathan, H., Grgicak, C. M., Medard, M., & Lun, D. S. (2015). NOCIt: A computational method to infer the number of contributors to DNA samples analyzed by STR genotyping. *Forensic Science International. Genetics*, *16*, 172–180. https://doi.org/10.1016/j.fsigen.2014.11.010

Székely, D., Corfixen, N. L., Mørch, L. L., Knudsen, S. W., McCarthy, M. L., Teilmann, J., Heide-Jørgensen, M. P., & Olsen, M. T. (2021). Environmental DNA captures the genetic diversity of bowhead whales (*Balaena mysticetus*) in West Greenland. *Environmental DNA*, *3*(1), 248–260.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L. P., & Bouvet, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, *24*(16), 3189–3194.

Toju, H., Tanabe, A. S., Yamamoto, S., & Sato, H. (2012). High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. *PLoS One*, *7*(7), e40863.

Tsuji, S., Maruyama, A., Miya, M., Ushio, M., Sato, H., Minamoto, T., & Yamanaka, H. (2020). Environmental DNA analysis shows high potential as a tool for estimating intraspecific genetic diversity in a wild fish population. *Molecular Ecology Resources*, *20*(5), 1248–1258.

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., Sawada, H., & Ushio, M. (2020). Quantitative evaluation of intraspecific genetic diversity in a natural fish population using environmental DNA analysis. *Molecular Ecology Resources*, *20*(5), 1323–1332.

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), e02036.

Uchii, K., Doi, H., & Minamoto, T. (2016). A novel environmental DNA approach to quantify the cryptic invasion of non-native genotypes. *Molecular Ecology Resources*, *16*(2), 415–422.

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.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet, N., Roset, N., Copp, G. H., Geniez, P., Pont, D., Argillier, C., Baudoin, J. M., … Dejean, T. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, *25*(4), 929–942.

Wei, N., Nakajima, F., & Tobino, T. (2018). A microcosm study of surface sediment environmental DNA: Decay observation, abundance estimation, and fragment length comparison. *Environmental Science & Technology*, *52*(21), 12428–12435.

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.

Wilcox, T. M., Zarn, K. E., Piggott, M. P., Young, M. K., McKelvey, K. S., & Schwartz, M. K. (2018). Capture enrichment of aquatic environmental DNA: A first proof of concept. *Molecular Ecology Resources*, *18*(6), 1392–1401.

Yao, M., Zhang, S., Lu, Q., Chen, X., Zhang, S. Y., Kong, Y., & Zhao, J. (2022). Fishing for fish environmental DNA: Ecological applications, methodological considerations, surveying designs, and ways forward. *Molecular Ecology*, *31*(20), 5132–5164.

Yates, M. C., Fraser, D. J., & Derry, A. M. (2019). Meta-analysis supports further refinement of eDNA for monitoring aquatic species-specific abundance in nature. *Environmental DNA*, *1*(1), 5–13.

Yates, M. C., Wilcox, T. M., Stoeckle, M. Y., & Heath, D. D. (2022). Interspecific allometric scaling in eDNA production among northwestern Atlantic bony fishes reflects physiological allometric scaling. *Environmental DNA*. https://doi.org/10.1002/edn3.381

Yoccoz, N. G., Nichols, J. D., & Boulinier, T. (2001). Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution*, *16*(8), 446–453.

Yoshitake, K., Fujiwara, A., Matsuura, A., Sekino, M., Yasuike, M., Nakamura, Y., Nakamichi, R., Kodama, M., Takahama, Y., Takasuka, A., Asakawa, S., Nishikiori, K., Kobayashi, T., & Watabe, S. (2021). Estimation of tuna population by the improved analytical pipeline of unique molecular identifier-assisted HaCeD-seq (haplotype count from eDNA). *Scientific Reports*, *11*(1), 1–12.

Yoshitake, K., Yoshinaga, T., Tanaka, C., Mizusawa, N., Reza, M., 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.

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