



*Annual Review of Ecology, Evolution, and Systematics*

# Uses and Misuses of Environmental DNA in Biodiversity Science and Conservation

Melania E. Cristescu<sup>1</sup> and Paul D.N. Hebert<sup>2</sup>

<sup>1</sup>Department of Biology, McGill University, Montreal, Quebec, Canada H3A 1B1; email: melania.cristescu@mcgill.ca

<sup>2</sup>Centre for Biodiversity Genomics and Department of Integrative Biology, University of Guelph, Ontario, Canada N1G 2W1

Annu. Rev. Ecol. Evol. Syst. 2018. 49:209–30

The *Annual Review of Ecology, Evolution, and Systematics* is online at [ecolsys.annualreviews.org](http://ecolsys.annualreviews.org)

<https://doi.org/10.1146/annurev-ecolsys-110617-062306>

Copyright © 2018 by Annual Reviews.  
All rights reserved

## Keywords

biodiversity, monitoring, environmental DNA, DNA barcoding, metabarcoding, metagenomics, environmental impact assessment

## Abstract

The study of environmental DNA (eDNA) has the potential to revolutionize biodiversity science and conservation action by enabling the census of species on a global scale in near-real time. To achieve this promise, technical challenges must be resolved. In this review, we explore the main uses of eDNA as well as the complexities introduced by its misuse. Current eDNA methods require refinement and improved calibration and validation along the entire workflow to lessen false positives/negatives. Moreover, there is great need for a better understanding of the “natural history” of eDNA—its origins, state, lifetime, and transportation—and for more detailed insights concerning the physical and ecological limitations of eDNA use. Although eDNA analysis can provide powerful information, particularly in freshwater and marine environments, its impact is likely to be less significant in terrestrial settings. The broad adoption of eDNA tools in conservation will largely depend on addressing current uncertainties in data interpretation.



**Environmental DNA (eDNA):** traces of DNA released by organisms into their environments and extracted without the isolation of target organism(s)

**Extracellular DNA (exDNA):** DNA released from prokaryotic and eukaryotic cells into the environment; may originate from physiologically active or dead cells

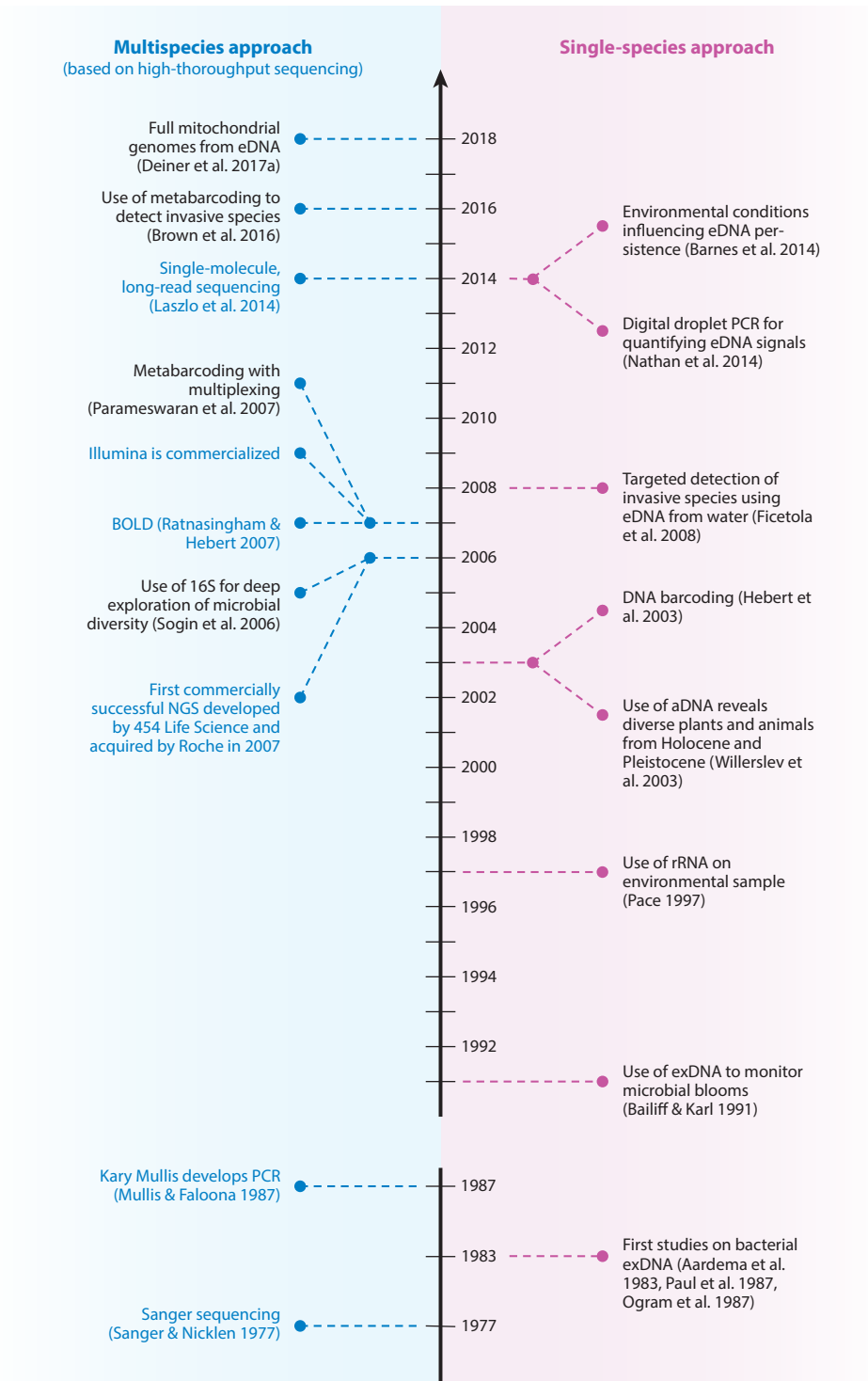
**Ancient DNA (aDNA):** DNA isolated from old specimens, recovered in the absence of obvious fossilized remains of organisms, and usually of low quality

## 1. INTRODUCTION

It has long been known that environments are rich reservoirs of DNA derived from resident organisms. The term extracellular DNA was first adopted to refer to DNA released by prokaryotic cells (and to some extent also eukaryotes) into the environment (Pietramellara et al. 2009, Ibáñez de Aldecoa et al. 2017). Early microbial studies revealed that extracellular DNA plays an important structural role in microbial films and that many bacteria release exonucleases that prevent its degradation (Ogram et al. 1987). Over the past decade, a new emphasis has developed, focused on using environmental DNA to identify metazoans (Ficetola et al. 2008) and to probe complex species assemblages across the trophic chain (Thomsen et al. 2012, Bohmann et al. 2014). Eukaryotic DNA extracted from environmental samples undoubtedly derives from diverse sources including decomposing organisms, shed epidermal cells, and body secretions (e.g., feces, urine, gametes). Although its physical state is poorly understood (Barnes & Turner 2016), it certainly occurs as both extracellular DNA in various states of degradation and as intact DNA in cellular form (Lodge et al. 2012, Taberlet et al. 2012). The transition from intracellular to extracellular DNA (either dissolved or particulate) is controlled by diverse physical and biological processes (Barnes & Turner 2016). Reflecting this complexity, the half-life of extracellular DNA varies dramatically, from a few days in warm water to many thousands of years in cold sediment with its concentration typically ranging from 1 to 100  $\mu\text{g L}^{-1}$ —higher in sediments than in water as well as in eutrophic versus oligotrophic habitats (Barnes & Turner 2016, Ibáñez de Aldecoa et al. 2017). The rise of this new wave of environmental DNA studies (**Figure 1**) has created some confusion in the literature because the acronym eDNA, long-used for work on extracellular DNA, has now also been adopted to refer to environmental DNA. As it seems important to resolve the confusion created by this dual usage, we propose that exDNA be employed to designate strictly extracellular DNA and eDNA be used to designate all DNA extracted from environmental samples (water, sediments, air) in the absence of their source organisms. Accordingly, we employ the latter acronym throughout the balance of this paper and do not consider studies using bulk samples of organisms without explicit indication.

Over the past decade, it has become increasingly clear that analysis of eDNA from natural environments can reveal information on the occurrence of targeted organisms (e.g., rare and endangered species, recently arrived invasive species) with high efficiency and sensitivity and without destructive sampling (Thomsen et al. 2012, Biggs et al. 2015). The adoption of eDNA-based techniques gained momentum when it was realized that this approach could, when coupled with high-throughput sequencing, reveal information on the biotic composition of entire ecosystems (Beja-Pereira et al. 2009, Blaaid et al. 2012, Clare 2014). Prior use of DNA to evaluate biological assemblages has involved the analysis of ancient DNA (aDNA) extracted from microfossils (Willerslev et al. 2003). Whereas aDNA from lake sediments, peats, permafrost soils, preserved gut content, and coprolites can aid the reconstruction of paleoenvironments (Willerslev et al. 2003, Birks & Birks 2016), eDNA aids the evaluation of biotic diversity in modern environments.

Methods for species identification based on eDNA have clear advantages over morphological approaches (Ficetola et al. 2008, Darling & Mahon 2011, Bohmann et al. 2014). Two obvious benefits relate to the capacity of these methods to reveal the presence of organisms that cannot otherwise be sampled (rare or ephemeral species) or recognized (cryptic species, larval stages, degraded specimens) (Caesar et al. 2006, Jerde et al. 2011). Moreover, eDNA-based methods permit noninvasive monitoring, reducing the unintentional transfer of invasive species linked to traditional sampling methods (Beja-Pereira et al. 2009, Valentini et al. 2016). Its high detection capability and sensitivity along with reduced costs in comparison to those of traditional methods create an unprecedented opportunity to advance biodiversity monitoring and environmental management (Sutherland et al. 2013, Cristescu 2014, Kelly et al. 2014b).



**Figure 1**

Timeline indicating major breakthroughs leading to metabarcoding approaches using eDNA. Abbreviations: aDNA, ancient DNA; BOLD, Barcode of Life Data System; eDNA, environmental DNA; exDNA, extracellular DNA; NGS, next-generation sequencing; rRNA, ribosomal RNA.

Annu. Rev. Ecol. Syst. 2018.49. Downloaded from www.annualreviews.org. Access provided by CASA Institution Identity on 09/12/18. For personal use only.



**Microarray:** used commonly to measure the level of expression in multiple genes simultaneously and used less frequently to identify species

**DNA barcoding:** identification of species using standardized DNA fragments and based on single-specimen sequencing

**Real-time quantitative PCR (qPCR):** uses the PCR reaction to detect targets of interest

Rapid adoption of eDNA-based techniques has been facilitated by the ongoing revolution in sequencing technologies (Shokralla et al. 2012). In fact, ingenious methods for decoding the rich information stored in eDNA have advanced so rapidly that some approaches have been displaced before being fine-tuned or fully validated. For example, microarray-based analyses are now used infrequently for species identification despite early enthusiasm for “biodiversity chips” designed to identify species from multiple phyla (Pfundler et al. 2004, Carr et al. 2008). However, these microarray-based methods suffered from a key disadvantage; they required detailed knowledge of sequence variation in the target taxa, making identification of rare geographic variants as well as divergent or newly encountered taxa impossible. As a consequence, despite their promise, DNA microarrays have now been supplanted for biodiversity analysis by direct sequencing.

The first eDNA studies employed Sanger sequencing coupled with specific primer sets to reveal the occurrence of target species (Ficetola et al. 2008, Jerde et al. 2011, Dejean et al. 2012). This approach is very similar to DNA barcoding, but it escapes the need to collect specimens of the target taxon once a species-specific primer set is available. However, a reference sequence library derived from the analysis of voucher specimens belonging to the taxonomic group under study is critical to support such work (Hajibabaei et al. 2007). Subsequent eDNA studies began to move beyond simple detection of a species to quantification of its abundance. The use of real-time quantitative PCR (qPCR) permits targeted detection of eDNA from a particular species, and quantification of variation in the copy number of eDNA molecules provides a sense of variation in its abundance among habitats (Kirshtein et al. 2007, Biggs et al. 2015). However, full realization of the impacts of eDNA-based methods for biodiversity assessment required high-throughput sequencing platforms. By allowing the simultaneous recovery of sequences from broad taxonomic assemblages, these platforms enabled a move from studies targeting single species to investigations probing entire communities (Shokralla et al. 2012, Cristescu 2014). These technical advances (Margulies et al. 2005, Glenn 2011, Goodwin et al. 2016) occurred when the research community was already strongly committed to building DNA barcode reference sequence libraries (Hebert et al. 2003). By 2015, the International Barcode of Life Project had delivered reference sequences for 500,000 species of animals and plants (Hebert et al. 2016), and coverage is expected to expand to 1.5 million species by 2025. This effort is important, as access to reference sequences is critical for eDNA studies, whether they target single species or entire biotic assemblages. The research community is now ready to adopt eDNA-based techniques as one key element in the employment of genetic methods to revolutionize how biodiversity assessments and conservation science are advanced (Lodge et al. 2012, Porter & Hajibabaei 2018).

Despite its very rapid development, eDNA-based methods require calibration, validation, and cautious interpretation (Bohmann et al. 2014, Cristescu 2014, Kelly et al. 2014a, Roussel et al. 2015). Although all elements of the eDNA workflow (sampling, DNA extraction, sequencing, data analysis) are constantly improving, every step in the analytical pathway requires calibration. There is also a strong need to explore the potential for misinterpretation of the results from eDNA studies, particularly when they underpin decisions relating to species conservation or habitat management. Although the need for vigilance is critical when eDNA results are employed as a basis for policy decisions (Darling & Mahon 2011, Creer et al. 2016, Goldberg et al. 2016), the same standard applies to studies that aim to advance our understanding of fundamental ecological and evolutionary processes.

Interpretation of many eDNA studies is currently impeded by our limited understanding of how various field and laboratory protocols influence the eDNA detection (Wilcox et al. 2016). Moreover, there is great uncertainty concerning the impacts of biotic and environmental conditions on the ecology of eDNA (its production, state, transport, and half-life within an environment) (Barnes & Turner 2016). Additionally, important limitations include the difficulty in obtaining

quantitative information (abundance data) from multispecies eDNA extracts when many applications require it (Kelly et al. 2014a, Biggs et al. 2015) as well as the lack of information on the developmental stage, sex, and size of individuals detected (Valentini et al. 2016). Another barrier to data interpretation results from gaps in taxon coverage in reference libraries, meaning that many sequences derived from eDNA analysis cannot be assigned to their source taxon (Cordier et al. 2017). Because of these limitations, the actual incidence of false positives and false negatives is likely high (Darling & Mahon 2011, Coissac et al. 2012, Roussel et al. 2015). Furthermore, the lack of standard protocols for eDNA sampling in the field and analysis in the lab (Kelly et al. 2014a, Rees et al. 2014, Flynn et al. 2015, Hunter et al. 2017) makes studies insular and comparative inferences difficult (Cristescu 2014).

Although recent reviews (Shokralla et al. 2012, Bohmann et al. 2014, Rees et al. 2014, Barnes & Turner 2016, Deiner et al. 2017b) have provided a good history of eDNA methods and their main applications, they have only briefly considered open questions and future challenges (Thomsen & Willerslev 2015). The present article addresses this gap by discussing the major uses and misuses of eDNA in biodiversity science and conservation with emphasis on primary sources of error. We outline a framework for understanding, identifying, and addressing them and consider both methodological and “process” errors (*sensu* Darling & Mahon 2011) that are not related to technical weaknesses but to “unknown” features of eDNA. Moreover, we identify and discuss efficient ways of benchmarking and validating the most commonly used eDNA-based methods.

## 2. CURRENT STATE AND APPLICATIONS

Improved access to sequencing facilities, enhanced computational support, and the growth in reference sequence libraries all have stimulated the surge in eDNA studies. Collectively, these advances are facilitating the uptake of this approach by research teams interested in using this method to advance understanding of species distributions and community ecology. This surge has also been motivated by the advantages of nondestructive sampling coupled with sensitivity, speed, and cost-effectiveness, and it has been further fueled by a general enthusiasm for transforming biodiversity and conservation science (Bohmann et al. 2014, Creer et al. 2016). The possibility of monitoring species composition and distributions in near-real time has arrived at a time when anthropogenic change is driving biodiversity loss and altering ecosystem functions in ways and at rates that are impossible to document with classical approaches (Cardinale et al. 2012, Martin et al. 2016). With eDNA analysis, it is possible to probe species composition, to reveal biodiversity trends, to detect endangered or invasive species, and ultimately to improve understanding of ecosystem-level processes.

### 2.1. Applications in Biodiversity Science and Conservation Biology

Because biodiversity management activities often involve protecting species assemblages and their supporting habitats, they routinely require monitoring biodiversity trends, identifying species at risk, assessing biosecurity risks, and preventing the introduction of invasive species (Simberloff 2005, Lodge et al. 2012, Martin et al. 2016). Although conservation studies have historically focused on the species level, emphasis is now shifting to communities and ecosystem-level processes (Swank & Vanlear 1992, Franklin 1993, Naeem 2002). This shift is happening as methods of molecular identification are transitioning from species-level approaches to ecosystem-level investigations, enabling efficient application of molecular methods to conservation studies and opening the way for interdisciplinary and integrative approaches.





**Metabarcoding:**

DNA-based method for identifying multiple species extracted from a mass collection of specimens or from eDNA analyzed via high-throughput sequencing

**2.1.1. Freshwater and marine samples.** Most eDNA studies have targeted aquatic environments, particularly freshwater habitats. This emphasis reflects the fact that eDNA occurs in solution in freshwater and is bound to particulates in terrestrial and aerial settings. As a consequence, eDNA derived from species occurring in aquatic habitats has a relatively homogeneous distribution, thus reducing the number of sampling points needed to characterize a habitat while also simplifying the extraction of eDNA.

Although few studies have examined its compositional variation through the water column, eDNA extracted from aquatic samples (often surface water) is generally considered to represent the diversity of species residing in and near the sampled habitat. Early studies focused on detecting target species of amphibians (Ficetola et al. 2008, Goldberg et al. 2011, Dejean et al. 2012, Thomsen et al. 2012), fishes (Jerde et al. 2011), mammals (Thomsen et al. 2012), and invertebrates (Gardham et al. 2014, Treguier et al. 2014). Employing PCR or qPCR to recover sequences from eDNA samples, this work revealed vertebrate species whose presence would otherwise have been overlooked in habitats ranging from wetlands to rivers and large stratified lakes. By contrast, there has been lower success in detecting invertebrates, such as crustaceans, especially when populations are at low density (Treguier et al. 2014).

More recent metabarcoding studies have tested the utility of eDNA analysis for rapid surveys of complex aquatic communities. This work has revealed a new set of challenges and opportunities when results are compared with those from traditional surveys. For example, Deiner et al. (2016) found that eDNA recovered from rivers often derives from species in the catchment basin, including terrestrial organisms, reflecting the transport of eDNA from land to water. Although assessing biodiversity for a large geographic area is useful in some contexts (such as the identification of biodiversity hot spots or the detection of aquatic networks or regions experiencing sharp biodiversity decline), such transfers of eDNA complicate applications that aim to determine species occurrences or turnover on a local scale (Deiner & Altermatt 2014, Sansom & Sasse 2017). The few eDNA studies on invertebrate communities are counterbalanced by the much greater number of studies that have examined bulk samples of benthic or planktonic invertebrates. For example, compared with morphological surveys, an extensive survey of bulk zooplankton samples from marine and freshwater Canadian ports revealed much higher levels of species diversity as well as strong temporal species turnover, particularly in the Arctic (Chain et al. 2016). Moreover, using a database of 124 aquatic invasive species, Brown et al. (2016) detected 19% of these taxa, half of which were detected at previously unreported locations. This extensive survey of commercial ports from four geographic regions (Atlantic, Pacific, Arctic, and Great Lakes) demonstrated the efficacy of metabarcoding for detecting invasive species in both marine and freshwater environments. If such approaches can be extended, eDNA could replace more costly approaches focused on the detection of single species.

The use of eDNA to detect species is considerably more challenging in marine environments than in freshwater owing to greater dilution, increased mixing, and higher salinity. For example, one of the first studies to employ a well-calibrated eDNA assay to detect marine mammals found that species detection was less consistent with this assay than with acoustic monitoring (Foote et al. 2012). Although other studies have efficiently detected fish species in marine settings (Thomsen et al. 2012), the challenges posed by a more physically dynamic environment suggest protocols need to be adjusted to marine conditions (Roussel et al. 2015). Relatively few eDNA surveys have been completed on freshwater and marine plants (but see Scriver et al. 2015, Fujiwara et al. 2016 for species-specific approaches in freshwater plants).

Superficial lake sediments, as well as lake cores, contain eDNA and aDNA of fishes (Turner et al. 2015), mammals (Gigu et-Covex et al. 2014), invertebrates (Gardham et al. 2014), and both terrestrial and aquatic plants (Anderson-Carpenter et al. 2011, Parducci et al. 2013). However,

interpretation is more complex for eDNA results from sediment analysis than for those from the water column, and underlying assumptions have not been adequately investigated. Because eDNA is more resilient when bound to sediments than when in aqueous solution, species occurrences are preserved for a longer duration (Turner et al. 2015), potentially revealing both current and past biotic assemblages. Moreover, eDNA can penetrate sediments, contaminating lower layers with more recent DNA molecules. However, lake sediments are saturated, so they should experience little transport of liquids through layers. As a consequence, eDNA should be less mobile in lake sediments than in terrestrial sediments and soils (Anderson-Carpenter et al. 2011), but sediment disturbances caused by the burrowing activities and tube construction of benthic organisms (Mermillod-Blonding 2011) may often cause deeper displacement of eDNA. The impact of such movement adds complexity to data interpretation, which is further enhanced by considerable variation in the depth of bioturbation among habitats (Teal et al. 2008).

**2.1.2. Terrestrial sediments and soil.** Recent studies suggest that eDNA extracted from sediments and soil can provide a relatively accurate record of the diversity of fungi (Geml et al. 2009), plants (Yoccoz et al. 2012), mammals (Andersen et al. 2011), and some groups of soil invertebrates (Bienert et al. 2012). However, these studies also revealed important caveats. For example, eDNA surveys of plants in various biomes conducted by Yoccoz et al. (2012) were generally consistent with aboveground surveys, but eDNA was also recovered from plants that were cultivated 40–50 years earlier. Thus, the uncertainty surrounding the geographic source and age of the eDNA complicates the interpretation of results from sediments and soil. Although most eDNA studies assume that the age of the DNA recovered from a sediment corresponds to the time of its deposition, DNA can penetrate lower layers (Haile et al. 2007) and can be transported across large geographic regions via wind-blown particles. Horizontal mass transport of wind- and water-driven materials used as a metric for local soil redistribution (Breshears et al. 2003) may also be useful in predicting the level of eDNA transport in various terrestrial ecosystems.

### 3. TRANSITION FROM TOOL DEVELOPMENT TO BROAD APPLICATIONS

#### 3.1. Applications in Management and Policy

Ongoing initiatives are promoting the adoption of eDNA analysis by bringing this approach to the attention of managers, stakeholders, and other end users. These efforts are often focused on explaining the uncertainties in data interpretation while promoting standardization and comparability (Darling & Mahon 2011). However, the use of eDNA still lags far behind current technical advancements. This is surprising because current biomonitoring programs are expensive, often have negative environmental consequences, and are difficult to implement owing to the lack of taxonomic expertise (Kelly et al. 2014b). Despite these factors, eDNA-based methods are only rarely used for environmental management. One encouraging example is the acceptance by the United Kingdom of eDNA qPCR results as evidence for the presence of the protected great crested newt, *Triturus cristatus* (Biggs et al. 2015). Another more controversial example involves the detection of eDNA in samples collected beyond the invasion front of Asian carp and past the electric barrier intended to prevent its dispersal into Lake Michigan, USA (Jerde et al. 2011, 2013). This result generated a heated debate regarding whether this species had actually invaded the Great Lakes (Darling & Mahon 2011). The uncertainty led to a proposal for separation of the Great Lakes and Mississippi River basin by closing the Chicago Sanitary and Shipping Canal. Although waterway operators and current users of the canal strongly opposed the plan, this highly



politicized debate brought eDNA into the public eye and ultimately resulted in its adoption as a key element of the Asian carp monitoring program (Darling & Mahon 2011). In a similar way, New Zealand and Australia were early advocates for the use of DNA barcoding to support bio-surveillance (Armstrong & Ball 2005, Darling & Blum 2007) and are in a strong position to adopt eDNA-based monitoring (Wood et al. 2013). However, effective adoption of eDNA tools will depend on addressing current uncertainties in data interpretation.

### 3.2. Sources of Uncertainty

The single most important problem confronting the adoption of eDNA analysis relates to the current uncertainty regarding rates of false positives (target species is absent, but its DNA is recovered) and false negatives (species undetected where they are present).

**3.2.1. Unknown rates of false positives and false negatives.** False negatives are a major concern for studies aiming to identify invasive, rare, or endangered species (Goldberg et al. 2016). The probability of detection clearly depends on many factors. Although the concentration of eDNA derived from each species in the environment is key to its detection, other factors intervene. The isolation of eDNA depends on sampling effort, extraction efficiency, barriers to analysis (e.g., PCR inhibitors), and the sensitivity of analytical methods (Goldberg et al. 2016). As a result, overlooked species can reflect the failure to recover eDNA from the target species owing to its low population density, inadequate sampling, low efficiency in DNA extraction, or inhibitors that prevent amplification. In other cases, the sequence is recovered, but bioinformatics protocols are poorly adjusted so rare sequences are discarded as artifacts. False negatives can be significantly reduced by increasing analytical effort (e.g., increasing the number of samples per locality, analyzing multiple DNA extracts per sample, and increasing the number of PCR cycles) and by including positive controls (**Table 1**).

False positives are also a cause for concern because they can reflect contamination of the sample at some stage in the protocol, poor primer design, probes that lack adequate specificity, or ambiguities introduced during sequencing processing or sequence assignments. The incidence of false positives due to methods errors (*sensu* Darling & Mahon 2011) can be evaluated if negative controls are used at all steps in the analytical chain. Such analysis employs well-validated reference sequences to clarify the source of contamination, thereby permitting adjustments in protocols and filtering steps to eliminate sequence artifacts. However, process errors can also create false positives that are more difficult to address because the source of the problem is often elusive. Such errors arise from general assumptions during the interpretation of results. For example, when eDNA is detected, most studies conclude that the source organisms are present in the sampled habitat, but their detection could actually reflect eDNA transported from elsewhere or from the mobilization of eDNA released centuries earlier.

Other steps to optimize protocols such as rigorous primer design and consistent use of negative controls as well as positive controls at environmentally relevant concentrations can significantly reduce the risk of false negatives and false positives (Darling & Mahon 2011, Piñol et al. 2015). They are also effective when the target species is known, but less applicable for studies that aim to recover entire species assemblages. Many studies compare eDNA results against direct (traditional) assessments of community composition to estimate the incidence of errors. Although such comparisons are valuable as a proof of concept (Roussel et al. 2015), efforts to employ them in broad biodiversity surveys in real time require affordable built-in approaches to validate results. One of the least expensive and most effective ways involves parallel analysis of mock communities to estimate rates of false positives/negatives (Cristescu 2014).





**Table 1** Problems associated with species detection based on eDNA, the primary methods affected, and potential solutions

Problems and limitations	Description of problems	Methods affected	Type of error	Possible solutions	Reference(s)
Wrong method	Because field, laboratory, and bioinformatics protocols all influence the detection probability of eDNA and subsequent biodiversity estimates, selecting optimal protocols is key. Lack of familiarity with the full range of analytical options and their impact on results can lead to the adoption of a poorly suited method	Microarrays PCR qPCR Metabarcoding	M	Perform more comparative studies of field and laboratory protocols; Adopt well-validated and well-justified methods. For example, the efficacy of all protocols selected for use should be validated through analysis of mock communities (artificial communities of known taxonomic composition) with enough species diversity and genetic variation to mirror diversity likely in the study system	Clare et al. 2016, Creer et al. 2016, Piggott 2016, Schmelzle & Kinziger 2016, Wilcox et al. 2016, Scott et al. 2018
False positives (species detected but not present)	False positives can reflect eDNA contamination due to improper handling of samples (M), movement from another environment, or resuspension from deep to superficial sediment due to bioturbation (P). False positives can also result from primers and probes that lack adequate specificity, errors in data analysis (resulting in oversplitting of operational taxonomic units and inflated biodiversity estimates) (M), or mutations that accumulate postmortem (P)	Microarrays PCR qPCR Metabarcoding	M & P	Optimize protocols by including positive and negative controls; Systematically use mock communities; Use of multiple markers or multiple primer pairs per marker; Select appropriate parameters for bioinformatics processing of sequences, which can significantly reduce the impact and frequency of errors; Compare results against direct (traditional) assessments of community composition	Darling & Mahon 2011, Coissac et al. 2012, Brown et al. 2015, Flynn et al. 2015, Leray & Knowlton 2015, Roussel et al. 2015
False negatives (species not detected but present)	False negatives can reflect a limited amount of eDNA in the sample; rapid degradation, dilution, or transportation; insufficient sampling; sequence variation at priming sites; or unequal amplification of the eDNA template (primer bias)	Microarrays PCR qPCR Metabarcoding	M & P	Generate species accumulation curves against sampling effort to confirm that enough samples have been analyzed to reach an asymptote; Determine the limit of detection; Maximize the extraction of signal by generating multiple eDNA extracts from each sample and by conducting multiple PCRs on each extract; Optimize protocols by including positive and negative controls and rigorously test the performance of these primers using positive controls; Analyze mock communities to estimate the rate of false negatives; Use multiple markers or multiple primer pairs per marker and employ degenerate primers as well as blocking primers (if needed) when the study targets a diverse taxonomic assemblage; Compare results against traditional assessments of community composition	Darling & Mahon 2011, Coissac et al. 2012, Cristescu 2014, Deagle et al. 2014, Ficetola et al. 2014, Pfiol et al. 2015, Roussel et al. 2015, Kraaijeveld et al. 2015, Creer et al. 2016

(Continued)



**Table 1** (Continued)

Problems and limitations	Description of problems	Methods affected	Type of error	Possible solutions	Reference(s)
Uncertain spatial resolution	Poor understanding of the impacts of biotic and environmental conditions on the ecology of eDNA (its production, transport and half-life within an environment) makes it difficult to rule out its origin from another environment Does the captured signal come from the local environment or represent regional diversity?	Microarrays PCR qPCR Metabarcoding	P	In river systems, use information on eDNA shedding, decay, sedimentation or resuspension rates as well as on basic physical characteristics of sites to predict downstream transport	Yoccoz et al. 2012, Bohmann et al. 2014, Deiner & Altermatt 2014, Turner et al. 2015, Barnes & Turner 2016, Sansom & Sasse 2017
Uncertain temporal resolution	Poor understanding of the ecology of eDNA makes it difficult to separate sequences derived from eDNA versus aDNA compromising temporal inferences. Does the captured signal come from existing or past biotic communities?	Microarrays PCR qPCR Metabarcoding	P	Directly compare eDNA-based methods to traditional surveys, because eDNA may offer a longer time perspective than does traditional samplings. Recover longer barcodes that may help to target more recent, less degraded eDNA; Analyze nucleotide misincorporation and substitution patterns to discriminate eDNA and aDNA; Use alternative RNA-based markers (based on complementary DNA derived from mitochondrial RNA and/or ribosomal RNA) to provide narrower spatiotemporal inferences, especially desirable in very sensitive assays (e.g., detection of invasive species, identifying very recent or rapid species turnover in response to environmental stressors)	Haile et al. 2007, Bohmann et al. 2014, Turner et al. 2015, Barnes & Turner 2016, Pochon et al. 2017
No quantitative data	Very difficult to obtain quantitative information (species abundance) from eDNA	Microarray qPCR PCR Metabarcoding	P	Use real-time qPCR and digital PCR to generate quantitative data. PCR-free approaches will help to quantify the variation in eDNA concentration among species. However, because of variation in body size with age and the exponential function in mitochondrial copy number with body size, it will not be possible to estimate the number of specimens in a habitat without a shift to recovering markers used in forensic analyses	Kelly et al. 2014a, Biggs et al. 2015, Clarke et al. 2017

(Continued)



**Table 1** (Continued)

Problems and limitations	Description of problems	Methods affected	Type of error	Possible solutions	Reference(s)
Limited ecological information	Lack of information on the developmental stage, sex, size of individuals detected using eDNA	Microarrays qPCR PCR Metabarcoding	P	Use life-stage- and sex-specific markers as a potential solution of a limitation difficult to overcome	Valentini et al. 2016
Limited taxonomic resolution	Incomplete reference libraries and sequences in these libraries that derive from misidentified specimens mean that a large fraction of easily detectable diversity remains unclassified and poorly linked to their ecology	Barcoding Metabarcoding	M & P	Use local reference libraries when few target species are investigated; Extend the global reference sequence library on the basis of well-identified voucher specimens	Kraaijeveld et al. 2015, Leray & Knowlton 2015, Cordier et al. 2017
Limited comparability	Lack of standardized protocols complicates comparing results among studies	Metabarcoding	M & P	Use benchmarking eDNA-based metabarcoding approaches; Follow precautions similar to those established for aDNA protocols; Directly compare new methods against existing ones in independent labs and under multiple conditions	Willerslev et al. 2004, Beja-Pereira et al. 2009, Kelly et al. 2014a, Rees et al. 2014, Brown et al. 2015, Flynn et al. 2015, Hunter et al. 2017

We differentiate between “methods” errors (M) linked to the technique employed and “process” errors (P) that apply to all techniques because they arise from poorly understood features of eDNA (sensu Darling & Mahon 2011). Abbreviations: aDNA, ancient DNA; eDNA, environmental DNA; qPCR, quantitative PCR.

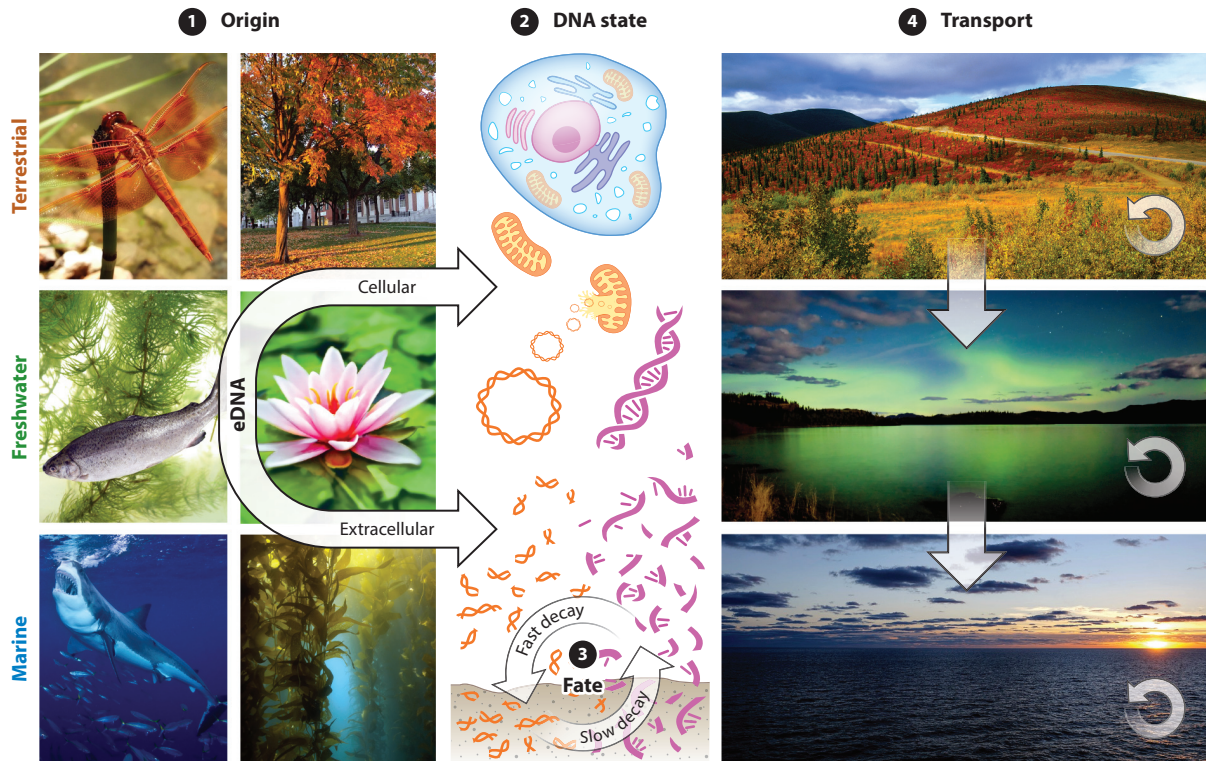


**3.2.2. Contamination in time and space.** Because eDNA in surface waters typically degrades within a few hours to days (Dejean et al. 2011, Thomsen et al. 2012), species detected by eDNA ordinarily indicate their recent presence (<1 month) and can reflect contemporaneous communities. Although this relatively short persistence is ideal for studies of current biodiversity (Lindahl 1993), eDNA in sediments can persist far longer and is often present at much higher concentration than is eDNA in the water column (Turner et al. 2015). Shaw et al. (2016) contrasted species richness values based on the analysis of eDNA extracted from water versus sediments and found that eDNA from water better reflects the present species composition. In fact, eDNA extracts from river sediments generated sequences of resident freshwater species, marine and estuarine species unlikely to occur at the sampled site, and freshwater species unrecorded for more than a century. Moreover, the ubiquity of aDNA in deep sediments, even when macrofossils are absent (Willerslev et al. 2003), indicates that sediments are better substrates for DNA preservation than is the water column; yet, few investigations have explored eDNA degradation rates in surficial sediments (Deiner et al. 2017b).

Because aDNA from sediments may be resuspended, particularly in rapidly flowing rivers, DNA extracted from water may often contain eDNA that reflects historical deposits. Separating recent eDNA from aDNA is not straightforward. Moreover, discriminating between eDNA (particularly its cellular form) and genomic DNA from small organisms inadvertently captured during sampling is difficult (Creer et al. 2016). However, DNA analyses of fossils have shown that their sequences can be distinguished from those with a recent source on the basis of their higher levels of nucleotide misincorporation (Höss et al. 1996). In particular, aDNA molecules show elevated levels of type 2 (cytosine-thymine/guanine-adenine) transitions (Sawyer et al. 2012), indicating a possible way of ascertaining if the sequence records for a particular species in an environmental sample derive from aDNA rather than eDNA.

Because eDNA moves readily through the environment (Barnes & Turner 2016), its geographic source is often difficult to establish (**Figure 2**). For example, downstream transport of eDNA in river systems is inevitably high (Roussel et al. 2015), and because the average velocity of the Amazon River is 2 km per hour, eDNA could be transported 1,500 km in 1 month, traveling one-quarter of the river's length. Deiner & Altermatt (2014) detected eDNA of two invertebrate species, *Daphnia longispina* and *Unio tumidus*, approximately 20 km downstream from the alpine lake where they were present. In addition, eDNA from terrestrial organisms living in the catchment area can be washed into aquatic systems revealing species that are clearly not residents. For example, analysis of aDNA from lake sediments revealed both terrestrial plant communities (Pansu et al. 2015) and livestock (Giguët-Covex et al. 2014), indicating that waterbodies are sinks for DNA originating in their watershed. Clearly, the biological, physical, and chemical factors responsible for the origin, movement, degradation, or persistence of eDNA need much more investigation (Barnes et al. 2014, Barnes & Turner 2016, Wilcox et al. 2016). Several recent eDNA studies have reported higher biotic diversity than has been revealed through traditional survey methods and have concluded that eDNA is a more sensitive approach (Dejean et al. 2012, Valentini et al. 2016). However, this conclusion fails to consider or control for the potential differences in timescale and geography reflected by the methods. To determine if eDNA is endogenous or exogenous, local or regional, present or past, we need a deeper understanding of its origin, state, transport, and fate (Turner et al. 2015). Moreover, such fine-scale temporal and geographic calibrations require alternative approaches like the use of RNA-based markers (based on complementary DNA derived from messenger or ribosomal RNA), whose persistence in the environment is much shorter (Pawlowski et al. 2016, Pochon et al. 2017).





**Figure 2**

Ecology of environmental DNA (eDNA): its (1) origin (derived from resident organisms), (2) state (cellular and extracellular form; free or particulate), (3) fate (slow decay in sediments or fast decay in water column), and (4) transport within and across environments (by wind, water currents, biotic activity, etc.).

**3.2.3. How quantitative is environmental DNA data?** Estimating the relative abundance of species remains a great challenge for eDNA-based methods (Valentini et al. 2016, Deiner et al. 2017b). Although eDNA concentrations are often positively correlated with population densities and biomass (Pilliod et al. 2013, Biggs et al. 2015), estimation of absolute abundance is problematic. Studies in well-controlled aquatic environments suggest that eDNA can be used to quantify population abundance if one employs species-specific primer sets and qPCR (Lacoursiere-Roussel et al. 2016). However, recovering abundance from whole communities using sequence counts is less promising owing to amplification bias among species during PCR and to species- and size-specific variation in the release of eDNA (Clarke et al. 2017). Linking eDNA reads to measures of species abundance requires a shift from standard PCR (Kelly et al. 2014a, Valentini et al. 2016). Digital PCR has the potential to improve quantification (Hindson et al. 2011), but it is no panacea because the count of eDNA molecules in any environment will depend on the complex interplay of biological (e.g., body sizes, release rates of eDNA) and physical (e.g., temperature, UV exposure) factors.

**3.2.4. From environmental DNA sampling to robust biodiversity estimates: selecting the best workflow.** Several studies have discussed important considerations and trade-offs that need to be considered before adopting any one of the most commonly used field, laboratory, and

**Digital PCR:**

alternative method to real-time qPCR that offers a more robust protocol for gaining an absolute count of sequences



bioinformatics protocols (Deiner et al. 2015, Piggott 2016). Protocols for each of these steps vary broadly in their suitability for application in specific environments (water, sediment, air), for particular taxonomic groups, and for the goal of the study (see Creer et al. 2016). However, insufficient knowledge on how particular workflows influence the detection probability of eDNA and resultant biodiversity estimates can lead to the adoption of suboptimal methods. Thus, we need more studies that compare the performance of new methods against existing protocols (Flynn et al. 2015, Clare et al. 2016). The most promising methods that emerge from such comparisons must then be stringently validated in several analytical facilities under multiple conditions (Beja-Pereira et al. 2009). For example, the efficacy of laboratory protocols could be validated through the analysis of mock communities with adequate taxonomic breadth (Brown et al. 2015).

**3.2.5. Major bioinformatics challenges.** Optimization of bioinformatics starts before data collection, as it begins with primer design. *In silico* analyses can help to identify markers that provide adequate resolution (Coissac et al. 2012) and can aid primer design (Clarke et al. 2017, Freeland 2017). An ideal metabarcoding gene region couples ease of recovery with high resolution of species in the target assemblage. Thus, marker genes must either have conserved primer-binding sites or incorporate degenerate bases to enable their recovery from diverse taxa. There should also be clear sequence divergence among closely related species, and a well-developed reference database should be available to maximize taxonomic assignments. No single marker will meet all these criteria because of the inevitable trade-off between the rapid evolution needed to optimize taxonomic resolution and the presence of conserved binding sites needed to optimize primer binding. Thus, taxonomic breadth and consistency can be increased by sequencing multiple independent markers (i.e., COI, 18S, 16S). However, because sequencing errors are far more difficult to detect with ribosomal genes than with the standard protein-coding barcode region COI, results based on ribosomal genes must be validated carefully (Flynn et al. 2015, Leray & Knowlton 2015). Supplemental barcode genes have the advantage of enabling sequences from eDNA analysis to be connected to other reference libraries such as ITS for fungi and 18S for protists.

An additional complication in data analysis arises from the fact that eDNA is often degraded (but see Turner et al. 2015, Deiner et al. 2017a), so primers often seek to amplify fragments of approximately 150 bp. By comparison, amplicons used in conventional DNA barcoding or metabarcoding are typically >500 bp, aiding discrimination of closely related species and often revealing intraspecific genetic diversity that is overlooked by short amplicons. Interestingly, recent studies suggest that longer eDNA molecules, likely derived from cellular DNA in environmental samples, may be more common than expected. The prospect of employing full-length barcodes for eDNA studies is now feasible because third-generation sequencers can generate high-fidelity sequences from single eDNA molecules. In fact, eDNA can be used to sequence multiple markers or even full genomes. For example, Deiner et al. (2017a) amplified large fragments (>16 Kb) from eDNA, recovering full mitogenomes. Such large data sets could be used to estimate population size and to advance population genetics and genomics (Barnes & Turner 2016, Sigsgaard et al. 2016).

By far the most important bioinformatics challenge lies in recognizing artefactual sequences that can be mistaken for rare species, greatly inflating biodiversity estimates (Kunin et al. 2010, Flynn et al. 2015). Sequence errors can be introduced at various steps in the analytical chain, by postmortem degradation in the environment, by replication errors during PCR, and during sequencing (Coissac et al. 2012). Sequencing platforms vary greatly with respect to their throughput, cost, error profile, and read structure (Goodwin et al. 2016). The transition from one platform to another requires bioinformatics adjustments since platform-specific bias during library construction and sequencing can impact downstream data interpretation. The relatively high error rates of



next-generation sequencing platforms are of general concern. Although not directly comparable, average error rates per base pair among different platforms vary by three orders of magnitude (0.01–16%) (Glenn 2011, Goodwin et al. 2016). Given this variation, the choice of analytical platform can substantially impact biodiversity estimates based on next-generation sequencing data. Few studies have directly tested the impact of sequencing platforms on surveys of microbial diversity using mock communities with known species composition or environmental samples with a well-known profile (Mahé et al. 2015, Tremblay et al. 2015). These microbial studies report that biodiversity estimates (particularly beta diversity) are not substantially altered by platform, but no equivalent studies have been conducted on multicellular organisms.

Because sequencing errors can be introduced at multiple stages, sequence processing remains an important issue. A suite of programs facilitates the identification and removal of both PCR and sequencing errors (reviewed in Coissac et al. 2012). Thus, sequence processing is often performed to reduce technical errors, but the bioinformatics steps can be an important source of error because appropriate parameters are often not known (Clare et al. 2016, Scott et al. 2018). Flynn et al. (2015) evaluated the performance of commonly used clustering methods on a series of mock communities and on more complex natural communities of zooplankton. Operational taxonomic unit (OTU) estimates varied by two and three orders of magnitude for mock communities and for natural communities, respectively, depending on the combination of methods used. Although this wide range of variation clearly indicates bioinformatics settings have a strong effect on OTU estimates, this effect can be slightly reduced in ecological analysis when the same parameters are used in all comparative analyses (Clare et al. 2016). Recently, new methods have been developed to overcome some of the problems related to the construction of OTUs. These methods use model-based approaches to remove spurious sequences and assign all unique sequences identified as biological variants. The retained amplicon sequence variants represent a higher resolution analog of the OTUs (Callahan et al. 2016).

These considerations indicate that the sequence-processing pipeline should be carefully optimized given the specific goals of the research program, such as estimating biodiversity or detecting a species of interest, because sequence abundance is often very low in the latter case. The performance of bioinformatics pipelines can be tested through simulation. For example, Scott et al. (2018) tested 1,050 combinations of parameters to determine the optimal parameter sets for particular research goals. They then tested the pipeline performance (detectability and sensitivity) by computationally inoculating sequences from 20 aquatic invasive species into 10 zooplankton community samples, revealing that optimal parameter selection often depends on the research goal.

**3.2.6. Difficulties with taxonomic assignments.** Taxonomic assignments are generally conducted using one of two distinct approaches. The more direct approach involves linking an unknown sequence with a Linnaean species by comparing the query sequence with reference sequences from a large assemblage of species in a database. The alternative approach involves delimiting species in the absence of reference data by defining groups of sequences that are likely to correspond to a particular taxonomic level. Because assignment methods can employ sequence similarity, phylogenetic inference, clustering algorithms, or coalescent-based approaches, selection of the appropriate method requires care (Porter & Hajibabaei 2018). The most common methods employ sequence similarity as a criterion, similar to the BLAST top-hit approach. However, false-positive rates are high when BLAST is employed, and the top hit is not always the closest phylogenetic neighbor (Koski & Golding 2001, Virgilio et al. 2010). The impact of this problem diminishes when the reference database includes sequences that are similar to the newly gathered sequence. Approaches involving species delimitation make use of arbitrary criteria of

---

**Operational taxonomic unit (OTU):** sequence clusters identified using a clustering algorithm based on sequence similarity and used as proxy for a taxonomic level

---



clustering and should be combined with other assignment criteria (Yeates et al. 2010, Coissac et al. 2012) that provide a statistical measure of confidence (Porter & Hajibabaei 2018).

Incomplete reference libraries and the presence of sequences derived from misidentified specimens mean that the species origin of many eDNA records remains uncertain. As a result, a large fraction of easily detectable diversity remains unclassified and, hence, unlinked to ecology. The use of customized, exhaustive reference databases for a target group of organisms can greatly improve taxonomic assignments (Harris 2003) and substantially increase taxonomic resolution (Valentini et al. 2016) while also drastically reducing computational time. Working with local reference libraries is particularly important in broad taxonomic studies that target rare species and require extensive blasting of individual sequences (Chain et al. 2016). For many specific applications, users must ensure that reference databases are up-to-date and contain entries for species of interest. An accurate taxonomic assignment provides a robust way of linking genotype to phenotype and to specific ecological roles (Creer et al. 2016, Ratnasingham & Hebert 2007).

## 4. FUTURE USE OF ENVIRONMENTAL DNA IN BIODIVERSITY SCIENCE AND CONSERVATION

### 4.1. Overcoming Skepticism: Improving Reproducibility and Comparability

For eDNA-based methods to gain endorsement as legitimate tools that monitor species distributions and estimate trends in their abundance, several important factors need to be addressed. First, accurate calibration and validation are essential to reduce both false positives and false negatives (Deiner et al. 2015, Schmelzle & Kinziger 2016). Improvements in validation steps are required at all stages of the analytical chain: sample collection, primer selection, high-throughput sequencing technology, bioinformatics, and taxonomic assignment. Studies without validation are likely to be exposed to multiple levels of errors (**Table 1**) with potentially synergistic effects.

Second, there is great need for critical and robust testing of the major assumptions involved in the use of eDNA, particularly around its origin, persistence, and transportation (Barnes & Turner 2016). Widespread skepticism, particularly in the management community, will be overcome only by refined techniques and strengthened protocols for data interpretation. Comparison is needed between results from eDNA analysis and those obtained through traditional approaches as well as among results generated by different labs under multiple conditions (Beja-Pereira et al. 2009). The key requirement is to achieve a fine balance between employing protocols that are best able to address the research question in the specific environment under study and promoting a consistent level of validation and standardization. With such steps in place, central problems can be addressed and results meaningfully compared.

### 4.2. Conclusions

The use of eDNA analysis as a tool for biodiversity science and conservation is gaining momentum. However, protocols still need to be standardized and the reproducibility of results demonstrated. Moreover, the physical and ecological limits of using eDNA require more investigation. Methods based on eDNA will not provide a universal solution to biodiversity assessment. They can provide powerful insights, particularly in freshwater and marine environments, but their impact is likely to be less significant in terrestrial settings. Another major challenge is the need for a coordinated effort, not only for benchmarking methods, but also for integrating traditional approaches (taxonomic and ecological data) while implementing and constantly refining new technologies. Studies focused on identifying biodiversity trends, biodiversity hot spots, and invasive species as

well as those focused on prioritizing conservation efforts or revealing ecosystem-level processes will greatly benefit from insights derived from the analysis of eDNA. To reach acceptable standards, precautions similar to those established for aDNA protocols must be adopted (Willerslev et al. 2004) and possible sources of errors quantified at each major step of the process.

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships funding, or financial holdings that might affect the objectivity of this review.

## ACKNOWLEDGMENTS

We appreciate the invitation from the Associate Editor, Dr. Daniel Simberloff, to write this review. Vasco Elbrecht, Joanne Littlefair, Xavier Pochon, and Frederic Chain provided helpful comments on earlier versions of the manuscript. This work was supported by Discovery Grants from the Natural Sciences and Engineering Research Council of Canada and by the Canada Research Chair program to M.E.C. and P.D.N.H.

## LITERATURE CITED

- Aardema BW, Lorenz MG, Krumbein WE. 1983. Protection of sediment adsorbed transforming DNA against enzymatic inactivation. *Appl. Environ. Microbiol.* 46:417–20
- Andersen K, Bird KL, Rasmussen M, Haile J, Breuning-Madsen H, et al. 2011. Meta-barcoding of ‘dirt’ DNA from soil reflects vertebrate biodiversity. *Mol. Ecol.* 21:1966–79
- Anderson-Carpenter LL, McLachlan JS, Jackson ST, Kuch M, Lumibao CY, Poinar HN. 2011. Ancient DNA from lake sediments: bridging the gap between paleoecology and genetics. *BMC Evol. Biol.* 11:30
- Armstrong KF, Ball SL. 2005. DNA barcodes for biosecurity: invasive species identification. *Philos. Trans. R. Soc. B* 360:1813–23
- Bailiff DM, Karl DM. 1991. Dissolved and particulate DNA dynamics during a spring bloom in the Antarctic Peninsula region. *Deep-Sea Res.* 38:1077–95
- Barnes MA, Turner CR. 2016. The ecology of environmental DNA and implications for conservation genetics. *Conserv. Genet.* 17:1–17
- Barnes MA, Turner CR, Jerde CL, Renshaw MA, Chadderton WL, Lodge DM. 2014. Environmental conditions influence eDNA persistence in aquatic systems. *Environ. Sci. Technol.* 48:1819–27
- Beja-Pereira A, Oliveira R, Schwartz MK, Luikart G. 2009. Advancing ecological understanding through technological transformation in noninvasive genetics. *Mol. Ecol. Res.* 9:1279–301
- Bienert R, de Danieli S, Miquel C, Coissac E, Poillot C, et al. 2012. Tracking earthworm communities from soil DNA. *Mol. Ecol.* 21:2017–30
- Biggs J, Ewald N, Valentini A, Gaboriaud C, Dejean T, et al. 2015. Using eDNA to develop a national citizen science-based monitoring program for the great crested newt (*Triturus cristatus*). *Biol. Conserv.* 183:19–28
- Birks HJB, Birks HH. 2016. How have studies of ancient DNA from sediments contributed to the reconstruction of Quaternary floras? *New Phytol.* 209:499–506
- Blaalid R, Carlsen T, Kumar S, Halvorsen R, Ugland KI, et al. 2012. Changes in the root-associated fungal communities along a primary succession gradient analysed by 454 pyrosequencing. *Mol. Ecol.* 21:1827–908
- Bohmann K, Evans A, Gilbert TP, Carvalho GR, Creer S, et al. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol. Evol.* 29:358–67
- Breshers DD, Whicker JJ, Johansen MP, Pinder JE. 2003. Wind and water erosion and transport in semi-arid shrubland, grassland and forest ecosystems: quantifying dominance of horizontal wind-driven transport. *Earth Surf. Process. Landf.* 28:1189–209



- Brown EA, Chain FJJ, Crease TJ, MacIsaac HJ, Cristescu ME. 2015. Divergence thresholds and divergent biodiversity estimates: Can metabarcoding reliably describe zooplankton communities? *Ecol. Evol.* 5:2234–51
- Brown EA, Chain FJJ, Zhan A, MacIsaac HJ, Cristescu ME. 2016. Early detection of aquatic invaders using metabarcoding reveals a high number of non-indigenous species in Canadian ports. *Divers. Distrib.* 22:1045–59
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johanson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13:581–83
- Caesar RM, Sorensson M, Cognato AI. 2006. Integrating DNA data and traditional taxonomy to streamline biodiversity assessment: an example from edaphic beetles in the Klamath ecoregion, California, USA. *Divers. Distrib.* 12:483–89
- Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings CV, et al. 2012. Biodiversity loss and its impact on humanity. *Nature* 486:59–67
- Carr SM, Marshall HD, Duggan AT, Flynn SMC, Johnstone KA, et al. 2008. Phylogeographic genomics of mitochondrial DNA: highly-resolved patterns of intraspecific evolution and a multi-species, microarray-based DNA sequencing strategy for biodiversity studies. *Comp. Biochem. Physiol. D* 3(1):1–11
- Chain FJJ, Brown EA, MacIsaac HJ, Cristescu ME. 2016. Metabarcoding reveals strong spatial structure and temporal turnover of zooplankton communities among marine and freshwater ports. *Divers. Distrib.* 22:493–504
- Clare EL. 2014. Molecular detection of trophic interactions: emerging trends, distinct advantages, significant considerations and conservation applications. *Evol. Appl.* 7:1144–57
- Clare EL, Chain FJJ, Littlefair JE, Cristescu ME. 2016. The effects of parameter choice on defining molecular operational taxonomic units and resulting ecological analysis of metabarcoding data. *Genome* 59:981–90
- Clarke LJ, Beard JM, Swadling KM, Deagle BE. 2017. Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. *Ecol. Evol.* 7:873–83
- Coissac E, Riaz T, Puillandre N. 2012. Bioinformatic challenges for DNA metabarcoding of plants and animals. *Mol. Ecol.* 21:1834–47
- Cordier T, Esling P, Lejzerowicz F, Visco J, Ouadahi A, et al. 2017. Predicting the ecological quality status of marine environments from eDNA metabarcoding data using supervised machine learning. *Environ. Sci. Technol.* 51:9118–26
- Creer S, Deiner K, Frey S, Porazinska D, Taberlet P, Thomas WK, et al. 2016. The ecologist's field guide to sequence-based identification of biodiversity. *Methods Ecol. Evol.* 7:1008–18
- Cristescu ME. 2014. From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends Ecol. Evol.* 29:566–71
- Darling JA, Blum JJ. 2007. DNA-based methods for monitoring invasive species: a review and prospectus. *Biol. Invasions* 9:751–65
- Darling JA, Mahon AR. 2011. From molecules to management: adopting DNA-based methods for monitoring biological invasions in aquatic environments. *Environ. Res.* 111:978–88
- Deagle BE, Kirkwood R, Jarman SN. 2014. Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Mol. Ecol.* 18:2022–38
- Deiner K, Altermatt F. 2014. Transport distance of invertebrate environmental DNA in a natural river. *PLOS ONE* 9:e88786
- Deiner K, Bik HM, Machler E, Seymour M, Lacoursiere-Roussel A, et al. 2017b. Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Mol. Ecol.* 26:5872–95
- Deiner K, Fronhofer EA, Machler E, Walser J-C, Altermatt F. 2016. Environmental DNA reveals that rivers are conveyor belts of biodiversity information. *Nat. Commun.* 7:12544
- Deiner K, Renshaw MA, Yiyuan LI, Olds BP, Lodge DM. 2017a. Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA. *Methods Ecol. Evol.* 8:1888–98
- Deiner K, Walser JC, Machler E, Altermatt F. 2015. Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biol. Conserv.* 183:53–63
- Dejean T, Valentini A, Duparc A, Pellier-Cuit S, Pompanon F, et al. 2011. Persistence of environmental DNA in freshwater ecosystems. *PLOS ONE* 6:e23398



- Dejean T, Valentini A, Miquel C, Taberlet P, Bellemain E, Miaud C. 2012. Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog *Lithobates catesbeianus*. *J. Appl. Ecol.* 49:953–59
- Ficetola GF, Miaud C, Pompanon F, Taberlet P. 2008. Species detection using environmental DNA from water samples. *Biol. Lett.* 4:423–25
- Ficetola GF, Pansu J, Bonin A, Coissac E, Giguët-Covex C, et al. 2014. Replication levels, false presences and the estimation of the presence absence from eDNA metabarcoding data. *Mol. Ecol. Res.* 15:543–56
- Flynn JM, Brown EA, Chain FJJ, MacIsaac HJ, Cristescu ME. 2015. Towards accurate molecular identification of species in complex environmental samples: testing the performance of sequence filtering and clustering methods. *Ecol. Evol.* 5:2252–66
- Footo A, Thomsen P, Sveegaard S, Wahlberg M, Kielgast J, et al. 2012. Investigating the potential use of environmental DNA (eDNA) for genetic monitoring of marine mammals. *PLOS ONE* 7:e4178
- Franklin JF. 1993. Preserving biodiversity: species, ecosystems, or landscapes. *Ecol. Appl.* 3:202–5
- Freeland JR. 2017. The importance of molecular markers and primer design when characterizing biodiversity from environmental DNA. *Genome* 60:358–74
- Fujiwara A, Matsuhashi S, Doi H, Yamamoto S, Minamoto T. 2016. Use of environmental DNA to survey the distribution of an invasive submerged plant in ponds. *Freshw. Sci.* 35:748–54
- Gardham S, Hose GC, Stephenson S, Chariton AA. 2014. DNA metabarcoding meets experimental ecotoxicology: advancing knowledge on the ecological effect of copper in freshwater ecosystems. *Adv. Ecol. Res.* 51:79–104
- Geml J, Laursen GA, Timling I, McFarland JM, Booth MG, et al. 2009. Molecular phylogenetic biodiversity assessment of arctic and boreal ectomycorrhizal *Lactarius* Pers. (Russulales; Basidiomycota) in Alaska, based on soil and sporocarp DNA. *Mol. Ecol.* 18:2213–27
- Giguët-Covex C, Pansu J, Arnaud F, Rey PJ, Griggo C, et al. 2014. Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nat. Commun.* 5:32311
- Glenn TC. 2011. Field guide to next-generation DNA sequencing. *Mol. Ecol. Resour.* 11:759–69
- Goldberg CS, Pilliod DS, Arkle RS, Waits LP. 2011. Molecular detection of vertebrates in stream water: a demonstration using Rocky Mountain tailed frogs and Idaho giant salamanders. *PLOS ONE* 6:e22746
- Goldberg CS, Turner CR, Deiner K, Klymus KE, Francis P, et al. 2016. Critical consideration for the application of environmental DNA methods to detect aquatic species. *Methods Ecol. Evol.* 7:1299–307
- Goodwin S, McPherson JD, McPherson JD, McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17:333–51
- Haile J, Holdaway R, Oliver K, Bunce M, Nielsen R, et al. 2007. Ancient DNA chronology within sediment deposits: Are paleobiological reconstructions possible and is DNA leaching a factor? *Mol. Biol. Evol.* 24:982–89
- Hajibabaei M, Singer GAC, Clare El, Hebert PDN. 2007. Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. *BMC Biol.* 5:24
- Harris JD. 2003. Can you bank on GenBank? *Trends Ecol. Evol.* 18:317–19
- Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. *Proc. Biol. Sci.* 270:313–21
- Hebert PDN, Hollingsworth PM, Hajibabaei M. 2016. From writing to reading the encyclopedia of life. *Phil. Trans. R. Soc. B* 371:20150321
- Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, et al. 2011. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* 83:8604–10
- Höss M, Jaruga P, Zastawny T, Dizdaroglu M, Pääbo S. 1996. DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Res.* 24:1304–7
- Hunter ME, Dorazio RM, Butterfield JSS, Meigs-Friend G, Nico LG, Ferrante JA. 2017. Detection limits of quantitative and digital PCR assays and their influence in presence-absence surveys of environmental DNA. *Mol. Ecol. Res.* 17:221–29
- Ibáñez de Aldecoa AL, Zafra O, Gonzalez-Pastor JE. 2017. Mechanisms and regulation of extracellular DNA release and its biological roles in microbial communities. *Front. Microbiol.* 8:1390
- Jerde CL, Chadderton WL, Mahon AR, Renshaw MA, Corush J, et al. 2013. Detection of Asian carp DNA as part of a Great Lakes basin-wide surveillance program. *Can. J. Fish. Aquat. Sci.* 70:522–26



- Jerde CL, Mahon AR, Chadderton WL, Lodge DM. 2011. "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conserv. Lett.* 4:150–57
- Kelly RP, Port JA, Yamahara KM, Crowder LB. 2014a. Using environmental DNA to census marine fishes in a large mesocosm. *PLOS ONE* 9:e86175
- Kelly RP, Port JA, Yamahara KM, Martone RG, Lowell N, et al. 2014b. Harnessing DNA to improve environmental management. *Science* 344:1455–56
- Kirshtein JD, Anderson CW, Wood JS, Longcore JE, Voytek MA. 2007. Quantitative PCR detection of *Batrachochytrium dendrobatidis* DNA from sediments and water. *Dis. Aquat. Organ.* 77:11–15
- Koski LB, Golding GB. 2001. The closest BLAST hit is often not the nearest neighbor. *J. Mol. Evol.* 52:540–42
- Kraaijeveld K, De Wever LA, Ventayol Garcia M, Buermans H, et al. 2015. Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. *Mol. Ecol. Res.* 15:8–16
- Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. 2010. Wrinkles in the rare biosphere: Pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* 12:118–23
- Lacoursiere-Roussel A, Rosebal M, Bertnatchez L. 2016. Estimating fish abundance and biomass from eDNA concentrations: variability among capture methods and environmental conditions. *Mol. Ecol. Res.* 16:1401–14
- Laszlo AH, Derrington IM, Ross BC, Brinkerhoff H, Adey A, et al. 2014. Decoding long nanopore sequencing reads of natural DNA. *Nat. Biotechnol.* 32:829–33
- Leray M, Knowlton N. 2015. DNA barcoding and metabarcoding of standardized sampling reveal patterns of marine benthic diversity. *PNAS* 112:2076–81
- Lindahl T. 1993. Instability and decay of the primary structure of DNA. *Nature* 362:709–15
- Lodge DM, Turner CR, Jerde CL, Barnes MA, Chadderton L, et al. 2012. Conservation in a cup of water: estimating biodiversity and population abundance from environmental DNA. *Mol. Ecol.* 21:2555–58
- Mahé F, Mayor J, Bunge J, Chi J, Siemensmeyer T, Stoeck T, et al. 2015. Comparing high-throughput platforms for sequencing the V4 region of SSU-rDNA in environmental microbial eukaryotic diversity surveys. *J. Eukaryot. Microbiol.* 62:338–45
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–80
- Martin J-L, Maris V, Simberloff DS. 2016. The need to respect nature and its limit challenges society and conservation science. *PNAS* 113:6105–12
- Mermillod-Blonding F. 2011. The functional significance of bioturbation and biodeposition on biogeochemical processes at the water-sediment interface in freshwater and marine ecosystems. *J. N. Am. Benthol. Soc.* 30:770–78
- Mullis KB, Faloona F. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* 155:335–50
- Nathan LM, Simmons M, Wegleitner BJ, Jerde CL, Mahon AR. 2014. Quantifying environmental DNA signals for aquatic invasive species across multiple detection platforms. *Environ. Sci. Technol.* 48:12800–6
- Naem S. 2002. Ecosystem consequences of biodiversity loss: the evolution of a paradigm. *Ecology* 83:1537–52
- Ogram A, Sayler GS, Barkay T. 1987. The extraction and purification of microbial DNA from sediments. *J. Microbiol. Methods* 7:57–66
- Pace NR. 1997. A molecular view of microbial diversity and the biosphere. *Science* 276:734–40
- Pansu J, Giguët-Covex C, Ficetola GF, Gielly L, Boyer F, et al. 2015. Reconstructing long-term human impacts on plant communities. An ecological approach based on lake sediment DNA. *Mol. Ecol.* 24:1485–98
- Paul JH, Jeffrey WH, DeFlaun MF. 1987. Dynamics of extracellular DNA in the marine environment. *Appl. Environ. Microbiol.* 53:170–79
- Parameswaran P, Jalili R, Tao L, Shokralla S, Gharizadeh B, Ronaghi M, et al. 2007. A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Res.* 35:e130
- Parducci L, Matetovici I, Fontana S, Bennett KD, Suyama Y, et al. 2013. Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia. *Mol. Ecol.* 22:3511–24
- Pawlowski J, Lejzerowicz F, Apotheloz-Perret-Gentil L, Visco J, Esling P. 2016. Protist metabarcoding and environmental biomonitoring: time for change. *Eur. J. Protistol.* 55:12–25



- Pfunder M, Holzgang O, Frey JE. 2004. Development of microarray-based diagnostics of voles and shrews for use in biodiversity monitoring studies, and evaluation of mitochondrial cytochrome oxidase I versus cytochrome b as genetic markers. *Mol. Ecol.* 13:1277–86
- Pietramellara G, Ascher J, Borgogni F, Ceccherini MT, Guerri G, Nannipieri P. 2009. Extracellular DNA in soil and sediment: fate and ecological relevance. *Biol. Fertil. Soils* 45:219–35
- Piggott MP. 2016. Evaluating the effects of laboratory protocols on eDNA detection probability for an endangered freshwater fish. *Ecol. Evol.* 6:2739–50
- Pilliod DS, Goldberg CS, Arkle RS, Waits LP. 2013. Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Can. J. Fish. Aquat. Sci.* 70:1123–30
- Piñol J, Mir G, Gomez-Polo P, Agust N. 2015. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Mol. Ecol. Resour.* 15:819–30
- Pochon X, Zaiko A, Fletcher LM, Laroche O, Wood SA. 2017. Wanted dead or alive? Using metabarcoding of environmental DNA and RNA to distinguish living assemblages for biosecurity applications. *PLOS ONE* 12:e0187636
- Porter TM, Hajibabaei M. 2018. Scaling up: a guide to high throughput genomic approaches for biodiversity analysis. *Mol. Ecol.* 27:313–38
- Ratnasingham S, Hebert PDN. 2007. BOLD: the Barcode of Life Data System. *Mol. Ecol. Res.* 7:355–64
- Rees HC, Maddison BC, Middleditch DJ, Patmore JR, Gough KC. 2014. The detection of aquatic animal species using environmental DNA—a review of eDNA as a survey tool in ecology. *J. Appl. Ecol.* 51:1450–59
- Roussel J-M, Paillisson J-M, Treguier A, Petit E. 2015. The downside of eDNA as a survey tool in water bodies. *J. Appl. Ecol.* 52:823–26
- Sanger FS, Nicklen ARC. 1977. DNA sequencing with chain-terminating inhibitors. *PNAS* 74:5463–67
- Sansom BJ, Sassoubre LM. 2017. Environmental DNA (eDNA) shedding and decay rates to model freshwater mussel eDNA transport in a river. *Environ. Sci. Technol.* 51:14244–53
- Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S. 2012. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLOS ONE* 7:e34131
- Schmelzle MC, Kinziger AP. 2016. Using occupancy modelling to compare environmental DNA to traditional field methods for regional-scale monitoring of an endangered aquatic species. *Mol. Ecol. Resour.* 16:895–908
- Scott R, Zhan A, Brown EA, Chain FJJ, Cristescu ME, et al. 2018. Optimization and performance testing of a sequence processing pipeline applied to detection of nonindigenous species. *Evol. Appl.* 11:891–905
- Scriver M, Marinich A, Wilson C, Freeland J. 2015. Development of species-specific environmental DNA (eDNA) markers for invasive aquatic plants. *Aquat. Bot.* 122:27–31
- Shaw JL, Clarke LJ, Wedderburn SD, Barnes TC, Weyrich LS, Cooper A. 2016. Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system. *Biol. Conserv.* 197:131–38
- Shokralla S, Spall JL, Gibson JF, Hajibabaei M. 2012. Next-generation sequencing technologies for environmental DNA research. *Mol. Ecol.* 21:1794–805
- Sigsgaard EE, Nielsen IB, Bach SS, Lorenzen ED, Robinson DP, Knudsen SW, et al. 2016. Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nat. Ecol. Evol.* 1:0004
- Simberloff D. 2005. The politics of assessing risk for biological invasions: the USA as a case study. *Trends Ecol. Evol.* 20:216–22
- Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, et al. 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere.” *PNAS* 103:12115–20
- Sutherland WJ, Bardsley S, Clout M, Depledge MH, Dicks LV, et al. 2013. A horizon scan of global conservation issues for 2013. *Trends Ecol. Evol.* 28:16–22
- Swank WT, Vanlear DH. 1992. Multiple-use management: ecosystem perspectives of multiple-use management. *Ecol. Appl.* 2:219–20
- Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH. 2012. Environmental DNA. *Mol. Ecol.* 21:1789–93
- Teal LR, Bulling MT, Parker ER, Solan M. 2008. Global patterns of bioturbation intensity and the mixed depth of marine soft sediments. *Aquat. Biol.* 2:207–18



- Thomsen PF, Kielgast J, Iversen LL, Wiuf C, Rasmussen M, et al. 2012. Monitoring endangered freshwater biodiversity by environmental DNA. *Mol. Ecol.* 21:2565–73
- Thomsen PF, Willerslev E. 2015. Environmental DNA—an emerging tool in conservation for monitoring past and present biodiversity. *Biol. Conserv.* 183:4–18
- Treguier A, Paillisson J-M, Dejean T, Valentini A, Schlaepfer MA, Rousell J-M. 2014. Environmental DNA surveillance for invertebrate species: advantage and technical limitations to detect invasive crayfish *Procambarus clarkii* in freshwater ponds. *J. Appl. Ecol.* 51:871–79
- Tremblay J, Singh K, Fern A, Kirton ES, He S, et al. 2015. Primer and platform effects on 16S rRNA tag sequencing. *Front. Microbiol.* 6:771
- Turner CR, Uy KL, Everhart RC. 2015. Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biol. Conserv.* 183:93–102
- Valentini A, Taberlet P, Miaud C, Civade R, Herder J, et al. 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Mol. Ecol.* 25:929–42
- Virgilio M, Backeljau T, Nevado B, De Meyer M. 2010. Comparative performances of DNA barcoding across insect orders. *BMC Bioinform.* 11:206
- Wilcox TM, McKelvey KS, Young MK, Sepulveda AJ, Shepard BB, et al. 2016. Understanding environmental DNA detection probabilities: a case study using a stream-dwelling char *Salvelinus fontinalis*. *Biol. Conserv.* 194:209–16
- Willerslev E, Hansen AJ, Binladen J, Brand TB, Gilbert MTP, et al. 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* 300:791–95
- Willerslev E, Hansen AJ, Poinar HN. 2004. Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends Ecol. Evol.* 19:141–47
- Wood SA, Smith KF, Banks JC, Tremblay LA, Rhodes L, et al. 2013. Molecular genetic tools for environmental monitoring of New Zealand's aquatic habitats, past, present and the future. *N.Z. J. Mar. Freshw. Res.* 47:90–119
- Yeates DK, Seago A, Nelson L, Cameron SL, Joseph L, Trueman JWH. 2010. Integrative taxonomy, or iterative taxonomy? *Syst. Entomol.* 36:209–17
- Yoccoz NG, Bråthen KA, Gielly L, Haile J, Eduards ME, Goslar T, et al. 2012. DNA from soil mirrors plant taxonomic and growth form diversity. *Mol. Ecol.* 21:3647–55

