

## REVIEW

## Methods for the extraction, storage, amplification and sequencing of DNA from environmental samples

Gavin Lear<sup>1\*</sup>, Ian Dickie<sup>2</sup>, Jonathan Banks<sup>3</sup>, Stephane Boyer<sup>4</sup>, Hannah L. Buckley<sup>5</sup>, Thomas R. Buckley<sup>1,6</sup>, Rob Cruickshank<sup>7</sup>, Andrew Dopheide<sup>6</sup>, Kim M. Handley<sup>1</sup>, Syrie Hermans<sup>1</sup>, Janine Kamke<sup>1</sup>, Charles K. Lee<sup>8</sup>, Robin MacDiarmid<sup>9</sup>, Sergio E. Morales<sup>10</sup>, David A. Orlovich<sup>11</sup>, Rob Smissen<sup>12</sup>, Jamie Wood<sup>12</sup> and Robert Holdaway<sup>12</sup>

<sup>1</sup>School of Biological Sciences, University of Auckland, 3a Symonds Street, Auckland 1010, New Zealand

<sup>2</sup>School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand

<sup>3</sup>The Cawthron Institute, Nelson 7010, New Zealand

<sup>4</sup>Insect Biology Research Institute (IRBI) – UMR 7261 CNRS / Université François-Rabelais de Tours, Parc Grandmont, 37200 Tours, France

<sup>5</sup>School of Science, Auckland University of Technology, Auckland 1010, New Zealand

<sup>6</sup>Landcare Research, Auckland 1142, New Zealand

<sup>7</sup>Department of Ecology, Lincoln University, Lincoln 7647, New Zealand

<sup>8</sup>School of Sciences, The University of Waikato, Hamilton 3240, New Zealand

<sup>9</sup>Plant and Food Research, Auckland 1025, New Zealand

<sup>10</sup>Microbiology and Immunology, University of Otago, Dunedin 9054, New Zealand

<sup>11</sup>Department of Botany, University of Otago, Dunedin 9054, New Zealand

<sup>12</sup>Landcare Research, Lincoln 7640, New Zealand

\*Author for correspondence (Email: g.lear@auckland.ac.nz)

Published online: 11 December 2017

**Abstract:** Advances in the sequencing of DNA extracted from media such as soil and water offer huge opportunities for biodiversity monitoring and assessment, particularly where the collection or identification of whole organisms is impractical. However, there are myriad methods for the extraction, storage, amplification and sequencing of DNA from environmental samples. To help overcome potential biases that may impede the effective comparison of biodiversity data collected by different researchers, we propose a standardised set of procedures for use on different taxa and sample media, largely based on recent trends in their use. Our recommendations describe important steps for sample pre-processing and include the use of (a) Qiagen DNeasy PowerSoil<sup>®</sup> and PowerMax<sup>®</sup> kits for extraction of DNA from soil, sediment, faeces and leaf litter; (b) DNeasy PowerSoil<sup>®</sup> for extraction of DNA from plant tissue; (c) DNeasy Blood and Tissue kits for extraction of DNA from animal tissue; (d) DNeasy Blood and Tissue kits for extraction of DNA from macroorganisms in water and ice; and (e) DNeasy PowerWater<sup>®</sup> kits for extraction of DNA from microorganisms in water and ice. Based on key parameters, including the specificity and inclusivity of the primers for the target sequence, we recommend the use of the following primer pairs to amplify DNA for analysis by Illumina MiSeq DNA sequencing: (a) 515f and 806RB to target bacterial 16S rRNA genes (including regions V3 and V4); (b) #3 and #5RC to target eukaryote 18S rRNA genes (including regions V7 and V8); (c) #3 and #5RC are also recommended for the routine analysis of protist community DNA; (d) ITS6F and ITS7R to target the chromistan ITS1 internal transcribed spacer region; (e) S2F and S3R to target the ITS2 internal transcribed spacer in terrestrial plants; (f) fITS7 or gITS7, and ITS4 to target the fungal ITS2 region; (g) NS31 and AML2 to target glomeromycota 18S rRNA genes; and (h) mICOLintF and jgHCO2198 to target cytochrome *c* oxidase subunit I (*COI*) genes in animals. More research is currently required to confirm primers suitable for the selective amplification of DNA from specific vertebrate taxa such as fish. Combined, these recommendations represent a framework for efficient, comprehensive and robust DNA-based investigations of biodiversity, applicable to most taxa and ecosystems. The adoption of standardised protocols for biodiversity assessment and monitoring using DNA extracted from environmental samples will enable more informative comparisons among datasets, generating significant benefits for ecological science and biosecurity applications.

**Key words:** biological heritage; biodiversity monitoring; community profiling; DNA primers; DNA sequencing; eDNA; environmental DNA; Illumina; metabarcoding; metagenomics; molecular ecology

## Introduction

The first widely-used DNA sequencing approach (Sanger et al. 1977) was sufficient to sequence the human genome (Venter et al. 2001), but the limited throughput of this technique remains a major constraint on its use for the analysis of complex DNA pools. Even following the release of high-throughput pyrosequencing platforms such as the 454 Genome Sequencer Instrument, capable of generating 25 million bases in a single 4-hour run (Margulies et al. 2005), DNA-based analyses of biological communities in multiple samples remained problematic due to difficulties associated with combining and later identifying DNA originating from many different samples. More recently, multiplex primer labelling approaches have been developed that, after DNA sequencing, allow the user to determine which DNA sequences originated from each of multiple DNA samples combined in a single solution prior to their analysis. Today, DNA from many hundreds of samples can be combined and analysed in parallel (Barberan et al. 2014, 2015; Shokralla et al. 2015). This ability to generate large amounts of sequence data from numerous samples in parallel offers huge potential for using DNA to monitor the biological diversity in any sample from which DNA can be extracted. Significantly, these methods are applicable to all organisms (e.g. Archaea, Bacteria, Protista, Fungi, Animalia and Plantae), all genetic markers (e.g. 16S, 18S, ITS, COI and *rbcL*) and all sample media (e.g. soil, water, air and tissue) such that DNA analysis protocols could provide a universal tool for future biodiversity and biosecurity assessments. The combination of the mass amplification of these genetic markers or ‘DNA barcodes’ by PCR, with high-throughput DNA sequencing to identify a mixture of organisms is most commonly referred to as ‘metabarcoding’. We seek to overcome the currently fragmented understanding of the identity and distribution of both native and introduced species, using a unified DNA metabarcoding approach for high throughput assessments of communities across all domains of life.

While DNA-based biodiversity assessment methods are not yet in widespread use beyond the microbial world, there are many potential benefits, as well as uncertainties, resulting from their application to a far wider range of organisms, as reviewed in Holdaway et al. (2017). For example, traditional observational techniques for biodiversity monitoring can be highly dependent on (and biased by) taxonomic and diagnostic expertise that is in scarce supply worldwide (Paknia et al. 2015). Cryptic species can be misidentified and whole taxa may be underrepresented or overlooked due to factors including their small size, nocturnal habits, occurrence in less-accessible habitats (e.g. below ground), or the non-random movement of study organisms in response to disturbance while being surveyed (Watson et al. 1995). In contrast, samples for DNA metabarcoding may be collected by non-specialists and may not require invasive sampling protocols since sampling or capture of whole, individual organisms is typically not required. Many environmental substrates are easy to sample (e.g. soils) and contain significant populations of micro-organisms, as well as small invertebrates, which will be represented in the DNA extracted from these substrates. In addition to the cellular DNA that may be directly extracted from communities of organisms, large quantities of ‘environmental DNA’ are continually excreted and shed in the environment by living organisms. For example, animals can be detected based on DNA excreted into environments from their urine (Valiere & Taberlet 2000), faeces (Kurose et al. 2005), hair and skin

(Henry & Russello 2011). Similarly, plant DNA originating from roots, root exudates and litter can provide information about plant community composition (Yoccoz et al. 2012). Consequently, DNA extracted from samples of soil, water or other material may simultaneously provide information about the occurrence, distribution and diversity of organisms and communities, across multiple branches of life. The detection of organisms from environmental DNA has an additional benefit in allowing the detection of transient organisms, which may be missed by traditional observational sampling. Once collected, environmental DNA can be stored for long periods, providing a library of sample material that can be accessed at any time for re-analysis. This library creates opportunities for investigations of taxa, genes, and hypotheses that were not considered in the original study. The DNA sequence analysis of large sample numbers using metabarcoding approaches is becoming a more cost-competitive method for biodiversity and biosecurity monitoring as sequencing technologies advance. With a plethora of DNA extraction, storage, amplification and sequencing methods available it is not possible to provide the exact costs associated with these procedures. As an indicative cost, and not including the costs of human resources, the processing and analysis of over 300 samples is achievable (as of September 2017) for under NZ\$12 000, excluding general sales tax (~NZ\$3500 for DNA extraction; ~NZ\$600 for DNA amplification; ~NZ\$600 for PCR purification, ~NZ\$7000 for analysis of 384 samples on an Illumina MiSeq DNA sequencing machine).

While DNA sampling and analysis holds much promise for use in biological heritage monitoring and assessment, a plethora of techniques are currently being used, not only for sample collection, but also the extraction, amplification, sequencing and storage of DNA from environmental samples. These methods, which vary among research groups focusing on different taxa and sample media, and even between individual researchers within these groups, mean that comparisons of data across studies are subject to multiple, poorly quantified biases. Additionally, it can take some time for new researchers to select from the often daunting list of sample processing options before commencing their own analysis. To address these shortcomings, we propose that standardised protocols should be promoted for the extraction, storage, amplification and sequencing of environmental DNA. These will help to reduce biases associated with the comparison of sample data collected by different researchers, and could provide additional opportunities for collaboration and sharing of data. In addition to providing a framework that existing researchers can choose to follow, we recommend here a standard set of methodologies that may also help to make DNA metabarcoding protocols more accessible to less experienced users.

We summarise the identified and perceived issues associated with the extraction, amplification, sequencing and storage of DNA from environmental samples for metabarcoding analyses, before identifying future research opportunities and suggesting a standard set of methodologies to be employed. Specifically, we summarise methods associated with the analysis of short DNA fragments, or ‘amplicons’, amplified for taxonomic purposes, which is often referred to as ‘metabarcoding’ (Escobar-Zepeda et al. 2015) and is distinct from metagenomics approaches. The latter seeks to directly analyse the genomes contained in an environmental sample and typically does not target individual genes for analysis. There are many advantages associated with adopting well-defined protocol standards for the generation and analysis of DNA

amplicon data. The recommendation of a standardised set of methodologies provides additional impetus for researchers to test the impact and implications of their use of alternative methods for assessments of community composition including error rates for the false positive and negative detection of organisms from sample media. Thus, the importance of biases

associated with various aspects of DNA metabarcoding may be better quantified and understood. A number of reviews have recently been published to synthesise achievements in environmental DNA research (Goldberg et al. 2016) and to highlight the advantages and limitations of these methods for biodiversity assessment, but largely focused on assessments of

### Glossary of terms

**Amplicon:** A piece of DNA that is the source or product of natural or artificial replication events, such as DNA fragments generated during **PCR**. Typically **PCR** amplicon lengths are no greater than 5000 nucleotide bases in length, although longer **PCR** products may be generated using specialised DNA **polymerases**.

**Blocking primer:** These are short DNA sequences, or ‘**primers**’ used to block the amplification of specific DNA fragments. They are most commonly used to maximise the detection of low-abundance DNA sequences. For example, to increase the detection rates of animal DNA in the gut of a carnivore, blocking primers may be used to inhibit amplification of the host carnivores’ DNA. A blocking primer is typically designed to bind to the unwanted DNA sequence, but is modified in such a way that it does not prime amplification during **PCR**.

**Chimeras:** In the context of molecular DNA studies, a chimera refers to any sequence that is formed when two or more sequences are joined together during **PCR**. Chimeric DNA sequences can artificially inflate diversity estimates and must be removed during bioinformatics analysis.

**Deep-sequencing:** The ‘depth’ of sequencing refers to the number of reads obtained by DNA sequencing. Deep-sequencing will provide many sequences per sample, whereas shallow-sequencing provides fewer sequences per sample, usually due to more samples being combined in a single **multiplexed sequencing** run.

**Degenerate primers:** A mixture of similar, but not identical **primers** where one or more of the nucleotide bases (A, C, G, T) in the **primer** sequence varies. Degenerate **primers** are used to increase **primer universality**.

**DNA barcode:** a short DNA sequence found in an organism that can be used to identify it.

**DNA extraction:** A process whereby the DNA is separated from the sample media. This is commonly achieved using both physical and chemical methods to lyse cells and to separate the DNA from any contaminants or inhibitory substances associated with the sample material.

**DNA methylation:** A method used by cells to control gene expression. A methyl (CH<sub>3</sub>) group is added to a DNA strand, effectively fixing the gene in an ‘off’ position with regards to gene expression, but without changing the DNA sequence.

**Environmental DNA (eDNA):** DNA that is collected from an environment, rather than from an individual. Most commonly, this term is used to describe DNA that is no longer located in living cells (e.g. excreted DNA and DNA within cellular debris), although frequently it is also used to describe DNA extracted from environmental media such as soil including DNA extracted from micro- and macro-organisms within the sample.

**False priming:** When one or both of the **primers** used bind to a region of DNA outside of the target area, leading to amplification of unintended gene or non-gene regions. This result is often caused by non-specific binding by one or more of the bases in the primer sequences.

**Flow cell:** In the context of this review, flow cells are a surface on which sequencing chemistry occurs and over which sequencing polymerases, nucleotides and buffers can be pumped. DNA may be hybridised to flow cells in low molar quantities before sequencing *en masse*.

**GC-content:** A term which refers to the portion of guanine or cytosine bases that are present in a genome, gene or gene region. High GC-content DNA is more stable and tolerant to high temperatures than low GC-content DNA, due to the triple hydrogen bonds associated with the GC base pairing.

**Hairpin:** A U-shaped loop that is created when base pairs are formed between two different sections of the same DNA or RNA strand.

**Metabarcoding:** A molecular biodiversity detection method that uses short genetic markers, or **DNA barcodes**, to identify the presence of, and distinguish between, organisms in a sample.

**Metagenomics:** The study of the all the genetic material associated with an environment or sample.

**Microbiome:** The genetic material of all the microorganisms present in a particular environment or sample.

**Mock communities:** Cells or DNA derived from a pre-defined range of organisms, often mixed in known concentrations. Mock communities and mock community DNA are usually provided in the form of a solution that can be incorporated into experiments to determine the accuracy of DNA sequencing approaches.

**Multiplexed sequencing:** A large number of different samples are combined and sequenced simultaneously. This is achieved by adding short DNA sequences or ‘tags’ to the DNA sequence from each sample prior to sequence analysis.

**Next-generation sequencing (NGS):** High-throughput DNA sequencing technologies that allow highly **multiplexed sequencing**, without the need for a cloning step.

**Nuclear ribosomal DNA (nrDNA):** DNA regions that originate in the nucleus of organisms, as opposed to plastid or mitochondrial DNA.

**Paired end read:** Paired ends refer to the two ends of the same DNA molecule. In paired end sequencing, one end of the DNA sequence is read before DNA sequence analysis is initiated from the other end of the DNA molecule.

**Paraphyletic:** A taxonomic term that refers to organisms descended from a common evolutionary ancestor, but does not include all the descendants.

**PCR purification/PCR clean-up:** The process of purifying amplicons for downstream analyses. During this step, components left over from the **PCR** that were not incorporated into the **amplicons** (e.g. residual **primers**, primer dimers, dNTPs and **polymerases**) are removed.

**Plastid sequences:** Originate from the plastids of organisms, rather than the nucleus. Plastids are organelles that have their own DNA and ribosomes, and are found in the cells of plants, algae and some protists.

**Polymerase chain reaction (PCR):** A molecular method whereby a specific DNA sequence is amplified across several orders of magnitude. The product of this reaction is called an **amplicon**. A common variant on this approach, quantitative PCR (qPCR), allows quantification of the number of target DNA sequences present in the original sample.

**Polymerase:** An enzyme that synthesises DNA. Modified versions of this enzyme are available and are used for PCR.

**Polyphyletic:** A group of organisms composed of unrelated organisms descended from more than one ancestor.

**Primer universality:** A characteristic of a **primer** pair that determines how suitable it is to amplify the same gene region in a wide variety of different species. Increased primer universality means the DNA of a greater variety of organisms may be amplified by **PCR**.

**Primer:** A short strand of DNA that is required to initiate DNA synthesis. In PCR, a forward and reverse primer are used in combination to target a specific gene region for amplification.

**Sequence identity:** The extent to which one DNA sequence matches another, usually presented as the percentage of nucleotides (A, C, G, T), that correspond between the two sequences.

**Sequencing:** A process whereby the order of nucleotides within a DNA sequence is determined.

**Sequencing read length:** The number of nucleotide bases reported for a fragment of DNA following DNA sequence analysis.

**Single Nucleotide Polymorphism (SNP):** When a single nucleotide (A, T, C or G) in one genome varies from that in another genome.

**Template:** The DNA, obtained through DNA extraction from a sample, that is added to a **PCR**. It will normally contain the region of interest to be amplified.

**Virion:** The infectious form of a virus, which transports the viral genome between cells.

**Virome:** The genetic material from all the viruses associated with a host or environment.

specific taxa (Elbrecht & Leese 2015; Aylagas et al. 2016) or sample media (Drummond et al. 2015; Klymus et al. 2107). The aims of this review include, but are not restricted to, the following: (i) to review recent practices for the extraction, storage, amplification and sequencing of DNA from a broad range of environmental samples for the detection of a broad range of taxa, (ii) to recommend standard procedures for DNA extraction, amplification and sequencing of key taxa from the broadest range of environmental samples, and (iii) to identify emerging methods, such as new DNA sequencing approaches and shotgun metagenomics of relevance for future biodiversity assessments using DNA metabarcoding.

## Review of current practices for extraction, storage, amplification and sequencing of DNA from environmental samples

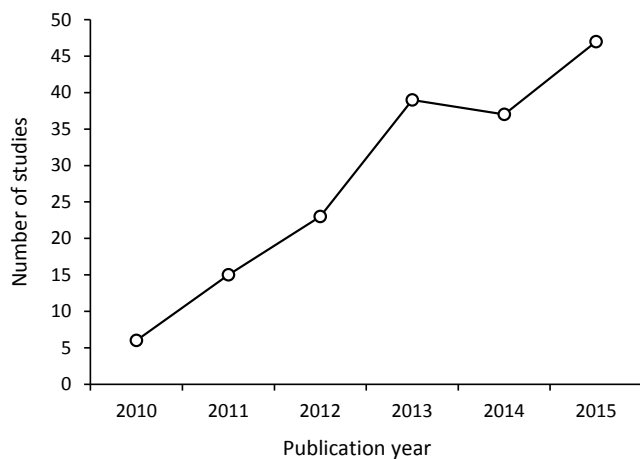
### Overview

We performed a literature review to summarise recent

approaches for the analysis of environmental DNA, or 'eDNA', from a broad range of taxa and environmental media. The aim of this analysis was not to conduct an exhaustive search of all eDNA literature, but rather to identify and summarise key methods for DNA extraction, storage, amplification and sequencing from environmental samples. Identification of methods consistently used for the analysis of particular sample media or taxa generates a framework on which to guide recommendations for future DNA metabarcoding research. The adoption of protocols that are both scientifically robust and in common use will allow newly-generated data to be directly compared to data from the largest number of existing studies. Conversely, where diverse approaches are used for similar research purposes (e.g. researchers target different gene regions to monitor the presence and abundance of the same taxa), these highlight areas in which the rational selection and recommendation of a standard method may be desirable.

## Methods

We performed a search of the ISI Web of Science Core Collection in August 2015 ([www.webofknowledge.com](http://www.webofknowledge.com)). We searched the literature for articles containing the terms 'environmental DNA' or 'eDNA' in the title, keywords or abstract. We refined our search terms to include only science and technology research published in the English language since 2010 and excluded books and conference proceedings by selecting only 'articles'. In total, we reviewed the full text of 584 articles. Articles that were determined to be reviews and perspectives were removed from the larger database. Articles



**Figure 1.** Number of eDNA studies published each year that were captured by our search of the ISI Web of Science Core Collection in August 2015 containing the terms 'environmental DNA' or 'eDNA'. We refined our search terms to include only science and technology research published in the English language since 2010 and excluded books and conference proceedings by selecting only 'articles'. Articles that were considered to be reviews and perspectives were removed from the larger database. Articles focused on 'extracellular DNA' were also removed.

focused on 'extracellular DNA' were similarly removed leaving a total of 167 articles for in-depth review.

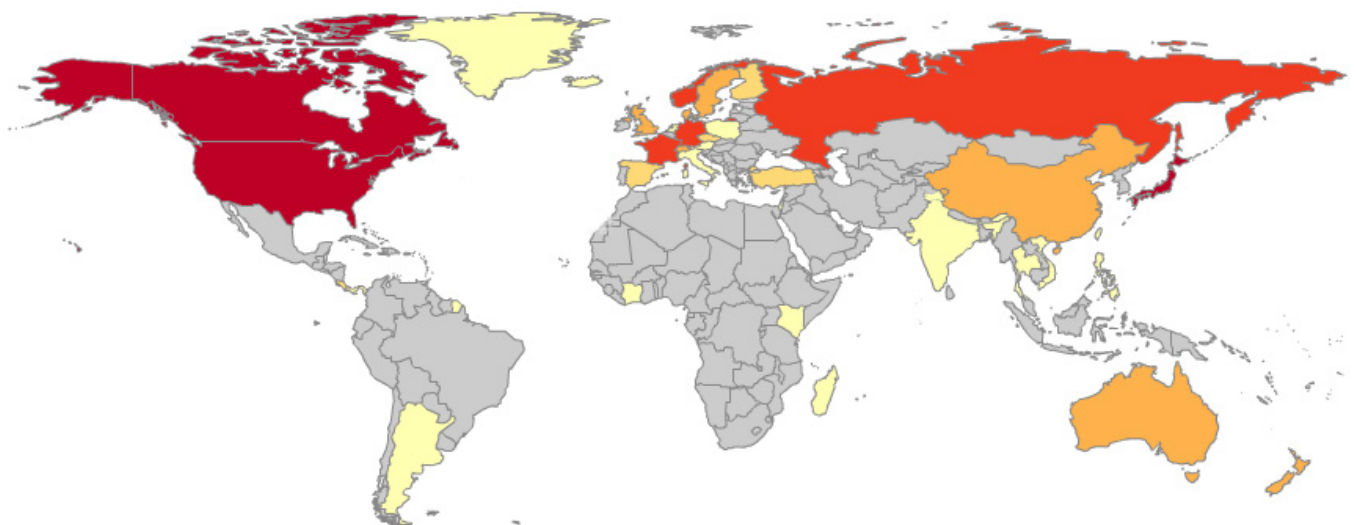
The terms 'eDNA' and 'environmental DNA' are largely redundant in studies of microbial communities and are used only infrequently by microbial researchers. As a consequence, we expect that studies on microbial DNA will be underrepresented due to our choice of search terms, even though almost all molecular studies on complex microbial communities can be considered as research on eDNA. Thus, it is important to note that our search terms were used to generate a varied dataset of eDNA-based studies focusing on a broad range of taxa and were never intended to capture all eDNA research. For example, there are 129 papers in the ISI Web of Science Core Collection up to 2015 that use the phrase 'metabarcoding' but not our search terms. Additionally, studies using alternative or more descriptive terms such as paleoenvironmental DNA (Rawlence et al. 2014) were not captured by our search terms meaning that research focusing on specific aspects of environmental DNA, such as ancient DNA may be underrepresented in our analysis.

## Summary of international eDNA research undertaken prior to August 2015

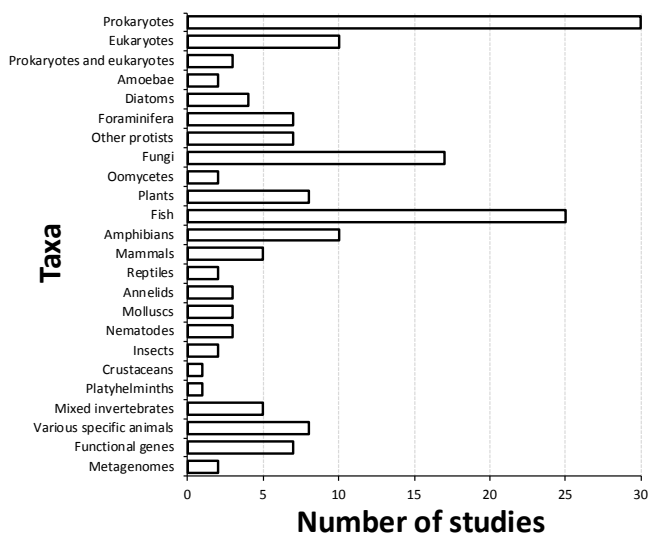
Of the 167 articles reviewed, ~75% were published in the previous 3 years; only six were published in 2010 (Fig. 1).

### Locations of prior investigations

Of the 167 articles reviewed, 44 were conducted in the United States of America, 11 in Japan and 11 in Canada (Fig. 2). Six studies were conducted in the open ocean, and eight were undertaken at a continental or global scale (i.e. data were selected from multiple countries). DNA metabarcoding studies were uncommon across the continents of South America and Africa. Five studies were included from New Zealand (Boyer et al. 2012; Collins et al. 2013; Pochon et al. 2013, 2015b; Teasdale et al. 2013); however, a large number of other studies were not included in our search results including Hamdan et al. (2011), Jangid et al. (2013), Hug et al. (2014), Koele et al. (2014), Sharp et al. (2014), Martinez-Garcia et al. (2015) and Morrison-Whittle & Goddard (2015).



**Figure 2.** Geographic locations of the 167 studies reviewed. Colours on map indicate the number of studies conducted in each country: (●)  $\geq 10$ ; (●) 8–9; (no countries belong to this category) 6–7; (●) 4–5; (●) 2–3; (●) = 1; (●) no data. Three studies were conducted that focused on samples taken from the Antarctic land mass (data not shown). The exact number of studies conducted in each country is shown in Table S1 in Supplementary Material.



**Figure 3.** Taxa targeted in 167 studies reviewed. ‘Functional genes’ refers to the amplification of single genes encoding for functional processes (e.g. nitrogen fixation), ‘Metagenomes’ refers to gene data obtained by shotgun metagenomics.

#### Organisms studied

Studies of prokaryote communities (largely of Bacteria and Cyanobacteria) were most common (e.g. Costa et al. 2015; Dong et al. 2015; Pal et al. 2015), followed by investigations of fish (e.g. Jerde et al. 2013; Takahara & Minimoto 2013; Janosik & Johnston 2015; Fig. 3). Ten studies focused on a broad range of eukaryotic biodiversity (e.g. Pawlowski et al. 2011; Baldwin et al. 2013), whereas only three of the studies targeted both prokaryotic and eukaryotic organisms (Xiao et al. 2014; Young et al. 2014; Kowalik et al. 2015). Just over 10% of the studies investigated fungi (e.g. Rao et al. 2012; Lazarus & James 2015). Research on micro-eukaryotes was relatively common, with multiple studies being conducted on the foraminifera, as well as diatoms and a number of other protist taxa (e.g. Bradford et al. 2013; Lejzerowicz et al. 2014; Zimmermann et al. 2015). Various invertebrate taxa were investigated in 18 studies (e.g. Bienert et al. 2012; Yu et al. 2012; Deiner et al. 2015). Plants were investigated in just eight studies (e.g. Parducci et al. 2013; Pansu et al. 2015b), mammals in five (e.g. Nichols & Spong 2014; van Bleijswijk et al. 2014) and birds (in combination with other organisms) in just one study (Thomsen et al. 2012b).

#### Studies conducted on invasive organisms

A total of 26 studies were identified as using environmental DNA to investigate the presence of invasive organisms (See Table S2 in Supplementary Material). With the exception of one study investigating a fungal pathogen (*Guignardia* sp.) of citrus (Hu et al. 2014), all investigations were undertaken in aquatic environments. A majority of these studies investigated the presence of exotic fish such as common carp (*Cyprinus carpio*; Eichmiller et al. 2014; Takahara et al. 2015), silver and bighead carp (*Hypophthalmichthys molitrix* and *nobilis*; Jerde et al. 2013; Farrington et al. 2015), African jewelfish (*Hemichromis lifalili*; Moyer et al. 2014), bluegill sunfish (*Lepomis macrochirus*; Takahara & Minimoto 2013) and zebrafish (*Danio rerio*; Collins et al. 2013).

Multiple studies highlight the advantages of sampling environmental DNA for improving occurrence and detection

estimates for invasive organisms. For example, Hunter et al. (2015) estimated detection probabilities in excess of 91% for Burmese python (*Python bivittatus*) with positive results reported outside of the leading northern edge of the organism’s known range. The analysis of environmental DNA is thought to offer substantial cost benefits over traditional methods for invasive organism detection. For example, Jerde et al. (2011) achieved a positive detection result for silver carp in just a few hours from the analysis of environmental DNA, compared to 93 days of effort the authors predicted would be required to obtain the same result by electrofishing. The authors were also able to confirm that the invasive carp were closer to the invasion of upstream lake systems than had previously been detected by traditional methods. Others have similarly reported that the analysis of environmental DNA is a superior approach compared to traditional survey methods for assessing the presence of invasive aquatic animals such as the bluegill sunfish (Takahara & Minimoto 2013) and the Chinese giant salamander (*Andrias davidianus*; Fukumoto et al. 2015).

#### Studies conducted on rare organisms

A total of 8% of studies were identified as using environmental DNA to investigate rare or threatened organisms (Table S2). Most of these studies investigated the presence of fish, including Chinook salmon (*Oncorhynchus tshawytscha*; Laramie et al. 2015), brook and bull trout (*Salvelinus fontinalis* and *S. confluentus*; Wilcox et al. 2013), and slackwater darter (*Etheostoma boschungii*; Janosik & Johnston 2015), as well as amphibians such as the great crested newt (*Triturus cristatus*; Rees et al. 2014; Biggs et al. 2015), Idaho salamander (*Dicamptodon aterrimus*), Rocky Mountain tailed frog (*Ascaphus montanus*; Goldberg et al. 2011; Pilliod et al. 2013) and eastern hellbender (*Cryptobranchus alleganiensis alleganiensis*; Olson et al. 2012; Spear et al. 2015). With the exception of the bull trout and slackwater darter, which are respectively classified as vulnerable and endangered (IUCN Red List version 2.3), these organisms may be locally rare, but are otherwise classified as low risk in terms of their conservation status. All of the studies that used environmental DNA to detect rare organisms took place in aquatic habitats.

Although the analysis of DNA from environmental samples is not in widespread use for the detection of rare organisms, authors point to multiple potential advantages for its use, such as non-invasive and greatly reduced sampling efforts, and in some cases confirming meaningful relationships between organism density and DNA amplification (Pilliod et al. 2013). In one example, eastern hellbenders (*Cryptobranchus a. alleganiensis*) were successfully detected using environmental DNA at densities approaching the lowest reported natural population densities (Olson et al. 2012). In many cases, the likelihood of false positive detection is reported to be low; potential biases for the incomplete detection of DNA can be quantified by formal estimation of DNA detection probabilities under occupancy modelling frameworks, as used by Moyer et al. (2014) and more recently by Furlan et al. (2016).

#### Study environments

To date, studies on environmental DNA have most commonly been conducted on samples collected from freshwater environments, with research on water column samples (Thomsen et al. 2012b; Bradford et al. 2013; Mao et al. 2014) far outnumbering studies based on freshwater sediments (e.g. Pansu et al. 2015b) or biofilm samples (Callejas et al. 2011; Zimmermann et al. 2015; see Table S3 in Supplementary

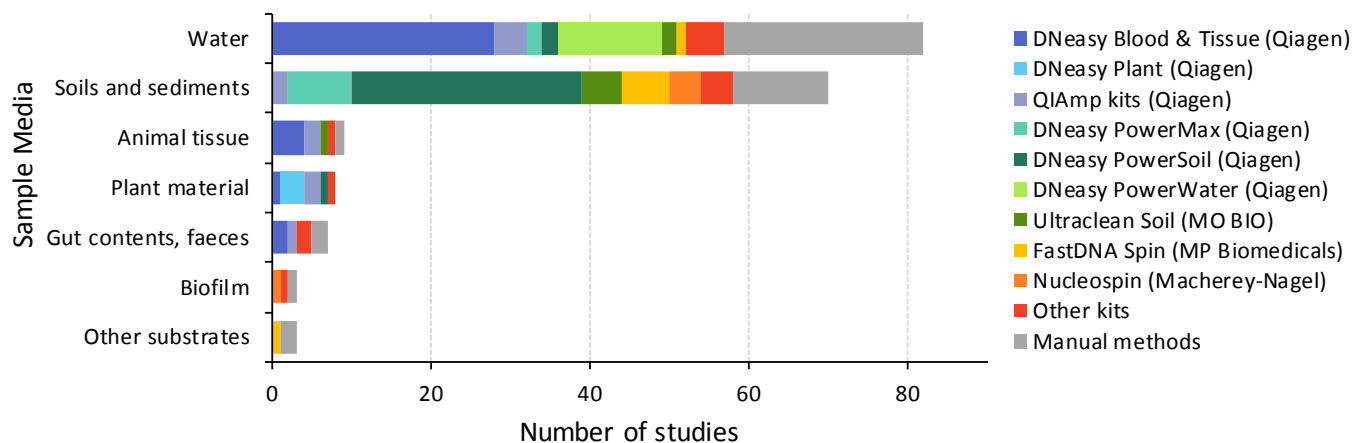
Material). These water column studies targeted a broad range of taxa including prokaryotes (e.g. Barberan & Casamayor 2014; Mao et al. 2014), fish (e.g. Jerde et al. 2013; Takahara & Minimoto 2013), amphibians (e.g. Goldberg et al. 2011; Olson et al. 2012; Rees et al. 2014), and invertebrates (e.g. Goldberg et al. 2013; Machler et al. 2014). In contrast, slightly more studies conducted in marine habitats focused on sediments (e.g. Nagahama et al. 2011; Pawlowski et al. 2011; Dong et al. 2015) compared to water samples (e.g. Stoeck et al. 2010; Thomsen et al. 2012a; Pochon et al. 2013). Micro-eukaryotes and fungi were the most common targets of marine sediment studies (e.g. Singh et al. 2012; Lejzerowicz et al. 2014), while the marine water column studies variously targeted prokaryotes (e.g. Cottrell & Kirchman 2012) or eukaryote organisms including fish (e.g. Thomsen et al. 2012a) and invertebrates (e.g. Thomsen et al. 2012a; Pochon et al. 2013). Terrestrial samples – mainly soil – were studied almost as often as marine samples, with the most common focus of these studies being the analysis of fungi (e.g. Teasdale et al. 2013; Lazarus & James 2015; Song et al. 2015) followed by prokaryotes (e.g. Lin et al. 2010; Kanokratana et al. 2011). Several of the soil-based DNA metabarcoding studies targeted earthworms (Bienert et al. 2012; Ficotela et al. 2015; Pansu et al. 2015a) or large vertebrates (Andersen et al. 2012). Gut and faecal material, collected in 5% of the studies, has been used to analyse the diets of herbivores (Hibert et al. 2013), carnivores (Boyer et al. 2015) and carrion feeders (Calvignac-Spencer et al. 2013). Several of the 167 DNA metabarcoding studies analysed pools of invertebrate specimens collected in malaise traps (Yu et al. 2012; Ji et al. 2013; Liu et al. 2013; Yang et al. 2014) or from soil samples (Yang et al. 2014), while others targeted invertebrate DNA extracted directly from soil (McGee & Eaton 2015) or aquatic habitats (Pochon et al. 2013; Cowart et al. 2015).

#### Summary of DNA extraction methods

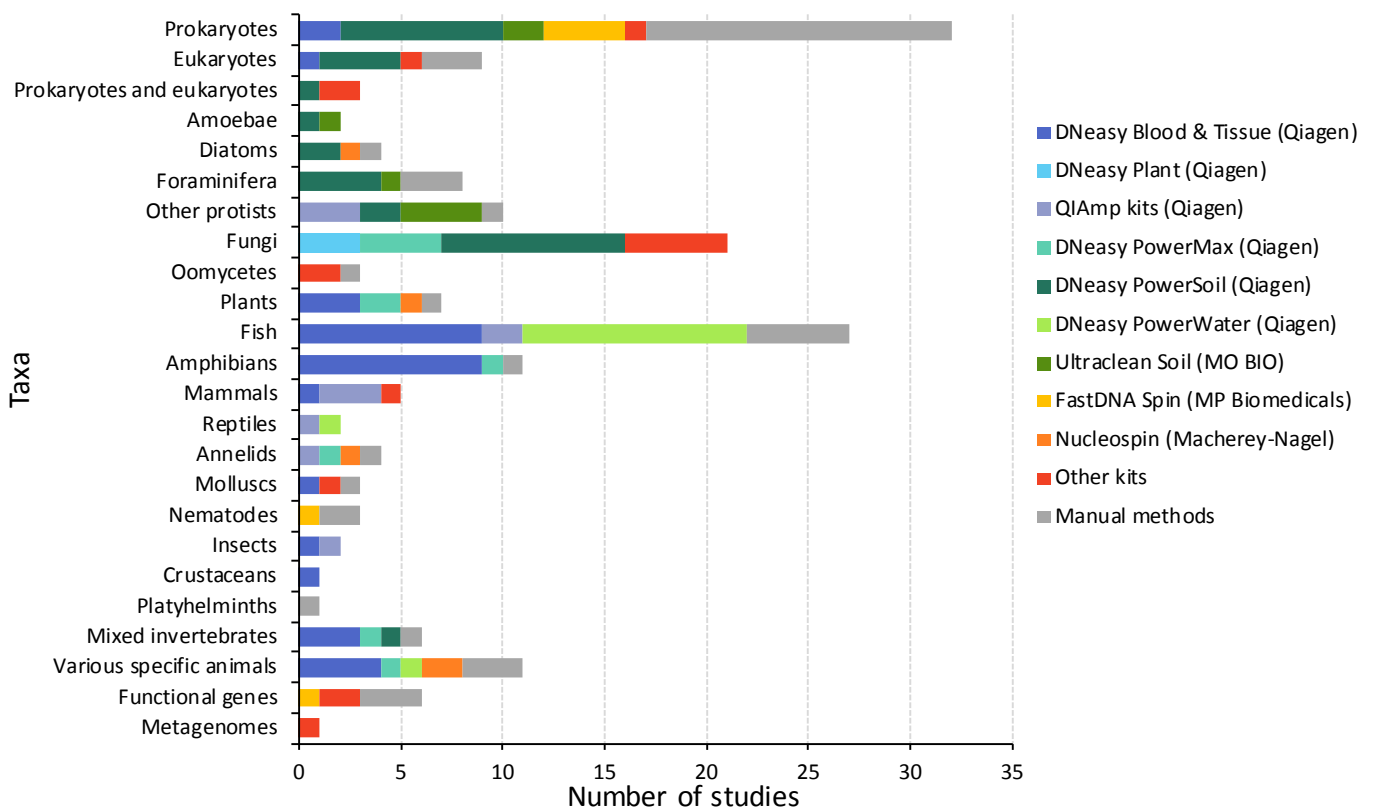
For the analysis of community composition from environmental DNA, the DNA must first be separated from the cellular material and the sample media (e.g. from soil particles) which can contain a wide variety of contaminants that may inhibit PCR. The choice of DNA extraction approach varied depending on the media from which the DNA was extracted, and the target organism under study (Fig. 4). PowerSoil<sup>®</sup> and PowerMax<sup>®</sup> Soil DNA isolation kits (now rebranded as DNeasy

PowerSoil and DNeasy PowerMax by Qiagen, Carlsbad, USA), as recommended by the Earth Microbiome Project (EMP; www.earthmicrobiome.org), were used in almost half of studies examining DNA from soil or sediment material (e.g. Pawlowski et al. 2011; Andersen et al. 2012; Teasdale et al. 2013; Lejzerowicz et al. 2014; Dong et al. 2015). In contrast, DNeasy Blood and Tissue kits (Qiagen) were commonly used to isolate and extract DNA from both marine and freshwater (e.g. Thomsen et al. 2012a; Goldberg et al. 2013; Spear et al. 2015; Takahara et al. 2015). DNeasy kits are available in a 96 well extraction format, making them more attractive for the analysis of a large number of samples. DNeasy PowerWater<sup>®</sup> DNA isolation kits (Qiagen) were used in ~20% of freshwater studies (e.g. Olson et al. 2012; Jerde et al. 2013; Deiner et al. 2015; Janosik & Johnston 2015). Although the number of plant-based studies captured by our analysis of the literature was low, DNeasy Plant Mini kits and QIAmp DNA Investigator kits (Qiagen) were used in multiple cases for the extraction of non-plant DNA (e.g. fungal DNA) from plant material (e.g. Bazzicalupo et al. 2013; Nichols & Spong 2014). A wide variety of manual (i.e. non-commercialised) methods were used across all types of sample media.

Some of the biases in the choice of methods to extract DNA from different sample media likely reflect differences in the organisms targeted from these different media (Fig. 5). Studies targeting macro-organisms, including fish and amphibians, most commonly extracted DNA using DNeasy Blood and Tissue kits (e.g. Takahara & Minimoto 2013; Laramie et al. 2015; Spear et al. 2015). In contrast, DNeasy PowerSoil<sup>®</sup> and PowerMax<sup>®</sup> kits were most commonly used in studies targeting the DNA of both prokaryotic and eukaryotic micro-organisms (e.g. Pawlowski et al. 2011; Lejzerowicz et al. 2014; Costa et al. 2015; Dong et al. 2015). Studies using DNeasy PowerWater<sup>®</sup> kits targeted DNA from fish (e.g. Jerde et al. 2013; Keskin 2014), but also amphibians (e.g. Olson et al. 2012) and an aquatic reptile (Burmese python; Hunter et al. 2015). Although DNeasy PowerSoil<sup>®</sup> kits were commonly used for prokaryotes (e.g. Griffin et al. 2013; Costa et al. 2015; Dong et al. 2015), a diverse range of manual methods (i.e. not using commercial kits) were used in the majority of studies (e.g. Callejas et al. 2011; Barberan & Casamayor 2014; Mao et al. 2014), indicating that the isolation of prokaryote DNA remains particularly poorly standardised. FastDNA Spin kits (MP Biomedical) have been used for the isolation of bacterial



**Figure 4.** DNA extraction kits and methods used for different sample media. Details of ‘other kits’ used for DNA isolation are provided in Table S4 in Supplementary Material.



**Figure 5.** DNA extraction kits and methods used in studies focused on the analysis of environmental DNA originating from different taxa.

DNA in a number of studies, but only twice since 2014 (Merlin et al. 2014; Wasaki et al. 2015).

#### Summary of gene regions used in prior analyses of environmental DNA

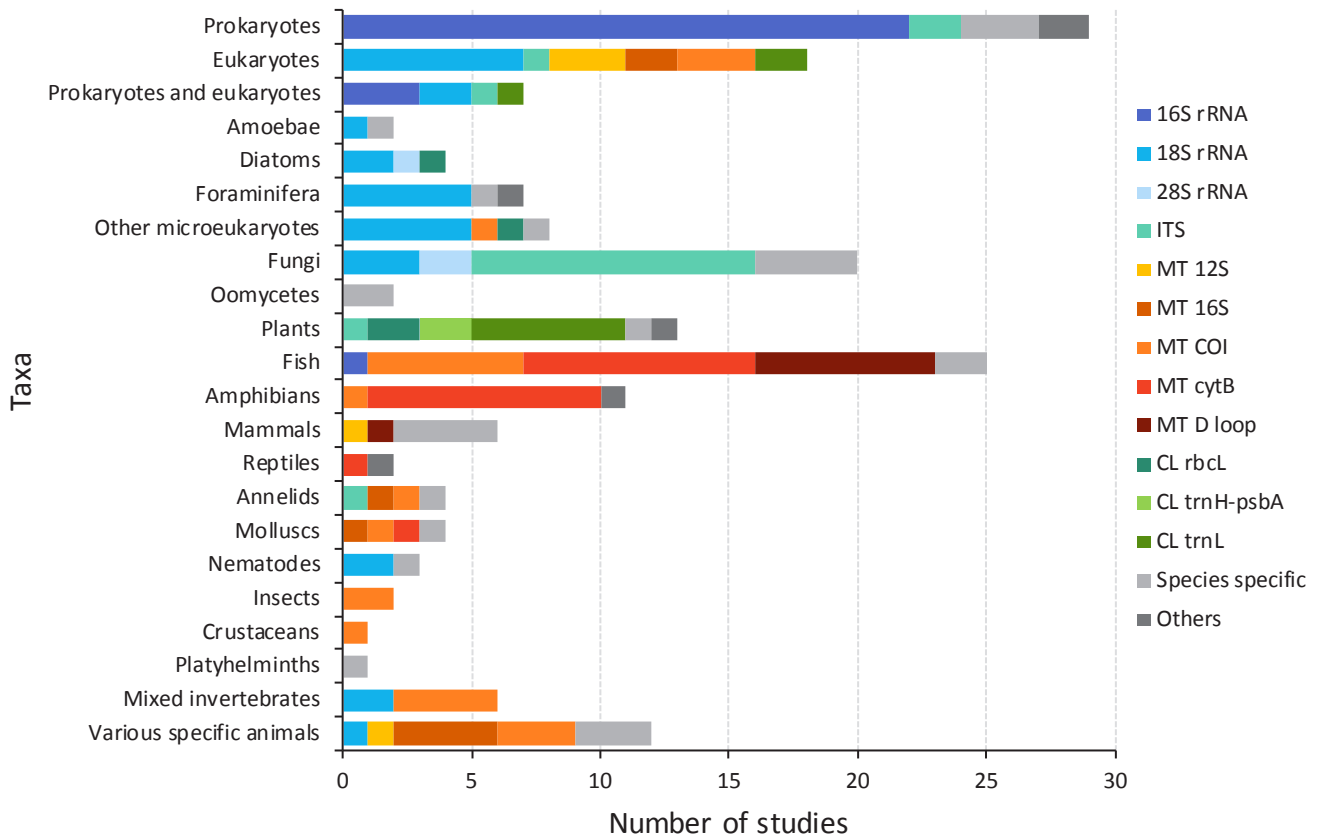
Polymerase chain reaction (PCR) is a method used to amplify a single or a few copies of a target piece of DNA, potentially to generate thousands to millions of copies. To initiate a PCR, short strands of DNA, or primers, are required. These short DNA sequences are chosen to bind either side of the gene region to be amplified, and therefore the choice of DNA primers dictates which DNA region is multiplied during PCR. As expected, the gene regions selected for amplification varied according to the target taxa (Fig. 6). Regions of the prokaryote 16S rRNA gene were used in almost all studies targeting the DNA of Bacteria and Archaea (e.g. Barberan & Casamayor 2014; Mao et al. 2014; Dong et al. 2015), unless primers were designed to detect individual species and genera (e.g. *Aeromonas hydrophila* strain Vah; Griffin et al. 2013). Five studies also chose to use shotgun metagenomic methods, which do not require the use of DNA primers, to identify prokaryote DNA from metagenomes (not shown in Fig. 6; Delmont et al. 2011; Nakai et al. 2011; Inskeep et al. 2013; Costa et al. 2015; Owen et al. 2015).

Investigations of amoebae, diatoms, and other micro-eukaryotes often used regions of the 18S rRNA gene as the DNA barcode (e.g. Pawlowski et al. 2011; Bradford et al. 2013; Zimmermann et al. 2015), which was also targeted in nematode-focused (e.g. Bhadury & Austen 2010; Kanzaki et al. 2012) and some arthropod-focused studies (e.g. Pochon et al. 2013; Yang et al. 2014). The 18S rRNA gene region was similarly amplified in several fungal analyses (e.g. Nagahama et al. 2011; Lazarus & James 2015), but the internal transcribed

spacer (ITS) region 1 (ITS1, located between 18S and 5.8S rRNA genes) was the more commonly targeted (e.g. Bellemain et al. 2010; Singh et al. 2012; Bazzicalupo et al. 2013; Song et al. 2015), with more recent studies switching to the ITS2 region (between 5.8S and 28S rRNA genes; Bazzicalupo et al. 2013). A majority of studies of plant DNA were based on the *trnL* UAA intron within the chloroplast transfer RNA, or rRNA gene (e.g. Hibert et al. 2013; Parducci et al. 2013; Pedersen et al. 2013). The DNA of arthropods was most commonly targeted by amplification of mitochondrial cytochrome *c* oxidase subunit I (COI) DNA (e.g. Yu et al. 2012; Ji et al. 2013; Machler et al. 2014), whereas mitochondrial cytochrome *b* (*cyt-b*) gene regions were the most frequent target for the detection and identification of amphibians (e.g. Goldberg et al. 2011; Olson et al. 2012; Pilliod et al. 2013; Spear et al. 2015).

There is a lack of consensus on target gene regions for the analysis of fish communities, with similar numbers of studies using the *cyt-b* (e.g. Minamoto et al. 2012; Thomsen et al. 2012a; Takahara & Minamoto 2013), COI (e.g. Collins et al. 2013; Keskin 2014; Laramie et al. 2015), and mitochondrial D-loop regions (e.g. Jerde et al. 2013; Farrington et al. 2015; Turner et al. 2015). Recently published studies have focussed on characterising fish communities using mitochondrial 12S rRNA genes (Evans et al. 2016; Olds et al. 2016; Shaw et al. 2016; Valentini et al. 2016). In addition, primers targeting the mitochondrial 16S rRNA gene were used in a minority of studies on fish, either alone (Deagle et al. 2013), in combination with mitochondrial 12S rRNA gene targets (Shaw et al. 2016) or with 12S rRNA gene and *cyt-b* gene targets (Evans et al. 2016; Olds et al. 2016). The mitochondrial 16S rRNA gene region has similarly been used to detect and identify Coleoptera (Epp et al. 2012), mammals (Ficotela et al. 2015) and earthworms (Bienert et al. 2012).





**Figure 6.** Gene regions targeted in studies focused on different taxa (excluding microarray or metagenomic studies). rRNA indicates ribosomal RNA genes, MT indicates mitochondrial genes, and CL indicates chloroplast genes.

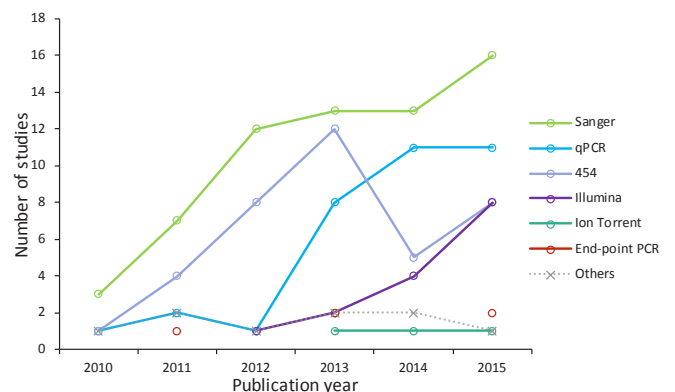
*Summary of PCR product purification approaches*

Following PCR, the reaction mixture must be ‘purified’ to remove remaining primers as well as PCR enzymes and salts. A variety of PCR purification approaches were used. The most commonly-adopted PCR purification approaches were the Agencourt AMPure XP system (e.g. Bazzicalupo et al. 2013; Pochon et al. 2015b; Song et al. 2015), the Qiagen MinElute PCR purification kit (e.g. Stoeck et al. 2010; Thomsen et al. 2012a; Calvignac-Spencer et al. 2013) and the Promega Wizard SV Gel and PCR Clean-Up system (e.g. Bhadury & Austen 2010; Callejas et al. 2011; Keskin 2014). The AMPure XP method was used only in high-throughput sequencing studies and the Promega Wizard system only in Sanger sequencing studies, whereas the Qiagen MinElute method was used for a range of applications.

*Summary of DNA sequence analysis methods used*

DNA sequencing is used to determine the order of nucleotides (A, C, G and T) within a DNA molecule and may be used to identify the genes of interest in a sample and potentially the organism from which the gene originated. Sanger sequencing remained the most common sequencing approach from 2010 through to 2015 (e.g. Bhadury & Austen 2010; Jerde et al. 2011; Minamoto et al. 2012; Collins et al. 2013; Keskin 2014; Janosik & Johnston 2015), despite the increasing availability and performance of high-throughput sequencing technologies (Fig. 7). While not providing any detailed community-based information, qPCR was used with increasing frequency for DNA detection and quantification (e.g. Pilliod et al. 2013; Takahara & Minamoto 2013; Merlin et al. 2014; Moyer et al. 2014; Farrington et al. 2015; Laramie et al. 2015; Spear et

al. 2015). The 454 pyrosequencing platform was used with increasing frequency prior to 2013 but declined in popularity after this date (e.g. Stoeck et al. 2010; Pawlowski et al. 2011; Yu et al. 2012; Parducci et al. 2013; Pochon et al. 2013; Yang et al. 2014; Zimmermann et al. 2015), coinciding with a rise in the use of Illumina sequencing platforms for DNA analysis; however, the maximum number of studies using the latter system in any year was just seven (Costa et al. 2015; Deiner et al. 2015; Dong et al. 2015; Ficotela et al. 2015; Pansu et al. 2015a; Pochon et al. 2015b; Song et al. 2015). The Ion Torrent platform was used just three times (all since 2013) among the 167 studies reviewed (Deagle et al. 2013; Young et al. 2014; Zaiko et al. 2015).



**Figure 7.** DNA sequencing approaches and other protocols used for the analysis of DNA from environmental samples.

## Conclusions

Based on our analyses of 167 research papers:

(i) Studies of freshwater habitats (water, sediment or biofilm) were more frequent than studies of marine or terrestrial habitats (45% compared to 23% and 22% of studies, respectively). DNA was extracted from water samples in 46% of the studies, mainly for the purpose of detecting specific animals using qPCR or Sanger sequencing methods. Analyses of DNA extracted from soils and sediments (40% of studies) targeted a wider range of organisms including plants, fungi, micro-eukaryotes, and prokaryotes, typically using either Sanger sequencing or high-throughput sequencing systems.

(ii) DNA was most commonly extracted from soil and sediment samples using DNeasy PowerSoil<sup>®</sup> and PowerMax<sup>®</sup> kits (Qiagen). Extraction of DNA from water samples was most commonly achieved using DNeasy Blood & Tissue kits, followed by DNeasy PowerWater<sup>®</sup> kits. An assortment of other kit-based and manual, or non-commercialised methods were also used for DNA extractions from these sample media. Several different methods were used for DNA extractions from other sample media, although DNeasy Blood & Tissue kits and DNeasy Plant kits were respectively used slightly more frequently than other approaches for extractions targeting animal or plant DNA.

(iii) Consistent gene regions were targeted in most studies of prokaryotes (16S rRNA gene), amphibians (mitochondrial cytochrome b, or *cyt-b* gene) and plants (chloroplast *trnL* intron); arthropods were usually analysed using primers targeting the mitochondrial COI gene. More studies of fungal communities targeted the internal transcribed spacer ITS1 region than the 18S rRNA gene. However, the 18S rRNA region was used in a variety of studies of micro-eukaryote and animal taxa. There is little consistency in primer targets for the analysis of fish DNA, with equal numbers of studies targeting mitochondrial COI, *cyt-b* and D-loop regions.

(iv) Our selective review identified only a limited number of studies using next-generation sequencing methods (seven Illumina-based studies, eight 454-based studies, and just one study using Ion Torrent out of 167 studies reviewed in total), with Sanger-based DNA sequencing remaining the most commonly used approach prior to August 2015.

## Identification of standard procedures for DNA extraction, storage, amplification and sequencing

### Sample storage prior to DNA extraction

It is generally best to keep samples cool and to carry out DNA extractions as soon as possible after samples are collected, to limit potential DNA degradation due to the material being removed from its original context. However, it is often impractical to carry out DNA extractions immediately after sample collection, and it may be desirable to retain samples for future analyses. A variety of methods are used to store samples, including cooling to 4, -20, or -80°C (depending on available facilities), drying, freeze-drying, or addition of preservative buffers, such as DMSO-EDTA. The suitability and feasibility of these approaches will depend on the taxa to be investigated, sample media, and the duration of storage. The simplest approach is to use cooling. Lauber et al. (2010) concluded that storage of microbiome samples at temperatures ranging from -20°C to -80°C for up to 14 days

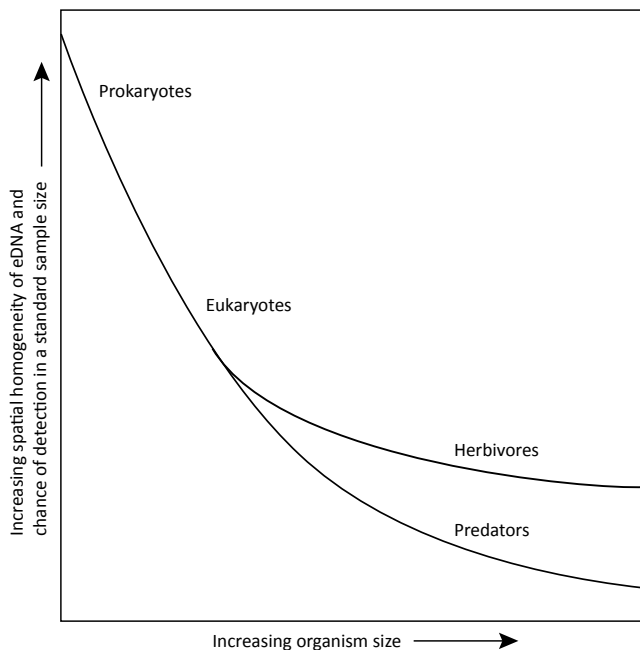
had little impact on the resulting inferences; similar findings have been reported in other studies (e.g. Carroll et al. 2012). Bainard et al. (2010) concluded that drying methods had an adverse impact on recovery of arbuscular fungal DNA, suggesting that samples should instead be frozen to prevent DNA degradation. We recommend the immediate storage of samples at ~4°C following collection. For field studies, this is most easily achieved by the transfer of samples to cool boxes containing ice or other frozen material. Cooled samples should be transferred to -20°C freezers within 48 h for short-term storage (e.g. weeks to months) and to -80°C freezers for longer term storage (e.g. months to years). In situations where the cooling of samples in the field is difficult, such as when sampling in remote locations, ambient temperature sample storage may be considered. Proprietary (e.g. RNAlater<sup>®</sup> and DNagard<sup>™</sup>) and non-proprietary (e.g. DMSO-EDTA) solutions can be used to stabilise nucleic acids for this purpose. A summary of these, and other approaches for storing sample DNA for molecular analyses is provided by Nagy (2010), where details on individual methods are described in terms of their adequacy for field work, optimal storage period (i.e. short-, medium-, or long-term), ease of use, health hazards and associated costs. RNA degradation by RNase enzymes is a significant concern during sampling handling and storage (Chomczynski 1992). Therefore, additional or different sample storage procedures may be required for studies in which the analysis of sample RNA is desirable. Readers are pointed to Kasahara et al. (2006) for more information on appropriate methods for sample storage for RNA preservation.

### Sample pre-processing

Before extracting DNA, pre-processing steps are often required, in order to (i) reduce the number of samples from which DNA is to be extracted, e.g. by combining multiple samples together, and (ii) reduce sample volumes required for processing by concentrating the target biomass required for DNA extraction. Procedures to concentrate the target biomass normally require its separation from the sample media and may simultaneously reduce concentrations of PCR-inhibiting substances in samples. Here, we review a number of common approaches used to process samples before and during DNA extraction.

#### *Reducing sample numbers for DNA extraction*

Natural biological communities are frequently complex, with heterogeneity observed over a wide variety of spatial scales (Ranjard et al. 2003). If the extraction of DNA from large sample numbers is not feasible, then large numbers of samples may be pooled together and mixed before one or more smaller subsamples are removed for DNA extraction (e.g. multiple soil cores of >10 g may be mixed together before a subsample, perhaps of <1 g, is collected for DNA extraction). However, wherever possible sample pooling should be avoided in favour of the analysis of more small subsamples. This is particularly the case in studies where the aim is to identify rarer members of the community, or to generate estimates of species richness. Sample pooling can mask a significant proportion of the detectable community, particularly organisms distributed with greater spatial heterogeneity (Manter et al. 2010). This masking occurs as locally dominant but spatially rare taxa become rare in the final pooled sample, rendering them undetectable. For example, Manter et al (2010) detected 67 more fungal taxa and 115 more bacterial taxa from the analysis of unpooled as compared to pooled soil samples collected from three plots (nine subsamples collected per plot). However, sample pooling



**Figure 8.** Schematic of the relationship between the size of an organism and the spatial homogeneity of its DNA in the environment, which is positively related to the likelihood of detecting that organism using DNA metabarcoding. Broadly, the spatial distribution of larger organisms such as multicellular eukaryotes, particularly predators, is less homogenous than that of smaller organisms such as prokaryotes, single-celled eukaryotes and herbivores. Consequently, the detection of larger organisms tends to require biomass concentration and/or bigger samples than are necessary for the detection of smaller organisms.

and homogenisation may be applicable where the main aims of a study are to determine the presence of specific taxa in a community, particularly if these organisms are spatially heterogeneous, such as is more likely to be the case for larger organisms and predators (Fig. 8). We recommend the analysis of as many individual samples as is possible, without pooling, particularly when estimates of taxon richness and diversity are key study aims.

#### *Reducing sample volumes for DNA extraction*

Biomass and DNA concentration may be appropriate for organisms that have high spatial heterogeneity (i.e. are at low population density relative to the sampled area) and have a correspondingly low chance of detection in a sample (Fig. 8). For example, predatory mites are typically less abundant in soil or leaf litter than herbivorous mites, and therefore may require concentration to achieve a sample representative of the community. Once concentrated samples have been obtained, DNA can be extracted from these using standard protocols.

The method of biomass concentration depends on the organisms or material in question. Where the aim is to concentrate environmental DNA (rather than cellular biomass, live or dead) from arbitrarily large volumes of terrestrial material, such as soil or leaf litter, DNA can be recovered using a saturated phosphate buffer solution (Bienert et al. 2012). After mixing the sample material with the buffer solution, the DNA is washed from the sample material, concentrated and extracted from one or more subsamples of the buffer solution using a conventional extraction kit (the Macherey-

Nagel Nucleospin Soil kit is recommended for this purpose; Taberlet et al. 2012). This simple protocol allows the efficient and economic recovery of homogenised pools of DNA from considerably larger sample volumes than can be processed using conventional extraction approaches. Such procedures inevitably bias against the extraction of DNA from more adherent organisms (e.g. biofilm dwelling bacteria may be underrepresented compared to their free-living counterparts; Garrett et al. 2008). Nevertheless, this approach is particularly relevant for the detection of large and/or sparsely distributed organisms and removes the need to process multiple small samples for DNA extraction and analysis. Where the aim is to concentrate cellular biomass, rather than extracellular DNA, methods of biomass concentration vary depending on the target organisms for analysis. Examples of biomass concentration methods for different taxa include the following.

**Trapping and collecting live organisms.** The composition of soil invertebrates is commonly assessed by first extracting the organisms from the soil using a combination of approaches including pitfall traps (Drummond et al. 2015) and modified Tullgren and Baermann funnels (Bao et al. 2012), each of which rely on the capture of organisms during active movement. Similarly, aerial insects may be captured using malaise traps (Yu et al. 2012). Aquatic invertebrates are commonly concentrated during capture using kick-net samplers (Machler et al. 2014). DNA is then extracted from the concentrated biomass.

**Flotation.** Flotation is used for the extraction of microarthropods from sandy soils, particularly for less-active taxa such as podurid Collembola (Geurs et al. 1991). This method is largely based on the difference in density of the animals and the flotation fluid (commonly heptane; Geurs et al. 1991).

**Centrifugation.** Centrifugation aids the collection of buoyant biomass (e.g. nematode cysts; Bellvert et al. 2008) from complex media such as soil. Density gradient centrifugation may also be used to separate small organisms such as bacterial cells from complex media, prior to downstream molecular processing (Dichosa et al. 2014).

**Filtration.** The biomass of microorganisms in both water (Lear et al. 2014) and air (DeLeon-Rodriguez et al. 2012) are commonly concentrated by filtration prior to extraction.

**Manual sorting.** Biological material such as roots, leaf fragments, and invertebrates may be manually picked from substrates such as soil, to increase the abundance of target biomass. Unwanted materials may be removed in a similar way, for example by sieving to exclude larger particles or inorganic material (e.g. stones) from the extraction process.

#### **DNA extraction protocols**

##### *General recommendations for the extraction of DNA from environmental samples*

As highlighted by our literature review, different protocols are used for the extraction of DNA depending on the sample medium. A variety of pre-treatment options are also used, depending on the sample media and volume (Table 1). Based on these observations, we make some broad recommendations for standard procedures for the extraction of DNA from different sample media. DNeasy PowerSoil<sup>®</sup> DNA Isolation kits, with a capacity of up to 0.25 g, are widely used in studies of soil, sediment, faecal material and leaf litter, and have been recommended for use by a number of international standards consortia following comparisons with a range of other methods (Gilbert et al. 2014). Consequently, we recommend the PowerSoil<sup>®</sup> DNA Isolation kit for extractions from the smallest samples (i.e.  $\leq 0.25$  g). The larger sample volume

**Table 1.** Suggested pre-treatment and DNA extraction protocols for a range of sample media.

Extraction media	Pre-treatment options	Recommended extraction kit and 'wet' sample volumes
<b>Soils/peats/sediments</b>	<p>If sampling from <b>large volumes of soil</b>, sample material may first be mixed. This is most commonly done by 'hand' or using kitchen blenders to homogenise soil slurries (Tien et al. 1999). Alternatively, environmental DNA may be recovered from large soil volumes using the <b>saturated phosphate buffer method</b> (Bienert et al. 2012).</p> <p><b>Allophanic soils</b> such as andisols bind environmental DNA very efficiently and may require a two-step pre-treatment (Huang et al. 2016).</p> <p><b>Other difficult samples</b> can benefit from the addition of lytic enzymes such as proteinase K, as in Bulat et al. (1998).</p> <p><b>If desiccated</b>, add ddH<sub>2</sub>O and rotate overnight to rehydrate. Otherwise, treat as soil (above).</p> <p>Centrifugation may be required to concentrate <b>watery soil and sediments</b>.</p>	<p>&lt;0.25 g: DNeasy PowerSoil<sup>®</sup>;  0.25–2.5 g: DNeasy PowerSoil<sup>®</sup>  RNA with DNA Elution Accessory;  2.5–7.5 g: DNeasy PowerMax<sup>®</sup>;  &gt;7.5 g: phosphate buffer<sup>1</sup>.</p>
<b>Faeces/ejecta</b>	<p>If desiccated, add ddH<sub>2</sub>O and rotate overnight to rehydrate.</p> <p><b>Carnivore:</b> Follow procedure for 'bone' below.</p> <p><b>Insectivore:</b> Follow procedure for 'chitin' below.</p>	<p>&lt;0.25 g: DNeasy PowerSoil<sup>®</sup>;  0.25–2.5 g: DNeasy PowerSoil<sup>®</sup>  RNA with DNA Elution Accessory;  2.5–7.5 g: DNeasy PowerMax<sup>®</sup>.</p> <p>DNeasy Blood and Tissue kit.</p> <p>DNeasy Blood and Tissue kit.</p>
<b>Leaf litter</b>	<p>For the analysis of prokaryotes and invertebrates, leaf litter is typically treated as soil (see above). For the analysis of plant DNA, leaf litter is typically treated as plant tissue (see below).</p>	<p>&lt;0.25 g: DNeasy PowerSoil<sup>®</sup>;  0.25–2.5 g: DNeasy PowerSoil<sup>®</sup>  RNA with DNA Elution Accessory;  2.5–7.5 g: DNeasy PowerMax<sup>®</sup>;  &gt;7.5 g: phosphate buffer<sup>1</sup>.</p>
<b>Water/ice</b>	<p>Filtration (normally at 0.22 µm). Enclosed cartridge filters may be used for improved sample storage following filtration (Urakawa et al. 2010).</p>	<p>DNeasy PowerWater<sup>®2</sup>.</p>
<b>Concentrated animal tissues</b>	<p><b>Chitin or keratin</b> digestion buffer (Campos &amp; Gilbert 2012).</p> <p><b>Bone</b> digestion buffer (Wood et al. 2016). Proceed to kit with ca. 1 mL of supernatant.</p> <p><b>Collagen (nematode cuticle):</b> concentrated live specimens<sup>3</sup> may be pre-treated with the digestion buffer of Zheng et al. (2003).</p>	<p>DNeasy Blood and Tissue kit (or DNeasy PowerSoil<sup>®</sup> kit for high inhibition samples, i.e. those with dark extracts).</p> <p>DNeasy Blood and Tissue kit.</p> <p>Supernatant purified using DNeasy Blood and Tissue kit.</p>
<b>Plant tissue</b>	<p><b>Roots</b> – grinding in bead mill, followed by modified PowerBead beating steps from the DNeasy kit.</p> <p><b>Leaves</b> – There seems to be no standard approach. Extraction is determined on a case-by-case basis.</p> <p><b>Pollen</b> – Grinding in liquid nitrogen or heat treatment (95°C, 10 min) to lyse cells and release DNA.</p> <p><b>Wood</b> – For a small number of large samples, drill powder into lysis buffer. For many small fragments grind in liquid nitrogen.</p>	<p>&lt;0.25 g: DNeasy PowerSoil<sup>®</sup>;  0.25–2.5 g: DNeasy PowerSoil<sup>®</sup>  RNA with DNA Elution Accessory;  2.5–7.5 g: DNeasy PowerMax<sup>®</sup>.</p>

<sup>1</sup> The phosphate buffer method is described by Taberlet et al. (2012) and Zinger et al. (2016).

<sup>2</sup> Where animal or plant derived DNA is the principal target, DNeasy Blood and Tissue or DNeasy Plant Tissue kits should be used instead.

<sup>3</sup> Wire basket method of Whitehead and Hemming (1965).

accommodated by the DNeasy PowerSoil<sup>®</sup> RNA Isolation kit means that samples of up to 2.5 g can be processed, when used in combination with an RNeasy DNA Elution Accessory kit to co-isolate DNA from the sample material. For even larger samples, the DNeasy PowerMax<sup>®</sup> Soil DNA Isolation kit, with a capacity of up to 5–10 g, is recommended. Terrestrial samples that exceed the capacity of the PowerMax<sup>®</sup> kit (>5–10 g) can be processed using the saturated phosphate buffer method (Bienert et al. 2012). A variety of methods, including PowerSoil<sup>®</sup> kits, have been used for DNA extraction from gut contents, faeces and plant tissues and yield high concentrations of quality DNA (Dineen et al. 2010; Wagner Mackenzie et al. 2015). Consequently, we also suggest the use of PowerSoil<sup>®</sup> or PowerMax<sup>®</sup> kits for these sample types, for consistency across different analyses. We recommend the use of PowerWater<sup>®</sup> kits for the recovery of DNA from microbial communities in both marine and freshwater, since this approach is widely used and is most similar to the PowerSoil<sup>®</sup> approach recommended for a variety of terrestrial media. However, where animal or plant-derived DNA is a main target of extractions from water, DNeasy Blood & Tissue and DNeasy PowerSoil<sup>®</sup> extraction approaches are suggested, respectively, allowing for better comparisons with sequence data collected directly from biological tissue. The preferential use of these and similar kits reduces the lab-to-lab variation that can arise from the use of non-commercial (i.e. non kit-based) methods, allowing access to comparable data from a wider group of users and creating consistency across studies carried out around the world. Alternative methods should nevertheless be used, where necessary, to maximise the quantity and quality of DNA extracted from difficult samples (e.g. soils containing elevated concentrations of heavy metals or humic organic matter).

#### *Extraction of DNA from soil, sediment and leaf litter*

##### General considerations

Early soil biodiversity studies relied on cell extractions prior to DNA isolation, followed by the removal of humic material by means of chromatography (Faegri et al. 1977; Torsvik 1980). These so called ‘indirect methods’ laid the groundwork for modern amplicon sequencing and metagenomics (Lane et al. 1985; Pace et al. 1986) and opened a window into the poorly-explored biodiversity of soils (Torsvik et al. 1990; Deagle et al. 2009). Subsequent method development, principally by Ogram et al. (1987), saw a shift towards ‘direct methods’ of extraction, which are the current standard. These methodologies frequently rely on the physical or chemical lysis of cellular material present in the sample media, combined with column-based DNA purification. The direct extraction of DNA from the original sample medium can result in higher (greater than an order of magnitude) DNA yields than indirect methods while retaining the molecular size of the DNA fragments within a range deemed suitable for DNA sequencing analysis (Miller et al. 1999; Miller 2001). However, DNA from other sources (e.g. plant debris) is frequently co-extracted in high concentration. Many comparisons of DNA extraction methods for soil, sediment and leaf litter samples have been published (Frostegard et al. 1999; Miller et al. 1999; Martin-Laurent et al. 2001; Miller 2001; Dineen et al. 2010; Mahmoudi et al. 2011) including for the recovery of ancient DNA (Haile 2012), each highlighting differences among methods in the community composition detected.

A wide variety of sample pre-processing steps are recommended before the extraction of DNA from soil and similar media. The selective removal of non-target material

such as stones, leaf litter or coarse root material, or alternatively the removal of sample material by size (i.e. by sample sieving) is common. As a minimum, the removal of larger inanimate material is advised (e.g. stones) to maximise the yields of DNA extracted. However, the details of sample processing approaches will vary depending on the research question as well as the target organism(s). For example, coarse root material would frequently be removed in studies where plant community DNA, or the DNA of plant endosymbionts and pathogens, is not under investigation.

DNA extraction is particularly difficult from samples containing high concentrations of clay and humic material. DNA binds strongly to clay particles (Frostegard et al. 1999; Cai et al. 2006) preventing the isolation of DNA into the extraction supernatant. Humic material has a similar size and charge to DNA, resulting in co-purification, as may be evidenced by the brown colour of some DNA extracts. The presence of humic material in DNA extracts inhibits the activity of some enzymes including DNA polymerases (Dong et al. 2006). Additionally, the co-extraction of humic material may interfere with DNA quantification by spectrophotometry, since both DNA and humic material exhibit optimal absorbance at both 230 and 260 nm. Fluorometric methods such as Qubit (ThermoFisher Scientific) are less affected by humic material, tending to provide more accurate estimates of DNA concentration in soil extracts. A number of commercial kits, including PowerSoil<sup>®</sup> DNA Isolation kits, are designed to remove PCR inhibitors from soil and similar material, including recalcitrant humic substances. In the case of DNeasy PowerMax<sup>®</sup> kits, addition of phenol to the extraction column may further improve the DNA recovery from clay-rich samples processed using this method (e.g. in step 1 of the standard protocol, add 10 ml PowerBead Solution with 5 ml phenol (phenol:chloroform:isoamyl alcohol pH 7–8); Charlotte Jordans, Geneworks Pty Ltd., pers. comm.). Alternative strategies for the removal of humic substances from soil DNA include the use of aluminium sulphate (Dong et al. 2006) and Sephadex columns (Tsai & Olson 1992b).

##### Extraction requirements and recommendations

DNeasy PowerSoil<sup>®</sup> kits are widely used and have been shown to be well optimised for DNA extraction from a variety of soils, including compost, sediment, clay and acidic soils (Roose-Amsaleg et al. 2001; Tedersoo et al. 2014). The same kits are also used for the extraction of DNA from leaf litter (Voříšková & Baldrian 2013). We recommend use of DNeasy PowerSoil<sup>®</sup> and PowerMax<sup>®</sup> kits for the extraction of DNA from soil, sediment and leaf litter (Table 1). Additional steps such as sample dilution or supplementing PCR mixtures with adjuvants such as BSA may be required for sample media containing elevated concentrations of PCR inhibitory substances.

#### *Extraction of DNA from faeces and egesta*

##### General considerations

Faeces or egesta (pellets regurgitated by predators, particularly insectivores) can be treated as the same sample media for purposes of DNA extraction (and so are hereafter referred to as faeces). There are three reasons why researchers may want to sequence DNA from such samples: (1) to determine to which species the faeces belong, for example, as a mechanism to identify ungulate species (Ramón-Laca et al. 2014); (2) because they want to determine the diet of the depositing animal (Vestheim & Jarman 2008; Shehzad et al. 2012; Boyer et al. 2013); or (3) because they are using the feeding

ecology of the depositor as a way of sampling the biodiversity in its environment (Kuch et al. 2002). This third approach envisions depositors as ‘environmental samplers’ and in effect faeces are ‘biodiversity capsules’ (Boyer et al. 2015) containing concentrated DNA from taxa consumed by the depositing species. While faecal DNA analysis can introduce sampling biases for biodiversity assessments, and requires an understanding of the ecology of the depositing species, it can be an excellent approach for detecting rare species in the environment. Each of these approaches has different assumptions and it is important that researchers are clear from the outset about what questions they aim to address. Care should be taken when sampling faeces, as many diseases and parasites that are transmissible to humans can exist in animal dung. Such precautions can involve simply wearing latex gloves when handling faeces and avoiding inhaling dust from dry faeces (wearing a dust mask). In exceptional circumstances (such as sampling bat guano from caves) it may be advisable to wear a respirator.

As the gastrointestinal tracts of most animals are excellent mixers, there may be no need to homogenise individual faecal samples. However, if small samples from multiple specimens are being combined for DNA extraction (e.g. invertebrate frass), they should be well mixed, using either a bead beater, or pestle and mortar. Larger volume samples may be mixed as a slurry with lysis buffer using, for example, stomacher laboratory paddle blenders (Abu Al-Soud & Rådström 1998).

If detecting the diet of the depositing species is of interest then the outside layer of each dung bolus represents a contamination risk because it has been in contact with the soil or other external substrate, and needs to be carefully removed (see the dung subsampling procedure of Wood and Wilmshurst (2011)). Obviously, this becomes more difficult for smaller specimens, and may be impossible for invertebrate frass, although UV light irradiation of the specimens may assist in reducing surface contamination in such instances (Cone & Fairfax 1993).

#### Extraction requirements and recommendations

Faecal DNA extraction is relatively straightforward, but differs for humic-rich (herbivore) and humic-poor (large carnivore) faeces (Table 1). In most cases, it is appropriate to extract DNA from faecal material using DNeasy PowerSoil<sup>®</sup> or PowerMax<sup>®</sup> kits (following rehydration if specimens are dry). However, we recommend the use of DNeasy Blood and Tissue kits for the analysis of carnivore faeces, allowing for better comparisons with sequence data collected directly from biological tissue. These sample materials may require decalcification and digestion steps prior to DNA extraction and isolation. Similarly, insectivore dung requires a chitin digestion step followed by use of a DNeasy Blood and Tissue kit, or DNeasy Powersoil<sup>®</sup> kit, if the sample media is also suspected to contain high concentrations of humic material (Table 1).

#### *Extraction of DNA from water and ice*

##### General considerations

A key issue with the extraction and amplification of DNA from water is low DNA concentration, which can require the filtration of many litres of water to obtain sufficient DNA from a single sample (Wilcox et al. 2016). Additionally, in lentic systems in particular, environmental DNA appears to be distributed somewhat patchily (Lear et al. 2014) which could yield false negative results for a given species of interest. For example, Furlan and Gleeson (2016) found that they needed to collect

up to 12 two litre water samples from each sampling station to detect redfin perch (*Perca fluviatilis*) that was known to be present in a lake. The mixing of running water will reduce spatial variability in the composition of environmental DNA; however, the downstream transport of DNA in lotic systems suggests that DNA sequence data may not only represent species present in the vicinity of sampling, but potentially from long distances upstream (Deiner et al. 2015).

#### Extraction requirements and recommendations

There are two distinct methods to collect and concentrate environmental DNA from water; precipitation followed by centrifugation (Turner et al. 2015), or filtration. Comparisons of the two methods on the same water samples have shown that higher concentrations of DNA are obtained by filtration methods (Deiner et al. 2015). However, there is a trade-off between filter pore size and the volume of water that can be filtered before the filter clogs. For this reason, it may be preferable to filter multiple small volumes of water through separate filters and later combine the DNA collected from each filter via ethanol precipitation (Santas et al. 2013). The optimal filter pore size is generally suggested to lie between 0.6 µm and 1.5 µm (Eichmiller et al. 2016; Minamoto et al. 2016). We suggest using 1.5 µm glass fibre filters (Type 934-AH) to filter water for assessments of vertebrate DNA but smaller filter sizes (e.g. 0.2 µm) are recommended for the capture of microbial biomass (Lear et al. 2014).

Once the water is collected, it should be filtered and stored as soon as possible since the quality and quantity of DNA present in water decreases rapidly (Thomsen et al. 2012a; Maruyama et al. 2014). For example, Maruyama et al. (2014) observed detectable concentrations of bluegill fish (*Lepomis macrochirus*) to decline by half in under 7 h at 20°C. Even when unfiltered water is frozen at -20°C, reductions of amplifiable DNA as great as ten-fold are reported in the literature (Cornelisen et al. 2012). Therefore, it is preferable to filter samples as soon as possible after sampling, and to preserve the filtered material at low temperature; studies have shown filters (and associated DNA) can be stored at -20°C for later DNA extraction without significant loss of amplifiable DNA (Gilpin et al. 2013). Other research suggests the fixation of filtered DNA with ethanol may allow sample material to be preserved at room temperature for many days (Minamoto et al. 2012; Thomsen et al. 2012a; Minamoto et al. 2016). However, since this approach is yet to gain widespread acceptance, its use for the preservation of rare target DNA is not recommended. For the collection of DNA or microbial biomass from small (i.e. less than one litre) volumes, on site-filtration is recommended using devices such as Sterivex filters (Wright et al. 2009) or customised portable filtration devices, such as those described by Yamanaka et al. (2016).

Comparative studies of DNA extraction effectiveness from filters suggest that the Qiagen DNeasy PowerWater<sup>®</sup> DNA extraction kit (a bead beating method) was less likely to extract PCR inhibitors along with the DNA compared with the Qiagen DNeasy Blood and Tissue kit, although the latter obtained higher concentrations of total DNA (Eichmiller et al. 2016). Thus, we recommend the use of the PowerWater<sup>®</sup> kit to extract DNA from water (Table 1), unless the goal of the study is to detect a specific animal (such as a rare or invasive fish species), in which case the DNeasy Blood and Tissue kit may give higher detection rates (Amberg et al. 2015).

### *Extraction of DNA from animal tissue*

#### General considerations

The extraordinary morphological and ecological diversity of animals presents some particular issues for DNA extraction. Many invertebrates (i.e. ecdysozoans, such as nematodes and arthropods) have hard, waterproof outer cuticles that may represent a barrier for the spread of DNA into the surrounding environment (Goldberg et al. 2013). This will be a particular problem for highly-sclerotised arthropods, such as weevils, which as a consequence may be under-represented in environmental DNA samples. Therefore, the DNA of animals, and invertebrates in particular, is often targeted directly from homogenised animal tissue rather than from true environmental DNA that has been shed or excreted into the environment.

The main types of animal tissue samples used for DNA metabarcoding include soil and litter invertebrates that have been separated from the substrate using a live extraction method such as Tullgren funnels, flotation (e.g. heptane flotation for mites) or centrifugation (e.g. sucrose centrifugation for nematodes) as well as invertebrates collected in traps. For example, litter-dwelling invertebrates can be collected in pitfall traps, flying insects in malaise traps and freshwater macroinvertebrates in nets using the kick sampling method. These samples reach the lab as bulk pools of specimens in preservative, such as ethanol or glycol, or as fresh or dry tissue. In contrast to many other forms of environmental DNA, these samples are likely to contain greater amounts of well-preserved DNA. However, DNA from large specimens may swamp that of small specimens, and it may be more difficult to retrieve the DNA from certain organisms that are particularly hard bodied and/or watertight. A way to minimise these potential biases is to ensure thorough mixing and digesting of the sample to maximise the chances of retrieving DNA from all specimens.

#### Extraction requirements and recommendations

Whole animal samples can be mixed by blending in a conventional food processor. For wet samples, this is usually followed by an incubation to evaporate residual water or ethanol (Hajibabaei et al. 2011). Samples in glycol need to be washed with ethanol prior to blending as glycol does not evaporate easily. Dry samples can be blended in a similar way, homogenised in a bead beater, or pulverised in liquid nitrogen. Elbrecht and Leese (2015) showed that using a commercial bead-beater (a Qiagen TissueLyser LT bead mill) led to the recovery of more taxa than when liquid nitrogen was used. The latter approach also provides more opportunities for contamination because samples are typically manually ground in open tubes rather than enclosed in sterile bead beating tubes.

Because animal tissue usually contains good quality DNA, extraction protocols developed for fresh tissue can be applied to these samples. The most widely used approach for DNA extraction from animal tissue is the DNeasy Blood & Tissue kit. Other kits appear to yield similar results as the DNeasy kit (Chen et al. 2010a). However, such comparisons have not been carried out in the context of biodiversity analyses based on DNA metabarcoding, and it remains uncertain whether particular extraction kits have the potential to selectively extract DNA from particular taxa over others in a mixed organism sample. Nonetheless, the DNeasy Blood & Tissue kit has been reported to provide good results from a very wide range of animal tissues and environmental media containing animal DNA and therefore is tentatively recommended for such applications (Table 1). Prior to processing with kits such as the DNeasy Blood & Tissue kit, specialist lysis buffers may be used, for example to extract DNA from nematodes

(Williams et al. 1992), or chitin or keratin-rich tissue, such as invertebrate exoskeletons or vertebrates (Campos & Gilbert 2012). Additional sample digestion steps are often required, depending on the organisms under study. To ensure standard DNA extraction approaches are used where possible, these additional processing steps should only be used where there is a clear need to maximise the extraction of DNA from target taxa.

### *Extraction of DNA from plant tissue*

#### General considerations

It has proven problematic to recommend a standardised DNA extraction protocol for plant tissues, due to the inherent variability of tissues within plants (e.g. at a broad level such as bark, leaves, roots, or at finer levels due to variation in turpene concentrations or the cuticle hardness in leaves) and the differing abilities of each extraction protocol/kit to deal with this variation. In addition, plants produce various secondary metabolites that can interfere with both the extraction of high quality DNA and subsequent PCR analyses (Kotchoni et al. 2011). In practice, DNA extraction protocols that are used to recover fungal DNA from within plant tissues are entirely transferable to the extraction of plant DNA from plant tissues. The greater diversity of DNA originating from different species of fungi makes it easier to assess the efficacy of DNA extraction procedures to lyse plant biomass. For example, DNA extraction method was shown to strongly influence the ectomycorrhizal fungal community structure detected in a Cameroon rainforest (Tedersoo et al. 2010), accounting for 15.7% of the observed variation.

#### Extraction requirements and recommendations

A range of methods are used to extract DNA from plant tissues, including fungal and bacterial endophytes (Guo et al. 2001; Maropola et al. 2015). Recent extractions of arbuscular mycorrhizal fungal (AMF) DNA from plant roots have used the DNeasy PowerSoil<sup>®</sup> kit (Martinez-Garcia et al. 2015; Padamsee et al. 2016), or the DNeasy Plant kit for AMF in the roots of *Ammophila arenaria* (Johansen et al. 2015). It is possible that the DNeasy PowerSoil<sup>®</sup> kit gives superior results in samples of tree roots that are higher in phenolic compounds in comparison to softer tissue like grass roots. DNeasy PowerSoil<sup>®</sup> or PowerMax<sup>®</sup> kits are routinely used for sample material, including plant tissues, that are high in inhibitory compounds (Dineen et al. 2010; Maropola et al. 2015), even though relatively low DNA yields are reported for some sample material (Maropola et al. 2015). While there is little evidence upon which to base favouring one extraction method over others, standardisation on one method shown to be adequate (or at least not inadequate) will improve the comparability of results obtained from different laboratories. For these reasons, we recommend the DNeasy PowerSoil<sup>®</sup> kit as a standard protocol for DNA extractions from plant material (Table 1).

### **Long-term storage protocols for DNA**

#### *Rationale for DNA storage*

There are many reasons for the long-term (*de facto* permanent) storage of DNA from sample extractions (Cary & Fierer 2014), even after DNA sequencing of sample material has been accomplished. A justification for long-term storage of DNA extracts is that the costs of sampling far exceed the costs of analysis and storage in most instances. The safe storage of DNA preserves its use for reanalysis without resampling,

either to confirm the reproducibility of results, or to complete alternative assessments on the sample material. The latter facilitates future-proofing through re-analyses using more comprehensive (e.g. microbial metagenomics) and emerging methodologies (e.g. recovery of complete microbial genomes) not currently available to individual researchers. In general, it is preferable to store the DNA extracted from environmental media (normally in volumes of <100 µl) rather than the raw sample itself because of the space and cost constraints associated with storing large numbers and volumes of the original sample material. The preferential storage of DNA extract material means it is very important to ensure that appropriate DNA extraction methods are used and recorded, since once DNA is extracted, the media from which the DNA is extracted is frequently discarded.

Centralised facilities for permanent DNA archival are deemed to be of particular benefit. Such facilities are designed to augment, rather than replace, the storage of DNA within individual research labs and to address common problems with the current practice in sample storage (e.g. loss and lack of access to samples in the deep freezers of individual research labs). The retrieval of samples stored in individual research labs may also be difficult or impossible due to personnel turnover, incomplete labelling, and lack of systematic inventory. The development and use of centralised facilities can provide more streamlined accounting systems to track the location and fate of sample material, and also provide a centralised repository for sample metadata. Commercial room-temperature storage solutions (e.g. those from Biomatrix Inc.) may provide attractive options for sample storage due to lower energy requirements, reduced likelihood of catastrophic failure, and reportedly better sample stability (Lee et al. 2011).

#### *Requirements and recommendations*

Centralised facilities for the room temperature storage of DNA are already used widely for the preservation of DNA for both clinical and forensic purposes (Ivanova & Kuzmina 2013). As the volume of sample material generated for studies of DNA continues to increase, similar facilities should be investigated for the storage of environmental DNA samples, particularly for datasets of national significance, for example prioritising the storage of DNA collected from remote or rare environments, and samples collected over long time series or spatial extents. In addition to providing long-term physical storage of sample DNA, DNA archive information should also be accessible through a secure internet portal, linked to a robust and secure metadata database. Ideally, the storage system should incorporate a management framework that meets the following requirements:

- (i) Mandatory submission of replicate DNA samples suitable for archiving as part of the regular reporting process.
- (ii) Standardised capture of metadata for every submitted DNA sample. We recommend the minimum information about a marker gene sequence, or MIMARKS, framework be followed as developed and recommended by the Genomics Standards Consortium (Yilmaz et al. 2011). This strategy outlines a checklist of core information required for the deposition of DNA sequence data including data pertinent to the investigation such as the project name and the sample environment, e.g. geographic location, collection date and biome. Further, we recommend sample volume and DNA extraction protocol information should be obligatory for samples to be accepted for long-term storage such that biases caused by the use of

different methodologies among studies can be quantified or at least better understood.

(iii) Consistent and standardised naming of samples to minimise ambiguity. Commonly, sample material will be stored using 'linked' or 'coded' naming procedures in which each sample has a unique numeric code linked to the sample metadata via an online database. Such database systems are already in widespread use to preserve DNA sample records in searchable formats (e.g. The NCBI BioSample Database; Barrett et al. 2012).

(iv) Validated protocols to ensure quality and quantity of DNA prior to submission.

(v) Clear and widely accepted mechanisms governing access to archived DNA samples and associated metadata. Such mechanisms must adequately address issues related to indigenous data sovereignty, property (intellectual and physical) ownership, and biosecurity and at the same time facilitate scientific research and the publication of findings.

(vi) An online, ideally cloud-based, portal for inventory of archived samples and associated metadata that facilitates collaborative editing, while providing granular access control by researchers, stakeholders, and the public. This means that different access rights can be provided for different groups, individuals and data, limiting the potential risks associated with unwanted user access to sensitive information.

#### **DNA amplification**

The ability of PCR-based methods to amplify millions of copies of DNA from a single DNA fragment makes PCR a powerful diagnostic tool capable of detecting the presence of even rare, or low biomass organisms. However, this sensitivity also necessitates that care be taken to avoid the generation of false positive results. These can arise from sample-to-sample contamination and, perhaps more commonly, by contamination with DNA amplified in previous PCR reactions, since this creates an abundance of template DNA in the laboratory environment. For this reason, careful consideration should be given to the design and operation of molecular laboratory facilities in which PCR is performed and appropriate controls used to confirm the successful amplification of the target gene. The particular problems associated with the contamination of ancient sample DNA by DNA of modern origin mean that a wealth of information is available on the handling of DNA in appropriately designed laboratories to avoid, or at least minimise, contamination in both pre- and post-PCR processing environments (Cooper & Poiner 2000; Knapp et al. 2012a; Pedersen et al. 2015).

#### *Physical separation of pre- and post-PCR processing*

To prevent the carryover of amplified DNA between samples and experiments, the setup of PCRs should take place in a dedicated area, ideally within a laminar flow cabinet with a UV light to deactivate DNA contaminating the surfaces of the work area. A dedicated set of laboratory consumables, equipment and personal protective clothing is recommended for use during PCR setup to ensure that previously amplified material or other contaminants do not enter the PCR setup area. Where possible, a single-direction laboratory workflow is recommended, which allows the movement of material from clean to contaminated areas, but not from contaminated to clean areas (McDonagh 2003). This workflow may involve the use of separate rooms or dedicated laboratory areas and



equipment for reagent preparation, nucleic acid extraction, PCR setup, and the purification and analysis of amplicons.

#### *Use of appropriate PCR controls*

Assay controls are vital for the detection of false results, as well as poor amplification by PCR. As a guide, the following controls are required or strongly recommended.

**Extraction controls (required).** Known positive and negative specimens for an assay are required to be extracted at the time study samples are processed, to confirm successful nucleic acid extraction and to check for contamination during extraction.

#### **Internal extraction controls (strongly recommended).**

Their inclusion is suggested as an approach to assess for extraction failures and PCR inhibition. Sample material should ideally be duplexed, with one sample being spiked with a quantity of biological material known to be amplified by the PCR conditions and primers used in the main study. Demonstration of the presence of the internal control sequence by PCR in spiked samples only can be used to validate negative results obtained from the corresponding non-spiked samples (i.e. to confirm negative results are not due to DNA extraction or PCR failure).

**Negative PCR controls (required).** These are required to confirm the absence of DNA contaminants in the reagent mix. The DNA template in one or more PCR replicates is often replaced with nuclease-free water for this purpose.

#### **Positive PCR controls (strongly recommended).**

Template DNA that has previously been shown to generate the correct amplification product is required to be included in one or more PCR replicates, to ensure the ability of the PCR to amplify target DNA, where present. Positive PCR controls should amplify consistently but weakly within an acceptable range; the generation of strong products may pose an unnecessary risk as a source of contamination.

**Mock community PCR controls (strongly recommended).** Where available, mock community DNA samples of known composition are strongly recommended to be included as positive PCR controls. The amplification and sequencing of mock community DNA can further be used to assess biases and errors caused by PCR and DNA sequencing processes.

#### Extraction controls

A drawback of the results of most published PCRs is that they lack appropriate extraction controls. Where no amplified product is detected following PCR, this could result from a lack of target sequence in the original sample media, or alternatively be caused by reaction inhibition due to (i) malfunction of the PCR thermocycler machine, (ii) incorrect PCR mixtures or primers, (iii) poor activity of the DNA polymerase enzyme, or (iv) the presence of inhibitory substances such as humic material in the original sample media (Hoorfar et al. 2004).

A common practice is to add a known amount of control DNA after DNA extraction. This approach is suitable to confirm PCR inhibition or failure, but has no value as a true extraction control. The better approach is to process the test sample and an internal control at the same time during DNA extraction. To achieve this, cells or biomass of a known concentration, rather than solutions of DNA, should be added to the sample media. The amplification of DNA from spiked samples, but not from non-spiked samples then signals an absence, or at least low abundance, of the target DNA sequence in the original sample. Proprietary internal extraction controls

are available for purchase (e.g. BioLine DNA Extraction Control 670). A problem arising from the use of these types of positive extraction control is that contaminant DNA from these controls can be difficult to distinguish from the amplified target DNA, potentially leading to false positive detection results. To address this issue, Wilson et al. (2016) suggest using synthetic oligonucleotides as positive controls. Replacing tissue derived controls with distinguishable short synthetic ones seems like a promising approach, reducing the risk of sample contamination, but such an approach has yet to gain widespread use in empirical studies. While the choice of control will depend on the target taxa and DNA sequence under investigation, we suggest that DNA extraction controls be considered obligatory for all metabarcoding analyses of environmental DNA.

#### Negative and positive PCR controls

The inclusion of negative PCR controls is essential in any PCR-based study. Negative controls reveal the presence of contaminating DNA in the laboratory reagents. The amplification of bacterial DNA in negative control reactions (i.e. in PCR mixtures where no sample DNA is added) is a common occurrence, resulting either from external contamination, or residual DNA present in the DNA polymerase solution (e.g. from *Thermus aquaticus* bacteria used to synthesise DNA Taq DNA polymerase). A number of approaches have been proposed for the removal of contaminating DNA from PCR reaction solutions including treatment with UV light (Ou et al. 1991), restriction endonucleases (Carroll et al. 1999) and ethidium monoazide (Patel et al. 2012), with various success. Certified DNA-free polymerases are available for purchase (e.g. MTP Taq, Sigma Aldrich Ltd.) and are recommended as a standardised approach to deal with the issue of DNA polymerase contamination by residual bacterial DNA.

We strongly recommend the inclusion of PCR positive controls. Options for positive controls are various and depend on the context of the study. Where the aim of the study is to determine the presence of one, or a small number of taxa, DNA extracts obtained from these target taxa are an appropriate control. Where the aim of the study is to explore the diversity of complex communities, DNA from 'mock communities' (i.e. where the DNA of multiple target taxa is combined in known ratios in a single solution) may be used to test the efficacy of PCR and biases associated with DNA amplification and sequencing. Positive PCR controls are not essential. However, in cases where no sample DNA can be amplified, amplification of DNA from the positive control sample is helpful to confirm if the PCR procedure failed for all samples (i.e. positive control DNA is not amplified) or if insufficient target DNA was present in the original sample media (i.e. only positive control DNA was amplified).

#### Amplification of mock community DNA

A number of studies, particularly of bacterial communities, highlight the utility of amplifying 'mock community' DNA in parallel to the sample DNA. This mock community DNA is normally comprised of a mixture of DNA derived from a small number of organisms and mixed together in known quantities (i.e. the concentration of DNA from each organism is known). The mock community DNA is treated the same as a sample DNA extract, and the DNA from it amplified at the same time the DNA from environmental samples is amplified. In this way, mock community DNA may be used as the DNA template for positive PCR controls. The inclusion

of amplified mock community DNA in a DNA sequencing sample run allows users to ensure that sequence data from each sample can be accurately identified (i.e. confirming that all of the organisms represented by the mock community DNA are detected by the methods used). The inability of PCR to detect DNA from certain members of the mock community, or their correct ratio, indicate that the abundances of similar taxa may be underrepresented from analyses of environmental DNA. Additionally, analysis of mock community DNA will confirm the presence of DNA erroneously associated with barcoded samples (i.e. 'false positive' sequence reads from DNA not known to be present in the mock community DNA, which may result from sample cross contamination in the laboratory or from DNA sequencing errors). The use of mock community DNA can further be used to reveal potential biases associated with both primer sequence selection, and additional biases associated with choices of DNA sequencing platform (Singer et al. 2016).

Synthetic mock bacterial community DNA, developed by the Human Microbiome Project, or HMP (The N. I. H. H. M. P. Working Group et al. 2009) includes 20 bacterial species in equal concentration, according to ribosomal copy number, and is currently provided free-of charge by BEI resources ([www.beiresources.org](http://www.beiresources.org)). This DNA provides a mechanism to monitor and quantify biases associated with DNA extraction and amplification efficiencies among bacteria (Highlander 2013). The use of mock community DNA is recommended by the HMP and is used by researchers investigating a wide range of sample media (Nelson et al. 2014). Various microbial community standards are also available through commercial providers (e.g. ZymoBIOMICS Microbial Community DNA Standards; ZymoResearch). However, their use appears to be restricted largely to prokaryote DNA. There may be significant advantages for the development and use of mock community DNA for a broader range of taxa, not just bacteria, to inform on the accuracy of DNA barcode sequencing to detect the correct taxa contributing to a sample of mixed DNA. Wherever possible, we recommend the use of mock community DNA as positive PCR controls.

#### *PCR conditions and DNA polymerases*

It is well understood that different PCR conditions and different DNA polymerases will have different amplification biases and errors (Brandariz-Fontes et al. 2015), which could impact the results of DNA metabarcoding studies. Ideally, a single enzyme and set of amplification conditions would be used across all DNA metabarcoding studies, but due to the large variation in sample media, primers, and taxa of interest, this is not possible. For these reasons, and because new enzymes are continually being developed, we hesitate to make specific recommendations, but where possible, we encourage standardisation across similar studies, particularly where there is likely to be a desire to make comparisons between them. For similar reasons, we make no recommendation regarding post-PCR purification methods. We recommend detailed reporting of these methods in publications to facilitate subsequent comparisons of species' detection and community composition across methods.

In addition to the impact of DNA polymerase, various PCR conditions are proposed to either cause or alleviate template amplification biases. Low template concentrations may lead to stochastic fluctuations in the interactions of PCR reagents (Polz & Cavanaugh 1998), whereas high PCR template concentrations may lead to false priming or poor DNA synthesis

due to the obstruction of large DNA polymerase molecules (Altshuler 2006). Because of confounding factors such as the presence of non-target DNA and PCR inhibitory substances, it remains hard to recommend optimal concentrations of template DNA for amplicon PCR.

In general, a greater number of PCR cycles are recommended for samples with low concentrations of DNA or high concentrations of inhibitors. While the importance of PCR cycle number continues to be debated (see Acinas et al. 2005), our recommendation is that the number of PCR cycles be maintained at the minimum required for DNA sequencing, following sample purification, as more cycle numbers leads to more point mutation errors in the DNA amplified (Qiu et al. 2001).

It is generally advisable to carry out replicate PCRs from environmental DNA extracts, to account for the effects of PCR variability. The number of PCR replicates that are generated should be carefully considered in relation to the objectives and design of the study. PCR replicates add substantial costs to a study as more reagents are required to undertake additional PCR and more kits must be used to purify the resulting PCR products. However, the costs associated with sequencing these extra samples are relatively small since additional sample material can be multiplexed on the same sequencing run. We recommend that triplicate PCRs (as a minimum) are carried out from each sample. While this may be sufficient to identify the dominant taxa that are present, a larger number of PCR replicates (up to ten or more) may be advisable in cases when the detection of organisms that are represented by comparatively few DNA molecules is required. Ideally, the PCR replicates from a sample or site would be individually barcoded for sequencing, meaning that all DNA sequence data can be individually identified as belonging to a single PCR. By not pooling individual PCR solutions, subsequent application of statistical techniques, such as accumulation curves (Deng et al. 2015), may help us to determine the probability of presence of a particular taxon in a sample and improve the accuracy of total biodiversity estimates. Ficetola et al. (2015) for example, suggested that eight individually barcoded PCR replicates may be appropriate, as this would allow the detection of taxa with a moderately low probability of DNA detection using an occupancy modelling approach.

#### *PCR additives and enhancers*

An astonishing array of different chemicals are frequently added to PCR reactions to deal with a range of different issues. Such issues include the presence of inhibitors in DNA extracts and DNA templates that are rich in guanine-cytosine (GC) base pairings, which bind more strongly than adenine-thymine (AT) bases, therefore requiring higher melting temperatures for successful PCR. DNA templates that are GC-rich can also hamper DNA amplification as secondary structures are formed, such as DNA hairpins (Creer et al. 2016); these may block DNA polymerases from efficiently replicating GC-rich DNA sequences. In many cases there is little justification for including these additives in PCR recipes other than empirical observations that they improve the yield and/or specificity of the amplification reaction. Some of the more common additives are betaine, bovine serum albumin (BSA), dimethyl sulfoxide (DMSO), dithiothreitol (DTT), 1, 2-propandiol, 7 deaza-dGTP, *E. coli* single-strand binding protein, ethylene glycol, formamide (and other low molecular weight amides), gelatin, glycerol, non-ionic detergents (NP-40, Triton X-100, Tween 20, Tween 40), T4 gene 32 protein and trehalose.

These are sometimes used in combination, for example Combinatorial Enhancer Solution (CES), which is a mixture of betaine, BSA, DMSO and DTT. A number of proprietary PCR additives are also available, e.g. BioStab PCR optimiser (Sigma), Perfect Match PCR enhancer (Agilent), Q solution (Qiagen) and Taq-stabilizer (Jena). Where humic material or other inhibitors are present in the DNA extract, relief is most commonly achieved by the addition of bovine serum albumin, or BSA, into the polymerase chain reaction mix (Jiang et al. 2005). The effects of these additives on amplification of DNA from environmental samples had not been systematically tested, which is a significant knowledge gap. It is possible that the inclusion of a particular additive could bias amplification in favour of particular taxa within a community sample. For this reason, it is important to standardise (or at the very least report) the use of PCR additives across studies that are likely to be compared with one another. PCR inhibition may also be alleviated by dilution of the DNA extract to reduce the concentration of inhibitors to a sufficient level, providing adequate DNA is retained in the extract for amplification to proceed (Tsai & Olson 1992a). If the aim is to maximise the amount of biodiversity amplified from a sample, then it may be worth considering pooling PCR products that have been amplified using different additives and conditions.

#### *Use of blocking primers to enhance PCR amplification of target DNA sequences*

Special amplification protocols may be needed to amplify the DNA of rarer sequences from mixed samples as the predominance of non-target DNA within a sample can bias or otherwise restrict molecular analyses (Polz & Cavanaugh 1998; Vestheim & Jarman 2008). For example, special PCR approaches may be desirable to avoid amplifying DNA from species depositing faecal material (e.g. universal vertebrate primers might not be suitable to analyse the contents of dog faeces), or to assess the stomach contents of predators while avoiding amplification of the predator DNA (Vestheim & Jarman 2008). In such instances, DNA from the depositing or 'host' species can overwhelm the dietary taxa (see Fig. 3 in Shehzad et al. 2012); however, this is not always the case because it is very dependent on the generality of the primer (Wood et al. 2016). A common method for preventing the depositing species' DNA from dominating the result is to use blocking primers, which can be designed to prevent or reduce amplification of DNA from the depositing species (Vestheim & Jarman 2008; Shehzad et al. 2012). A blocking primer is typically designed to bind to the predator DNA but is modified in such a way that it does not initiate PCR amplification. Commonly, three carbon (3C) spacers are incorporated into the sequence of a blocking primer. Spacer 3C modifications to the 3'-end of a primer act as an effective blocking site for the DNA polymerase, preventing the further amplification of that sequence. Primers can be purchased with 3C modifications from a wide range of DNA primer suppliers (e.g. Integrated DNA Technologies, Sigmaaldrich). The usefulness of blocking primers has been demonstrated in studies of microbial (Zulian et al. 2016), environmental (Wilcox et al. 2014) and ancient DNA (Boessenkool et al. 2012).

#### **Sequencing platforms and approaches for the high-throughput analysis of DNA metabarcodes from environmental samples**

##### *Sequencing platforms*

Among available high-throughput sequencing technologies,

Illumina platforms are favoured for most amplicon sequencing applications. The costs associated with sequence analysis using each available technology is hard to quantify because of the wide range of sequence analysis options available. For example, a wide range of Illumina machines are currently available for use and the cost of each varies depending on factors including the sequencing chemistry used (which impacts the average sequence read length) and DNA multiplexing options. For an in-depth summary of sequencing platform attributes, including average sequence read length, throughput (i.e. Gb of data generated), runtime, error rate, instrument costs and cost per sequencing run, readers are referred to Goodwin et al. (2016), and updates of the Field Guide to Next-Generation DNA Sequencers (Glenn 2011).

##### Illumina

Illumina sequencers were first introduced to the market in 2007 (Shokralla et al. 2012). They have been used for a wide range of studies, including to investigate the ability of DNA metabarcoding to indicate the fish species present in water (Kelly et al. 2014) and to analyse the structure of bacterioplankton communities (Tiirik et al. 2014); it is the sequencing platform of choice for the EMP (Gilbert et al. 2014).

Illumina manufactures several different sequencing machines, which differ in their read lengths and sequencing outputs. The HiSeq platform was one of the first introduced, and comes in several different models. They are currently capable of producing paired-end reads of 150 bp in length each and can generate up to 6 billion reads, or 1800 Gb of sequencing data per run (van Dijk et al. 2014). The Illumina MiSeq platform, introduced in 2011, is a benchtop sequencer that can produce reads of varying lengths, dependent on the chemistry used. Currently, the maximum it can achieve is paired-end reads of 300 bp in length (Salipante et al. 2014). This platform can generate up to 15 Gb of data, which translates to around 25 million reads (van Dijk et al. 2014). Illumina also offers the NextSeq platform, which was introduced in 2014. Similar to the MiSeq, this is a benchtop sequencer, but it has a larger output range of 120 Gb, although the maximum read length is shorter at just 150 bp (Reuter et al. 2015).

The biggest advantage of the Illumina sequencers is the low cost per sequence (costs can be as low as US\$7.00 per Gb data using Illumina's HiSeqX platform; Buermans & den Dunnen 2014), which has made it very popular as it allows for high throughput and large coverage (Caporaso et al. 2012). The ability to perform paired-end reads is also advantageous because after merging the two reads, the overlapped region should have fewer sequencing errors (Schloss et al. 2016). The analysis of amplicons that exceed the paired-read length of Illumina systems prevents concatenation of the forward and reverse reads. However, Illumina sequencing platforms are known to have issues when sequencing DNA libraries with low genetic diversity, such as samples containing exclusively 16S rRNA gene amplicons. To avoid this problem, sequence diversity is often increased, by adding genomic DNA from the phage PhiX (Kozich et al. 2013). This results in a loss of sequence recovery from the sample DNA because over 10%, and perhaps as much as 50% of the capacity of an Illumina sequencing run may have to be allocated to PhiX DNA (Kozich et al. 2013). An alternative strategy, is to use a phasing amplicon sequencing approach to shift sequencing phases among DNA originating from samples by adding various additional nucleotide bases to the forward and reverse primers. Although not currently in widespread use, this approach does

appear to ameliorate the problems caused by low sequence diversity, in some cases improving sequence base read quality by 10%, raw sequence throughput by 15% and the number of effective reads by nearly 50% (Wu et al. 2015). Illumina MiSeq platforms are commonly used for amplicon DNA sequencing and a wide range of instrument and sequencing chemistry options are available from Illumina. At the present time the analysis of up to 386 samples on an Illumina MiSeq machine can generate about 17.5 million DNA sequence reads, for ~NZ\$3000, although cost will vary among DNA sequencing providers, and increased charges normally apply if the second stage of two-step PCR is completed by the DNA sequence provider (i.e. the process of labelling DNA from individual samples for multiplexed analysis).

#### Ion Torrent

In 2010, Ion Torrent, a division of Life Sciences, released their Personal Genome Machine (PGM), with the Ion Proton model following 2 years later. This technology has been used to investigate bacterial and archaeal communities in anaerobic digesters (Whiteley et al. 2012), to determine how fungal communities change after forest fires (Brown et al. 2013) and to investigate the diet of big brown bats (*Eptesicus fuscus*) (Clare et al. 2014). The PGM is a relatively low cost instrument capable of producing sequences of 400 bp in length, and can generate 1 Gb of data per run (Reuter et al. 2015). In comparison, the Ion Torrent machine is a more expensive instrument (perhaps comparable to the cost of an Illumina NextSeq machine), capable of producing up to 10 Gb of data per run, but the sequences are shorter, at just 200 bp (Reuter et al. 2015). For this reason, the PGM is more suitable for most laboratory applications, unless very large numbers of sequence reads are required. The Ion Torrent S5 model, released in late 2015, can produce up to 20 million 400 bp reads (6–8 Gb of data) or up to 80 million 200 bp reads (10–15 Gb of data) per run. This machine is cost-competitive with Illumina MiSeq and possibly NextSeq platforms, but at present a broad user community has not been established for molecular ecology.

The Ion Torrent platforms lack expensive components such as optical equipment, instead relying exclusively on the detection of hydrogen ions, which reduces the manufacturing cost, and therefore the upfront investment required to purchase the machines, making them more accessible to scientists (Rothberg et al. 2011). It is also the fastest benchtop sequencer available to date (van Dijk et al. 2014). However, the semiconductor sequencing technology suffers from high error rates for homopolymers because there is not a perfect correlation between the number of identical bases incorporated and the observed voltage change (Bragg et al. 2013). It has been reported that premature truncation of the reads occurs and that this can be due to biases associated with not only the orientation of the read (forward or reverse) but also the organism the template strand originated from (Salipante et al. 2014). This could adversely affect community profiles generated using this platform. At the present time the analysis of samples on an Ion Torrent PGM should generate about 5 million DNA sequence reads of ~400 bp length for ~NZ\$2000, although cost will vary among DNA sequencing providers. Increased charges normally apply for highly multiplexed sample analyses if additional sample processing is required by the DNA sequence provider.

#### Other sequencing technologies

As well as the sequencing technologies detailed here, several

others are, or have been, available. These are not described here as they are not considered optimal for eDNA research. This includes: Sanger sequencing, which was first developed in 1977 (Sanger et al. 1977); SOLiD technology, an NGS platform introduced by Applied Biosystems (Life Sciences) in 2007 (Liu et al. 2012); Helicos, a single molecule sequencer released in 2008 (Harris et al. 2008); and PacBio which was introduced by Pacific Biosciences in 2011 and performs single molecule sequencing (SMS), therefore being referred to as a ‘third generation’ sequencing technology (Eid et al. 2009). Additionally, the 454 sequencing system, which was the first next-generation sequencing technology to become available to scientists (Margulies et al. 2005), is not discussed further, as it is no longer supported or produced by the manufacturer (Roche). The cost-per-nucleotide of 454 sequencing data was markedly higher than for Illumina and Ion Torrent systems, and as Illumina or Ion Torrent read lengths increase, uptake of these platforms for fungal community analysis has similarly increased (Lindahl et al. 2013).

As well as redundant sequencing technologies, we are yet to consider emerging technologies. A very promising sequencing technology, which has shown significant advances in both read length and accuracy (Jain et al. 2016), is the MinION platform, produced by Oxford Nanopore Technologies. This sequencer is novel in that it is a small, hand-held device and as such is the first truly portable sequencing platform (Mikheyev & Tin 2014). It also has the advantage of allowing real-time targeted sequencing, both in terms of being able to sequence in the field, as well as conducting real-time analysis during the sequencing run (Jain et al. 2016). While it can generate reads as long as 98 kB, it is again plagued by extremely high error rates (Laver et al. 2015). Mikheyev and Tin (2014) reported that less than a quarter of reads generated from resequencing the lambda phage genome could be mapped to a reference sequence, with less than 10% identified. When using the two direction reads protocol and a more recent version of the MinION chemistry (R7), Laver et al. (2015) reported an error rate of 38.2%, thereby demonstrating significant improvement with new iterations. While the advances in Nanopore sequencing are exciting, error rates must be further improved before the technology can be considered accurate or reliable. It is currently best suited to clinical applications, for example for the identification of infectious agents. However, it has been used to identify bacterial species present in synthetic communities with low diversity, and was shown to accurately identify the majority of species present (Brown et al. 2017). Such results suggest that, with more development and improvements, this hand-held sequencer could one day be used successfully for eDNA research.

#### *Introducing index sequences onto DNA fragments to enable high-throughput analysis of multiple samples in a single DNA sequencing run*

If any of the NGS technologies described above were employed to analyse a single environmental sample, they would likely generate sequence data well in excess of what would be required, due to the high sequencing outputs. Therefore, a more cost and time effective approach is to pool multiple samples into a single sequencing run to get sequencing data on all samples simultaneously, at a more appropriate, ‘shallower’, sequencing depth. To achieve multiplexing, a specific tag, or barcode should be added to each DNA fragment to provide information on which sample that particular fragment came from. DNA barcodes can be added by PCR (either a one-step

or two-step process), or through ligation reactions (Meyer et al. 2008). Barcodes, or tags, can be applied either to one end only, or to both ends of the DNA to be sequenced. However, DNA labelled by tag sequences appended to primers are prone to cross-contamination as the tags may dislodge and switch location among amplicons (Esling et al. 2015). For this reason, double-tagging each DNA fragment is preferable as it allows for the detection of these ‘mistagging’ events; any sequence with unexpected tag combinations can be identified as ‘mistags’ and discarded (Philippe et al. 2015).

The one-step PCR approach for multiplexing samples requires a large set of primers, each with their own unique DNA barcode already included (Knapp et al. 2012b). Each sample will then be amplified by a different set of primers, resulting in all the PCR products from a single sample having a unique combination of barcodes at either end. While this approach is simpler in terms of the labour needed, it does require researchers to invest in a large number of unique primers, proportional to the number of samples that are to be included on a single sequencing run. As sequencing platforms change and improve, large sets of barcoded primers may become redundant and would have to be replaced. This will most likely occur as users wish to amplify longer gene regions to take advantage of advances in DNA sequencing technology. However, if using standard desalted primers, as recommended by Illumina (Illumina undated), the costs of primer purchase are unlikely to be prohibitive (e.g. approximately NZ\$25 for each primer of 50 bp length, at a concentration of 25 nmole). Alternative approaches that use a single PCR step, but do not require the use of pre-barcoded primers are also used, where the target gene region is amplified and barcodes are added in a single PCR using a linker sequence (Clarke et al. 2014). Two-step PCR approaches allow users to amplify their gene region of interest using standard primers carrying overhang adaptor sequences at their 5' ends. The overhang adaptor sequences facilitate the later addition of barcodes to DNA sequences originating from each sample during a second PCR (Bybee et al. 2011). If a new set of primers are required, or desired, the same set of barcoding tags can be used for the new primers. Therefore, investing in these barcodes, rather than barcoded primers, is likely to be more cost-effective.

In theory it is possible to multiplex thousands of samples in a single sequencing run (Caporaso et al. 2012). However, in most cases, researchers will limit the number of samples multiplexed to less than 384, and this is what we also recommend. Sample numbers are often restricted to ensure an adequate average number of DNA sequence reads per sample, which could exceed 20 000 on the correct Illumina MiSeq platform (Kozich et al. 2013), but is frequently far less, particularly after low quality sequence data is discarded. The Illumina Nextera XT Library Preparation kit, commonly used to add the barcoded primers during the second PCR step, is currently limited to the analysis of 384 samples.

#### *Recommendation of a standardised approach for DNA sequencing*

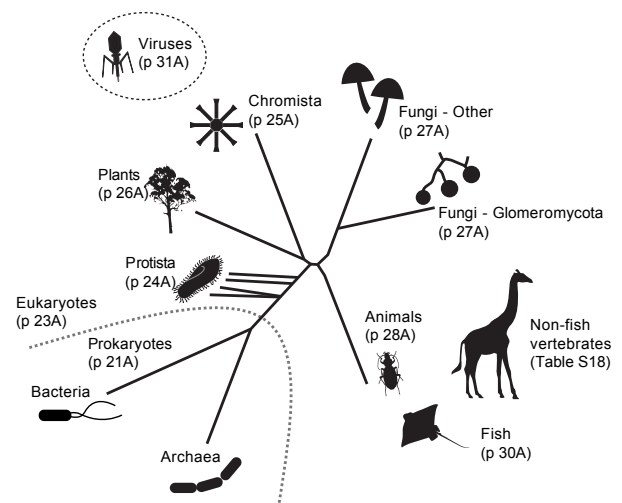
Choosing the most suitable sequencing platform for a study is based on several considerations, including the quality of the sequences, the number of sequences that can be obtained at what cost, and the length of the sequences. The reason for the first consideration is clear; the better the quality of the sequences, the more reliable the data. The second consideration is especially important for large-scale studies, which aim to analyse many samples, as is often the case with

DNA metabarcoding studies. Sequencing platforms with a greater output will be more suitable to multiplexing a greater number of samples, and therefore will be more cost-effective. Finally, longer sequences are easier to assign to taxa (Wang et al. 2007), therefore, the length of sequence produced must be considered. While different studies may benefit more from the use of some platforms than others, it would be largely beneficial to use a standardised sequencing approach to make inter-study comparisons as compatible as possible.

The most attractive candidates are the Illumina and the Ion Torrent sequencing systems. The lower cost per sequence associated with Illumina sequencing, and the higher error rates associated with Ion Torrent sequencing, especially the taxon-specific biases identified by Salipante et al. (2014), indicate that Illumina would currently be a good choice for a standardised sequencing platform. Illumina sequencing is already the preferred platform for many studies performing amplicon sequencing, and as the read length of this sequencing technology increases, it will become suitable for many more DNA metabarcoding studies (Lindahl et al. 2013). For this reason, we currently recommend use of the Illumina DNA sequencing platform for DNA metabarcoding, in conjunction with a two-step PCR barcoding approach for the analysis of sample DNA multiplexed into a single sequencing run, as described in the Illumina 16S Metagenomic Sequencing Library Preparation guidelines (Illumina undated). The sequence analysis of 384 samples, including costs associated with the addition of barcoded tags in the second PCR step, is achievable for ~NZ\$7000 on Illumina MiSeq machines, operated by commercial DNA sequencing providers.

#### Standardised approaches for the amplification of DNA from different taxa

In this section, we identify gene regions and primers suitable for the analysis of different taxa for their DNA ‘barcodes’ (Fig. 9). In an ideal world, a single primer for a gene region might



**Figure 9.** Taxa for which individual DNA amplification protocols are provided. The amplification procedures outlined here will not detect all species, or necessarily provide resolution to species level, but are intended to provide a broad overview of biological diversity using DNA metabarcoding approaches. The page numbers associated with different taxa indicate the location of in-depth DNA amplification protocols within this document.

be used to amplify all prokaryotes or eukaryotes. However, specific taxa may or may not be resolved by taking this general approach. Hence, we identify methods that target various taxa at different levels of resolution. There are many reasons why users may wish to deviate from the protocols outlined below, including for the amplification of shorter DNA sequence regions as may be desirable for the analysis of degraded ancient DNA (Pedersen et al. 2015).

The general PCR protocols outlined in this document can be adapted for use on any PCR amplicon, providing the length of the amplicons generated falls within the acceptable range of the DNA sequencer and sequencing chemistry, i.e. in the present document we assume the use of Illumina MiSeq v3 kits to generate  $2 \times 300$  bp paired end runs, meaning amplicons should, where possible, not exceed  $\sim 500$  bp. After identifying suitable primers, Illumina ‘overhang sequences’ must be added to the locus specific primer sequences. These primer overhang sequences allow the later addition of Illumina sequencing adapters, which are essential for DNA sequencing using this platform, and the barcoding indices used to identify the sample origin of each DNA sequence. Ideally, the locus specific portion of the primer (i.e. the primer sequence corresponding to the target DNA sequence, excluding the overhang region), should have a melting temperature of  $60\text{--}65^\circ\text{C}$  (Illumina undated).

For example, a researcher may wish to use the primers NC1 (forward) and NC2 (reverse) to target the ITS-2 rRNA gene sequence of nematodes, as demonstrated by Avramenko et al. (2015). The sequence of these primers is as follows:  
 NC1: 5'-ACGTCTGGTTCAGGGTTGTT-3' (Gasser et al. 1993)  
 NC2: 5'-TTAGTTTCTTTTCTCCGCT-3' (Gasser et al. 1993).

As shown below, Illumina Nextera overhang sequences (underlined) must be added to the gene-specific primer sequences (bold) following the guidelines presented in Illumina (undated) to facilitate Illumina MiSeq DNA sequence analysis.

NC1 5'  
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG  
**ACGTCTGGTTCAGGGTTGTT** 3'  
 NC2 5'  
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG  
**TTAGTTTCTTTTCTCCGCT** 3'

Once amplified by PCR, the quality and quantity of the product should be assessed by running the DNA on a gel, or using a Bioanalyzer Instrument (Agilent), or equivalent instrument. Where PCR products are confirmed, products originating from each sample should be individually purified (e.g. using AmpPure XP beads as per the Illumina (undated) protocol). They may then be sent to a DNA sequence provider for the attachment of Illumina sequence adapters (allowing PCR products to bind to the Illumina flow cell prior to sequencing) and dual indices (allowing the sample origin of each DNA sequence to be ascertained even after the PCR products from multiple samples are pooled in a single sequencing run). Alternatively, users may wish to complete the full protocol as outlined by Illumina (undated).

### Prokaryotes

Prokaryotes, which can be divided into the two domains of Archaea and Bacteria, are often considered to be ubiquitous in their distribution, occurring in a wide range of environments. DNA-based methods have been used to characterise the community composition of prokaryotes across a wide range of environments including soils (Griffiths et al. 2011), surface

waters (Lear et al. 2013), groundwater (Sirisena et al. 2013) and air (Fierer et al. 2008), including large or national scale surveys (Griffiths et al. 2011; Lear et al. 2013; Ranjard et al. 2013). The resulting datasets have generated valuable insights into variabilities in prokaryote communities across land-uses and landscapes, as well as key environmental drivers or correlates of prokaryote community composition such as variation in climate (Lear et al. 2013), pH (Lauber et al. 2009) and pollution gradients (Yergeau et al. 2012). While these large-scale investigations have traditionally relied on DNA-fingerprinting methods, DNA metabarcoding and sequencing strategies are becoming more common. For example, high-throughput amplicon sequencing has recently been used to investigate continental-scale distributions of Bacteria (Barberan et al. 2015) and community responses to stresses such as fumigation (Wei et al. 2016).

The compositions of prokaryote communities in soil and water have been demonstrated to respond in a somewhat predictable manner to changes in land-use and management, as well as to specific pollution and management events (Lear et al. 2009; Yergeau et al. 2012; Ancion et al. 2014). This has prompted investigation of the potential of bacterial community data as a biological indicator of environmental conditions (Lau et al. 2015; Hermans et al. 2017). Such approaches are suggested to be of particular value in situations where traditional biological indicator communities (e.g. fish, macroinvertebrates) perform poorly, such as in artificial or enclosed systems, or for the assessment of highly degraded environments (Ancion et al. 2014).

Overall, Bacteria and Archaea contribute to the generation and maintenance of many ecosystem processes, and therefore are important contributors to many industries that rely on our environment. Given their significance, the detection and quantification of these organisms in a wide range of ecosystems is extremely valuable; DNA metabarcoding makes this possible.

### *Current practices for the analysis of prokaryote communities with DNA barcodes*

The majority of studies seeking to characterise the composition and diversity of prokaryote communities analyse 16S rRNA genes, which encode for a component of the 30S small subunit of prokaryotic ribosomes. The 16S rRNA gene is commonly used in phylogenetic studies because it is highly conserved among species of Bacteria and Archaea. However, some Archaea, particularly thermophiles, contain 16S rRNA gene introns in otherwise highly conserved regions, which may impact the annealing of some ‘universal’ primers used to target this gene (Parada et al. 2015). Comparison to genome-based taxonomies show that the taxonomic resolution of this marker gene alone is not sufficient to classify to ‘species level’ using even full-length sequences (Konstantinidis & Tiedje 2005; Richter & Rossello-Mora 2009).

A number of international sequencing efforts, including the EMP, have developed, and continue to update, protocols for the extraction, processing and sequencing of 16S rRNA genes for the analysis of prokaryote communities (Caporaso et al. 2012). The approach recommended by the EMP is in widespread use and in July 2014 had already been used to catalogue at least 30 000 samples for community analysis, from more than 200 collaborators collecting samples from over 40 biomes (Gilbert et al. 2014). More recently, officially supported protocols for the amplification of 16S rRNA genes have been provided by Illumina (undated). This protocol targets the V3 and V4 region of bacterial 16S rRNA genes and largely excludes the

amplification of archaeal DNA. Despite the widespread use of 16S rRNA genes as an amplification target for prokaryote DNA, chloroplast as well as mitochondrial rRNA genes are amplified by 16S rRNA gene specific primers (Rodríguez-Marconi et al. 2015), an observation commonly used to support the endosymbiotic bacterial origin of mitochondria and chloroplasts (Ochoa de Alda et al. 2014). For this reason, care should be taken to remove chloroplast and mitochondrial DNA sequence data before analyses are undertaken that may otherwise lead to inflated estimates of prokaryote community diversity.

The accuracy of 16S rRNA gene amplicon analysis is strongly dependent on the choice of primers. The impact of chimeras, differing specificities and error rates, among other issues, should be considered. Primer choice has been debated extensively and hundreds of primers have been examined covering different regions of the 16S rRNA gene, as well as varying specificities (Schloss et al. 2011; Soergel et al. 2012; Klindworth et al. 2013; Kozich et al. 2013; Tremblay et al. 2015). Commonly used primers include those adapted and used by the EMP, namely 515F/806R (Caporaso et al. 2011, 2012), 515F/806RB (Apprill et al. 2015) and 515F-Y/926R (Parada et al. 2015), all of which target the V3-V4 region of the gene (see Table S5 in Supplementary Material). This region has been shown to result in fewer chimeras (Haas et al. 2011) and lower error rates (Kozich et al. 2013), but with potentially lower OTU detection in specific environments and from mock community DNA (Parada et al. 2015). Although these are general observations, the most informative gene region may vary between ecosystems, largely as reference databases may contain many sequences derived from some environments, but not others (Soergel et al. 2012). The clear consensus is that no method or primer pair are perfect for all sample media, and these differences can only be determined by preliminary assessments (Schloss et al. 2011; Soergel et al. 2012). Comparisons of amplicon sequencing results to metagenomes is a good approach to assess primer bias against important taxa (Klindworth et al. 2013). Additionally, comparison of primer sequences to DNA sequences found in comprehensive databases like SILVA (Quast et al. 2013) and Greengenes (McDonald et al. 2012) allow the theoretical ability of the chosen primer set to amplify sequences in the databases to be tested. Regardless of primer choice, the use of consistent primers will likely provide the most reliable comparisons of data across environments (Kozich et al. 2013; Tremblay et al. 2015). As a result, we recommend adoption of the EMP primers targeting the V3-V4 region, which correlates strongly with community profiles determined by shotgun sequencing (Tremblay et al. 2015).

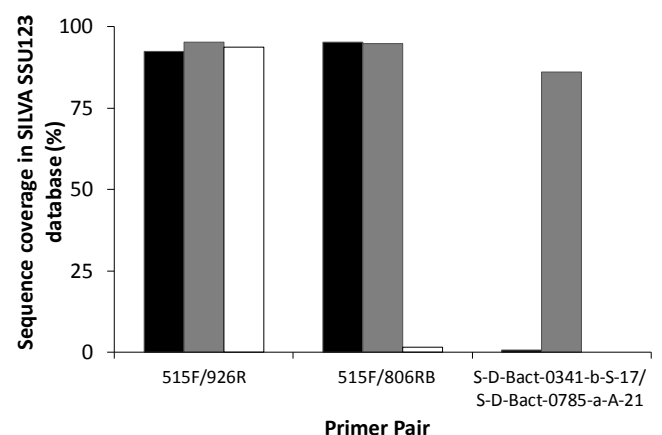
The currently recommended primer pair by the Earth Microbiome Project (Caporaso et al. 2012) for the universal amplification of all prokaryotic organisms consists of forward primer 515F (5' GTGYCAGCMGCCGCGGTAA 3') and reverse primer 806RB (5' GGACTACNVGGGTWTCTAAT 3'). Our own *in silico* analysis of these primers (performed by Janine Kamke; see Figure 10) suggest they cover 94.8% of all Bacteria and 95.2% of Archaea sequences without losing specificity for prokaryotic organisms. Our analysis allowed one mismatch of the non-redundant SILVA small subunit 16S rRNA gene database (release SSU 123) (Pruesse et al. 2007) using TestPrime (Quast et al. 2013). About 1.5% of the sequences identified by our *in silico* tests belonged to 18S rRNA gene sequences in the database.

The primers 515F and 806RB target a 301 bp region (V3 and

V4) of the ~1,500 bp prokaryote 16S rRNA gene. Alternative primers amplifying longer regions (e.g. 515F/926R) have been tested for use and have resulted in more accurate assessments of relative abundances of mock community DNA (Parada et al. 2015). However, according to our own *in silico* tests, this alternative primer combination showed poorer specificity for prokaryotes, matching 92.4% archaeal and 95.3% bacterial 16S rRNA gene sequences in the database, but also 93.7% of eukaryote 18S rRNA gene sequences in the SILVA SSU 123 release. Hence, the primers appear more biased towards the amplification of non-prokaryote DNA, while DNA from Archaea may not be as well represented in the final sequence analysis compared to data generated using the primer pair 515F/806RB (Figure 10). Inevitably, coverage for individual sub-phyla will vary between different primer combinations. For example, the correction of a single base mismatch is observed to increase the apparent abundance of Thaumarchaea in field samples (Parada et al. 2015). Therefore, depending on the expected microbial community and taxa of interest, one primer pair might be favoured and lead to higher accuracy for a specific sample, but this needs to be assessed on a case by case basis.

#### Recommendation

Based on (i) our analyses of these commonly used primer pairs for the purpose of gaining the highest coverage of prokaryotes (both Bacteria and Archaea), while excluding the DNA of eukaryotes, and (ii) the popularity of the already highly standardised EMP approach, we recommend the primers 515F/806RB and the Amplification Protocol version 6\_15 of the EMP (Caporaso et al. 2012) for the routine analysis of prokaryote communities from 16S rRNA gene sequence data. A more detailed description of our recommended protocol can be found in see Appendix S6 in Supplementary Material. The primer pair 515F/806RB (Apprill et al. 2015) include unpublished modifications (e.g. later iterations on the approach recommended by Caporaso et al. 2012) to remove known biases with prior iterations, including poor amplification of DNA from Crenarchaeota/Thaumarchaeota and the Alphaproteobacterial clade SAR11.



**Figure 10.** Percentage of (■) archaeal, (■) bacterial and (□) eukaryotic taxa in the small subunit SILVA database (release SSU123) having 16S rRNA gene sequences that are predicted to match three commonly used primer pairs.

## Primer Details

**515F:** 5' GTGYCAGCMGCCGCGGTAA 3'

**806RB:** 5' GGACTACNVGGGTWTCTAAT 3'

Note: Only the prokaryote-specific primer sequences are shown above. Additional primer components, including multiplex barcodes and platform-specific sequencing adaptors are required for high-throughput DNA sequencing. A more detailed protocol is provided in Appendix S6.

## Eukaryotes

The eukaryotes are extremely diverse, ranging in size from microscopic single-celled organisms to the largest plants and animals on Earth. Microscopic eukaryotes, such as protists and fungi, are present in nearly all habitats, and have important roles in ecological processes such as decomposition and nutrient cycling. Plants and algae carry out photosynthetic primary production, and provide much of the physical structure of ecosystems. The animals include an extraordinarily diverse range of sizes and morphologies, from microscopic invertebrates to large mammals, birds and reptiles. Overall, eukaryotes are larger and more complex but less numerous than prokaryotes. Plants and animals are typically less abundant and more heterogeneously distributed than protists and fungi, but they make up most of the visible biosphere.

### *Current practices for the analysis of eukaryote communities with DNA barcodes*

The most widely used target for metabarcoding of a broad range of eukaryotes is the ribosomal 18S rRNA gene (see Table S7 in Supplementary Material), which is present in all eukaryotes as highly-expressed multiple copies and encodes for a component of the eukaryotic small ribosomal subunit (40S). It consists of nine variable regions interspersed with conserved regions, making it relatively straightforward to identify PCR primer sites spanning different variable regions, or combinations of variable regions, resulting in amplicons of a variety of lengths (~150 to >1000 bp). It is not easy to design primers that include all branches of eukaryote life. The polyphyletic make-up of protists, for example, means that 18S primer combinations are likely to overlook one or more branches of this group. The 18S rRNA gene region generally lacks sufficient resolution to discriminate species (with some exceptions), but is effective for resolving lineages at higher taxonomic levels.

Various PCR primers, targeting most regions of the 18S rRNA gene, have been used in one or more environmental metabarcoding studies since 2010. These studies have most commonly focused upon the analysis of protists, but fungal and metazoan taxa are usually detected also. The EMP recommends the primer pair Euk\_1391f/EukBr which amplifies approximately 150 bp of the V9 region of 18S rRNA genes. These primers are based on those developed by Amaral-Zettler et al. (2009) and are described as primarily targeted at protists, but they also match organisms from other major eukaryote taxa. This gene region was used by Ramirez et al. (2014) in their assessment of protist biodiversity in Central Park, NY, soils. In other cases, the V3 region (180 bp), and V5–V7 region (535 bp) have been used to analyse protist communities in a lake and a geothermal feature, respectively (Nolte et al. 2010; Meadow & Zabinski 2012).

The V4 region is the longest and most variable section of the 18S rRNA gene, which means it has higher resolution than other 18S rRNA regions. The Consortium for the Barcode

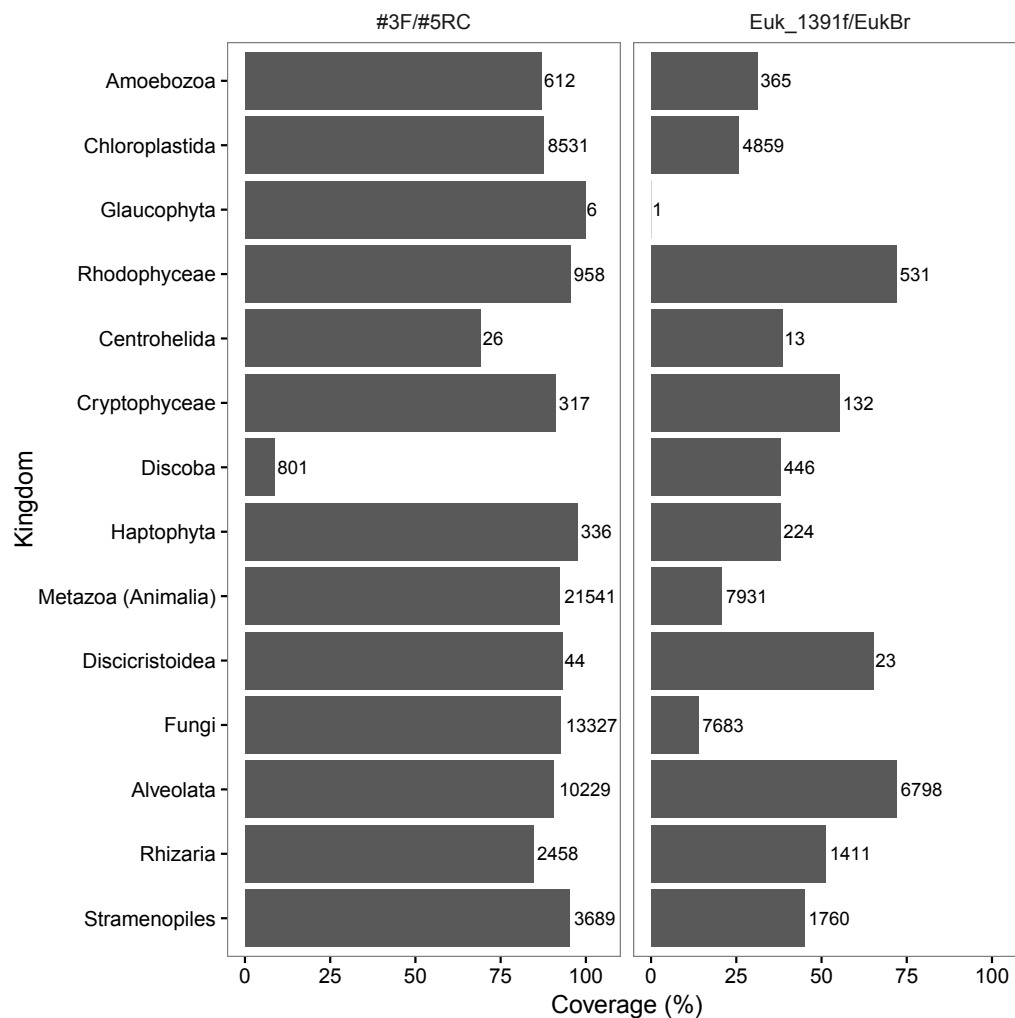
of Life (CBOL) protist working group recommended this region as a target for DNA barcoding of protists, albeit in combination with various other genes to achieve species resolution (Pawlowski et al. 2012). Hadziavdic et al. (2014) carried out a full characterisation of conserved and variable regions across the 18S rRNA gene, resulting in the identification of optimal universal primers spanning the V4–V5 region. Similarly, primers targeting the 18S V4–V5 region were used by Bates et al. (2013) to analyse micro-eukaryote diversity in soils. The regions amplified by these primer sets are about 600 bp to 650 bp (including primers); however, they exceed the ideal length for sequencing on the Illumina MiSeq system.

Different 18S rRNA regions have been used in several metabarcoding studies with greater focus on metazoan taxa. For example, the 18S V1–V2 region (400 bp) was targeted with primers SSUF04 and SSUR22 in a metabarcoding study of meiofaunal biodiversity in marine sediments (Fonseca et al. 2010). The same primers were also used by Creer et al. (2010) in analyses of marine and tropical forest meiofaunal biodiversity, along with primers targeting the 18S V7–V8 region (NF1 and 18Sr2b; 330 bp). In another case, primers that target only the V7 region (~140 bp) were used to analyse eukaryote biodiversity in marine sediment (Chariton et al. 2010).

Various 18S (and 28S) primers intended for metabarcoding metazoan organisms were identified by Machida and Knowlton (2012). One of these primer pairs spans the 18S V4–V5 region, and was recommended as having the highest efficacy of those proposed, but as for Bates et al. (2013) and Hadziavdic et al. (2014), this results in an amplicon of approximately 600 bp including primers, exceeding the optimal read length for the Illumina MiSeq system. A second suggested primer pair (#3/#5RC) that spans the 18S V7–V8 region (about 330 bp, or 375 bp including primers) has been tested in environmental metabarcoding analyses of soil DNA, and was found to amplify sequences from a broad range of metazoans (including arthropods, nematodes, platyhelminthes and annelids), fungi (including Ascomycota, Basidiomycota, and Glomeromycota), protists (including Alveolata and Amoebozoa) and Chromista (Drummond et al. 2015). The #3/#5RC primer pair was assessed as having much better coverage of eukaryote taxa represented in the SILVA rRNA database ([www.arb-silva.de](http://www.arb-silva.de)) than the EMP-recommended primers Euk\_1391f/EukBr, based on the TestPrime function (allowing one mismatch but none within five bases of the 3' end of each primer; Fig. 11). For example, #3F/#5RC matched 91%, 87%, 84%, 93% and 92% of available sequences respectively in the kingdoms Alveolata, Amoebozoa, Rhizaria, Fungi, and Metazoa, whereas Euk\_1391f/EukBr matched only 72%, 31%, 51%, 14% and 21% of sequences, respectively, in these kingdoms (albeit based on a smaller sample due to fewer sequences available covering the V9 region). Discoba was the only kingdom for which Euk\_1391f/EukBr primers (38%) had better coverage than #3F/#5RC (9%). A similar approach also indicated that the #3F/#5RC primers had better coverage of eukaryote taxa than either SSUF04/SSUR22 or NF1/18SR2b. The V7–V8 region amplified by the #3F/#5RC primers was observed to result in somewhat fewer OTUs than the V9 region amplified by Euk\_1391f/EukBr, based on clustering at 97% identity of *in-silico* amplicons derived from about 2700 SILVA database sequences that covered both regions.

There is potential for adaption of 18S rRNA gene sequencing protocols to new technologies, due to the distribution of conserved regions (and known primer binding sites) throughout the molecule. For example, the proximity of





**Figure 11.** Percentage of 18S rRNA sequences from different eukaryote kingdoms in the small subunit SILVA database (release SSU128 RefNR 99; Quast et al. 2013) that are predicted to match primer pairs #3/#5RC and Euk\_1391f/EukBr. Values indicate the numbers of SILVA database sequences from each kingdom that include the target regions of each primer pair, and bars represent the proportions of those sequences that each primer pair matches.

the V7-V8 and V9 regions means that there is a reasonable prospect that the entire V7-V9 region could be analysed, with a modest increase in the sequence read length. Similarly, an increase in read length would make analysis of the V4-V5 region more feasible.

#### Recommendation

The V4 has been recommended as the best 18S rRNA target due to it having the most variability and the highest resolution (Hadziavdic et al. 2014). Unfortunately, suggested primer sites that span this region tend to produce amplicons of lengths longer than ideal for sequencing on the Illumina MiSeq platform. There is some consensus around the use of primers that target the 18S V9 region (about 150 bp), but this is mainly based on efficacy for analysis of protist biodiversity. The V7-V8 region, at over twice the length of the V9 region, can be feasibly sequenced using the MiSeq system, and appears to be useful for analysis of a broader range of eukaryotes. Hence, we recommend the primers #3 and #5RC for amplification of the broadest range of eukaryote DNA from the 18S rRNA region, following the protocol of Machida and Knowlton (2012). In many cases, the amplification of DNA from all Eukarya may not be necessary or even desirable, for example where the aims of monitoring are

to confirm the richness of fish species in a locality. Protocols for the detection of DNA from specific taxonomic ranks of organisms within the kingdoms Plantae, Animalia, Fungi and Chromista, the glomeromycota division of Fungi, as well as the paraphyletic groupings of fish and protists, are provided in later sections of this document.

#### Primer details

**Forward primer #3:** 5'GYGGTGCATGGCCGTTTSKTRGTT3'

**Reverse primer #5RC:** 5'GTGTGYACAAAGGBCAGGGAC3'

Note: Only the eukaryote specific primer sequences are shown above. Additional primer sequences including multiplex barcodes and platform-specific sequencing adaptors are required for high-throughput DNA sequencing.

A more detailed protocol is provided in Appendix S8 in Supplementary Material.

#### Protists

The protists are a paraphyletic collection of single-celled eukaryotes. Here we follow Cavalier-Smith (2010) in excluding the Chromista from the protists because the Chromista and Plantae together form a monophyletic group (the corticates).

Compared to prokaryotes, protists are less well studied using molecular methods, and they are often overlooked when investigating microbial communities. However, this should not be considered a reflection of their lack of importance; it has been argued that protists are crucial in ‘sustaining all life on planet Earth’ (Corliss 2004, p. 12). In the soil they represent a vast pool of biodiversity, similar to that of Bacteria (Bates et al. 2013), and they are important members of trophic chains and nutrient cycles in both terrestrial and aquatic environments (Sherr & Sherr 2002; Ribblett et al. 2005). Despite the challenges that remain for the study of protists, the continuing advances in next-generation sequencing technologies means that we are in a better position than ever to increase our knowledge around the biology, biogeography and ecosystem contributions of these abundant organisms. Several protist taxa (diatoms, foraminifera, ciliates and testate amoebae) are already under investigation, or in use as ecological indicators. For example, foraminiferal metabarcoding has been used to assess the impact of activities as diverse as salmon farming (Pochon et al. 2015a) and oil and gas exploration (Laroche et al. 2016) on the health of aquatic environments. Such studies could greatly improve our ability to assess environmental impacts (Pawlowski et al. 2016).

#### *Current practices for the analysis of protist communities with DNA barcodes*

Because of their paraphyletic state ancestral to other eukaryotes, it is difficult to target protists as a single group (Geisen et al. 2015). One option is to use general eukaryote primers and, where the goal is to focus exclusively on protists, simply exclude non-protist sequences from analysis of the results (Bates et al. 2013), noting also that general Eukaryote primers may miss 50% or more of eukaryotes (Lentendu et al. 2014). Alternatively, primers can be used that are specific to particular protist taxa, such as the Amoebozoa. A recent review of primers for protist 18S rRNA gene sequencing by Adl et al. (2014) lists 193 different primer pairs that amplify DNA from different subgroups of protists. Few of these have been used in metabarcoding applications, with the exception of a small number used by Lentendu et al. (2014). Most of the primers listed by Adl et al. (2014) are likely to amplify non-target sequences, but the degree to which this will compromise results is not yet clear.

Recent analyses of a broad range of protist biodiversity have used general primers targeting a variety of 18S regions, including the V3 region (Nolte et al. 2010), the V4-V5 region (Bates et al. 2013), the V5-V7 region (Meadow & Zabinski 2012), and most commonly, the V9 region (Amaral-Zettler et al. 2009; Ramirez et al. 2014). The V9-targeted primers Euk\_1391f/EukBr are recommended by the EMP for analysis of eukaryotes including protists. Our own analysis (Fig. 11) suggests that the general eukaryote-targeted primers #3F/#5RC (Machida & Knowlton 2012) provide much improved coverage of protists (other than those within the Discoba lineage) than Euk\_1391/EukBr.

#### *Recommendation*

At present, the use of general eukaryote primers, reviewed above, is likely the most cost effective and efficient approach for the analysis of protist organisms (i.e. compared to the use of multiple primer pairs for different protist taxa), despite potentially capturing only a subset of the total protist diversity. Therefore, we recommend using the same protocols are used as for general eukaryotes.

#### **Primer details**

As for eukaryotes.

#### **Chromista**

The kingdom Chromista as circumscribed by Cavalier-Smith (2010) is distinct from the protists and includes the Cercozoa, rhizaria, and oomycetes among other taxa. On land, cercozoans have been described as the most abundant eukaryote in soils, while oomycetes play critical roles as pathogens (e.g. causing sudden oak death, kauri dieback and potato blight). In freshwater systems, didymo (*Didymosphenia geminata*) is an influential invasive species (Bothwell et al. 2014), while oomycetes and other Chromista are important pathogens and free-living saprotrophs.

#### *Current practices for the analysis of chromist communities with DNA barcodes*

Specific primers have been used for 454 DNA sequence analysis of cercozoans and oomycetes (see Table S9 in Supplementary Material), but not for the analysis of the Chromista as one taxon. Efforts to metabarcode oomycetes from soils have had mixed success. Coince et al. (2013) observed only a few different types of oomycete, primarily *Saprolegnia* sp. in French beech forest soils using a nested PCR approach with ITS primers. However, Sapkota and Nicolaisen (2015) demonstrated that with sufficiently high annealing temperature, the amplification of non-oomycete sequences can be minimised.

#### *Recommendation*

At present there are no robust primers for the entire Chromista kingdom. Like protists, this taxonomic group could be detected using general eukaryote primers. However, because of their critical importance as plant pathogens, we suggest that including at least the oomycetes in metabarcoding should be a priority for most soil and plant tissue sampling. The primers and protocols of Sapkota and Nicolaisen (2015) appear to be the most suitable published method to date, but testing using Illumina sequencing is needed. These primers target the non-coding internal transcribed spacer region ITS1, which lies between two coding regions, the 18S and 5.8S rRNA genes.

Our recommended approach for the amplification of the ITS1 rRNA internal transcribed spacer region from Chromista DNA broadly follows the amplification protocol of Sapkota and Nicolaisen (2015). The protocol uses a semi-nested PCR in which a second round of PCR is used to amplify a smaller region of the spacer, using the product obtained from a first PCR as a reaction template. A more detailed description of our recommended protocol can be found in Appendix S10 in Supplementary Material.

#### **Primer details**

Nested PCR 1:

**ITS6:** 5' GAAGGTGAAGTCGTAACAAGG 3'

**ITS4:** 5' TCCTCCGCTTATTGATATGC 3'

Nested PCR 2:

**ITS6:** 5' GAAGGTGAAGTCGTAACAAGG 3'

**ITS7:** 5' AGCGTTCTTCATCGATGTGC 3'

Note: Only the Chromista specific primer sequences are shown above. Additional primer sequences including multiplex barcodes and platform-specific sequencing adaptors are required for high-throughput DNA sequencing.

A more detailed protocol is provided in Appendix S10.

## Plants

Plants are the major primary producers in terrestrial ecosystems and in the oceans; their ecological, economic and cultural importance needs no explanation. Generally speaking, the land plants are far better characterised than the green algae, many of the latter being microscopic. The green algae are paraphyletic with respect to the land plants, and contain vastly more phylogenetic and genomic diversity. Resolution of plant DNA barcodes is generally not to the species-level except in contexts where the species present can be *a priori* restricted to a small set of global diversity (e.g. specific geographic areas or ecological contexts with known floras). For countries such as New Zealand, this limitation is exacerbated by the presence of a number of native land plant flora lineages that have undergone recent and rapid species radiations and make up a large proportion of the species-level diversity, e.g. *Veronica* (Wagstaff & Garnock-Jones 1998), *Raoulia* (Smitsen et al. 2004), *Coprosma* (Wichman et al. 2002) and *Myosotis* (Winkworth et al. 2002).

### *Current practices for the analysis of plant communities with DNA barcodes*

Some DNA metabarcoding studies reporting plant diversity have used 18S rRNA primers targeting Eukarya in general, leading to only broad taxonomic assignments (i.e. to Order or Phylum; Bradford et al. 2013; Drummond et al. 2015). Others have used a part of the *rbcL* gene (encoding the large subunit of ribulose 1, 5 bisphosphate carboxylase/oxygenase) for land plants or green algae (Stoof-Leichsenring et al. 2012) or the *trnL* intron in whole or in part for land plants (Hiiesalu et al. 2012; Drummond et al. 2015). A comparison of studies using these different gene regions strongly suggests that the *trnL* intron provides far greater resolution than lengths of *rbcL* suitable for metabarcoding with the current Illumina technology (Yoccoz et al. 2012). The plastid DNA region encompassing the chloroplast *trnL* intron and *trnL-trnF* intergenic spacer (used either together or each in isolation) have been used extensively in land plant systematics as well as some metabarcoding projects (Lang et al. 2014; Pornon et al. 2016; Yang et al. 2016). Together the typical amplicon length of this region (c. 1 kb) is too great for current Illumina based metabarcoding techniques. While the *trnL* intron is relatively conservative in length, this is not the case for the *trnL-trnF* intergenic spacer, making the latter unsuitable for use in isolation either in standard DNA barcoding or metabarcoding. A drawback of the *trnL* intron is that the most commonly used primer combination (i.e. the c and d primers of Taberlet et al. 1991) do not work in at least some ferns (Trewick et al. 2002) although alternative primer combinations targeting this region are considered to have potential for exploring fern diversity (Chen et al. 2013). Moreover, and perhaps more importantly, Drummond et al. (2015) report that when used in soil, the majority of sequences produced by these primers were of prokaryotic origin.

The consortium for the barcode of life (CBOL) recommends the combination of sequences from two plastid coding regions *rbcL* and *matK* for barcoding land plants. Of these two, *matK* is particularly prone to problems with primer universality, but these are also found with *rbcL* primers, particularly with primers designed for spermatophytes (seed plants) performing poorly for other lineages. Furthermore, the amplicons generally used for standard barcoding using these markers are too long (at least in some species) for current Illumina based metabarcoding techniques. Neither of these gene regions is favoured from the

point of view of discriminating power, having relatively slow substitution rates compared to the candidates discussed below. For green algae, CBOL recommends a region of the plastid *tufA* gene (encoding for the plastid elongation factor, Tu), which, to our knowledge, has not been used in metabarcoding studies.

For degraded plant DNA samples (e.g. extracellular DNA from soil or the gut contents of animals) the short (10– c.150 bp) P6 loop of the *trnL* intron (Taberlet et al. 2007; Ando et al. 2013, 2016) is the preferred target. However, the taxonomic resolution provided by sequence analysis of the P6 loop region is lower than that of the whole *trnL* intron. Nevertheless, the resolution provided by the P6 region is generally higher than that of existing alternative molecular markers (Taberlet et al. 2006) and will generally provide taxonomic assignment to the family level.

The nuclear internal transcribed spacer (ITS1 and ITS2) regions of plants typically have faster substitution rates than plastid sequences, giving them the potential for improved resolution of taxa. Of the two regions, ITS2 shows the greater conservation of length and secondary structure. This makes it a better option for metabarcoding studies as DNA sequence alignment and therefore comparison is more easily achieved among sequences of the same approximate length. Fahner et al. (2016) compared DNA sequence recovery, annotation and sequence resolution among the DNA markers *matK* (the maturase K plastid gene), *rbcL*, ITS2 and the *trnL* P6 loop. The *trnL* intron had only 69% coverage compared to ~82% coverage for the other markers. ITS2 showed the most correct taxonomic assignments of the known sequences, followed by *matK*, *rbcL* and *trnL*.

In considering database coverage of reference sequences, it is useful to consider the identity of species under investigation. For example, in New Zealand the *rbcL* sequences of native flora are available for the great majority of vascular plant genera (with many of the gaps coming from orchids), but species level representation is poor, reflecting the tendency of this gene to vary little, if at all, within genera. Markers such as *matK* and the *psbA-trnH* intergenic spacer are represented only sporadically. In contrast, for the naturalised flora of New Zealand, the *rbcL* marker is well surveyed at the species level (>70%) and near complete at the genus level (>95%). Similarly, nuclear ITS DNA sequences are available for most species naturalised in NZ (>70%) and nearly all genera (95%). Figures for the *trnL* intron are considerably lower at present for species (<30% species) but still near complete for genera (98%). The most appropriate gene marker to use may vary depending on the communities being investigated, but within a New Zealand context, the use of nuclear ITS DNA barcodes would appear preferable.

### *Recommendation*

We recommend the use of ITS2 for DNA metabarcoding studies including plants. It is typically present in genomes in high copy number, is flanked by conserved rRNA genes, is an appropriate length in most organisms for current Illumina sequencing and is universally present. We recommend the primers S2F and S3R of Chen et al. (2010b) as capable of amplifying ITS2 from Embryophyta (land plants) noting that these primers appear to work poorly for ferns; also the specificity of these primers to detect and differentiate among different taxa of green algae requires further empirical confirmation.

### Primer details

**S2F** 5' ATGCGATACTTGGTGTGAAT 3' (Chen et al. 2010b)

**S3R** 5' GACGCTTCTCCAGACTACAAT 3' (Chen et al. 2010b)

Note: Only the plant specific primer sequences are shown above. Additional primer sequences including multiplex barcodes and platform-specific sequencing adaptors are required for high-throughput DNA sequencing.

A more detailed description of our recommended protocol can be found in Appendix S11 in Supplementary Material.

### Fungi

In forming links between soil and plants, fungi are key mediators of plant community structure (van der Heijden et al. 2008), nutrient cycling (Read & Perez-Moreno 2003), carbon acquisition (Deslippe et al. 2016) and plant life and death, thus performing essential ecosystem services (Gianinazzi et al. 2010). Metabarcoding fungi will underpin future exploration of functional diversity of this ecologically diverse group of organisms. Over 100 000 fungal species have been described (Schoch et al. 2014). Extrapolations of fungus-to-plant ratios in a boreal forest in Alaska (Taylor et al. 2014) suggest up to six million species of fungi exist globally, of which 98% are undescribed, although recent work on a global scale (Tedersoo et al. 2014) indicate that diversity may be over-estimated by 2.5 fold. Regardless of uncertainty in diversity estimates, there are vast numbers of fungi yet to be discovered and described. For example, New Zealand has over 6300 described fungal species (Johnston 2010), with an estimated 22 000 species in total (Buchanan et al. 2004).

#### *Current practices for the analysis of fungal communities with DNA barcodes*

The internal transcribed spacer region (comprising ITS1, 5.8S and ITS2) is accepted as the optimal barcoding region for most fungi on the basis of a thorough analysis by Schoch et al. (2012) (see Table S12 in Supplementary Material). Other targets, particularly the 18S nuclear ribosomal small subunit rRNA gene are also used (Chu et al. 2016), but unlike the bacterial 16S rRNA gene homolog, which is often used for bacterial diagnostics, the 18S region has fewer hypervariable regions (Schoch et al. 2012). This limits its ability to differentiate closely related taxa. A fungal-specific forward primer was developed for ITS1 by Gardes and Bruns (1993) and was used in many early metabarcoding studies. Using 454 pyrosequencing it was possible to amplify the entire ITS1–5.8S–ITS2 region using primers ITS1F and ITS4 (White et al. 1990) and sequence either end to obtain ITS1 and ITS2 fragments (e.g. Koele et al. 2014).

The entire ITS1–5.8S–ITS2 region is too long for Illumina-based metabarcoding techniques, making it necessary to choose to amplify either ITS1 or ITS2. Both ITS1 and ITS2 have been used for metabarcoding and recover somewhat different communities of fungi (Nilsson et al. 2009). In a direct comparison of ITS1 and ITS2 in leaf endophyte communities Bazzicalupo et al. (2013) ITS2 had greater interspecific variability, than ITS1. Conversely, Monard et al. (2013) found that ITS1 recovered greater diversity of fungi than ITS2, and that the two regions differed in what fungal taxa were recovered. ITS2 has more data available in the GenBank sequence database than ITS1 (Nilsson et al. 2009). Further, the downstream gene region from ITS2 (28S) is more variable than the 18S and 5.8S regions (Nilsson et al. 2009; Schoch

et al. 2012), so ITS2 may be more adaptable (by extension of the barcode from ITS2 into the large rRNA subunit (LSU)) to methodological improvements in sequencing length. A comparison of ITS and 28S (LSU) regions to assess fungal diversity associated with three forest tree species (Bonito et al. 2014) showed that they both gave comparable results in terms of species diversity, but that the use of LSU barcoding data was limited by a smaller database of reference sequences with many misidentifications. Nonetheless, as a source of more data, LSU may be a reasonable choice for expansion when sequencing read length increases in the future.

#### *Recommendation*

The fITS7 and gITS7 forward primers of Ihrmark et al. (2012) allow fine resolution among Dikarya (comprising the fungal phyla Ascomycota and Basidiomycota) when used in conjunction with the reverse primer ITS4 (White et al. 1990) to amplify the ITS2 region. The effective use of gITS7 may be limited in samples with high plant DNA concentrations because it amplified a higher proportion of plant sequences compared with fITS7 (Ihrmark et al. 2012). Because the two primers differ only in a single base (which is degenerate in gITS7), data collected with either primer should be comparable, but this should be undertaken with some caution. Therefore, we recommend fITS7 for most samples where plant DNA is abundant, with the caveat that gITS7 may be preferable where concentrations of plant DNA are low and greater inclusivity of fungi is a priority. ITS4 is recommended as the reverse primer.

### Primer details

**fITS7** 5' GTGARTCATCGAATCTTTG 3' (Ihrmark et al. 2012)

**gITS7** 5' GTGARTCATCGARTCTTTG 3' (Ihrmark et al. 2012)

**ITS4** 5' TCCTCCGCTTATTGATATGC 3' (White et al. 1990)

Note: Only the fungal specific primer sequences are shown above. Additional primer sequences including multiplex barcodes and platform-specific sequencing adaptors are required for high-throughput DNA sequencing.

A more detailed description of our recommended protocol can be found in Appendix S13 in Supplementary Material.

### Glomeromycota

Glomeromycota are under-represented in metabarcoding studies of soil, even where 'fungal' primers are used (reviewed by Dickie & St. John 2016). The Glomeromycota form arbuscular mycorrhizas with 78% of land plants (Brundrett 2009), including many agricultural and horticultural plant species and underpin plant nutrient uptake and growth. Therefore, these fungi are critically important to include in many metabarcoding studies, particularly in productive landscapes, despite comprising relatively few species (Davison et al. 2015).

One major drawback of the fITS7 primer recommended for fungi (above) is that it matches only 68% of Glomeromycota sequences in GenBank (Ihrmark et al. 2012). Further, the ITS region is considered sub-optimal for identification of Glomeromycota (Hart et al. 2015). Therefore, specific methods for inclusion of Glomeromycota are needed if this critical taxon is not to be overlooked.

#### *Current practices for the analysis of Glomeromycota communities with DNA barcodes*

The molecular methods for metabarcoding of Glomeromycota

are comprehensively addressed by Hart et al. (2015), who conclude that while there is no perfect gene region for the Glomeromycota, the V3–V4 18S rRNA gene region is currently the most widely used (Table S14 in Supplementary Material) and is well supported by a comprehensive database of sequences (Opik et al. 2010).

The universal eukaryote primer NS31 and the Glomeromycota specific primer AML2 (Lee et al. 2008) have been increasingly adopted for 454-pyrosequencing. AML2 has advantages over previously used AM1 primers in being more specific to Glomeromycota (Van Geel et al. 2014). The amplification length of NS31 and AML2 is longer than ideal for Illumina sequencing (~560 bp). Further, the AML2 primer can in some cases form a strong hairpin with Illumina adapters, preventing PCR (Wakelin & Dickie, unpubl. data). The length of the NS31–AML2 fragment can be trimmed using an internal primer, such as the WANDA primer (Dumbrell et al. 2011) which results in a fragment 23 bp shorter.

An alternative 18S rRNA gene primer pair, AMV4.5NF/AMDGR, was used for 454 pyrosequencing by Lumini et al. (2010) and by Cui et al. (2016) for Illumina sequencing of Glomeromycota. These primers amplify a much shorter (~258 bp) fragment internal to NS31–AML2. In a direct comparison using 454 pyrosequencing, Van Geel et al. (2014) showed these primers tended to preferentially amplify Glomeraceae at the expense of other major families (Ambisporaceae, Claroideoglomeraceae and Paraglomeraceae) of Glomeromycota.

#### Recommendation

Further work is required before a strong recommendation for Glomeromycota methods can be made. The NS31–AML2 primer combination has major advantages over other primer sets in terms of specificity, but the amplified fragment may be slightly too long for reliable Illumina sequencing with current technology and has issues with hairpins that need resolution. Nonetheless, we believe this is likely to be the eventual preferred primer set. PCR requirements for these primers, adapted for use on a 454-sequencing platform are available from Davison et al. (2015).

#### Animals

Animals constitute a very diverse taxon, with widely contrasting morphology and lifestyles, and inhabit almost all ecosystems. Most animal species are invertebrates, which represent about 97% of the 1.2 million animal species described to date. However, the number of undescribed (or cryptic) species remains debated, with estimates ranging from 2 to 80 million (Mora et al. 2011; Caley et al. 2014). Due to the high mobility of many animal taxa, the wide range of body size, abundance and distribution patterns including the restricted distribution of certain taxa to remote or difficult to access habitats (e.g. underground, pelagic zone, canopy), biodiversity assessments of animals can be challenging. DNA sequencing of environmental samples is recognised as a means to alleviate limitations around sampling cryptic and difficult to detect species for more than 10 years (Tringe & Rubin 2005). Yet, with the notable exceptions of fish and some aquatic invertebrates, the use of environmental DNA to study metazoan communities is still in its infancy, and lags well behind that of most other taxa; little work has been done on environmental DNA of animal origin in sample media other than water.

#### Current practices for the analysis of animal communities with DNA barcodes

The Folmer fragment (Folmer et al. 1994), or DNA barcoding region, of the mitochondrial cytochrome c oxidase subunit I gene (COI) is likely to be the first choice of marker for animal metabarcoding studies (Table S15 in Supplementary Material) due to (a) the availability of general primers already designed to target the gene, and (b) large quantities of reference data in publicly accessible databases, such as GenBank (Clark et al. 2016) and the BOLD identification engine (Ratnasingham & Hebert 2007). For most animals, sequences of this gene region can be identified to species with a high level of accuracy and precision, if suitable reference data are available for that taxon (Hebert & Gregory 2005).

Deagle et al. (2014) point to a number of problems that they see with using COI for metabarcoding of animals. In particular, they highlight sequence variability in primer sites and the consequent lack of truly universal primers, and a likely need for a cocktail of several more specific primers when attempting to identify all of the animals in a sample. The authors suggest instead using nuclear and mitochondrial ribosomal RNA genes, although they do not recommend any specific primers. Deagle et al. (2014) recognise that the resolution of nuclear ribosomal RNA genes in animals will not be as high as for COI, but claim that mitochondrial rRNA genes are likely to have a similar resolution to COI while allowing for design of more highly conserved primers. The mitochondrial 16S rRNA gene has proven useful for the analysis of environmental DNA from a broad range of taxa, including arthropods and annelids (Table S16 in Supplementary Material). For example, Kartzinell and Pringle (2015) were able to identify arthropods in the diets of a generalist vertebrate predator from faecal samples using a 107 bp fragment of this gene. Pansu et al. (2015a) used a different c. 70 bp fragment of the mitochondrial 16S rRNA gene to amplify earthworm DNA from soil samples. To identify earthworm species from DNA extracted from the faeces of giant carnivorous land snails, Boyer et al. (2013) designed primers for mitochondrial 16S rRNA genes using a sliding window approach (Boyer et al. 2012) implemented in the R package ‘spider’ (Brown et al. 2012) that were specific to earthworms and did not amplify mollusc DNA. In fact, these primers are so specific that they amplify 16S rRNA genes from earthworms that are native to New Zealand (Oligochaeta: Megascolecidae and Acanthodrilidae), but not the common exotic earthworms *Eisenia fetida*, *E. andrei*, *Lumbricus terrestris*, which are found in New Zealand but belong to a different family of oligochaetes (Lumbricidae). Others have used vertebrate specific 12S rRNA gene primers to identify key species of fish, mammals and birds present in seawater, but false positive and negative results are often reported following their use (Kelly et al. 2014; Port et al. 2016).

Other genes useful for developing taxon-specific primers include the nuclear 18S and 28S ribosomal RNA genes because, similar to mitochondrial COI gene regions, they are relatively easy to align, to design universal primers from and are the product of slow evolutionary rates (Machida & Knowlton 2012). Sint et al. (2014) used the 18S rRNA gene to develop two sets of multiplex PCR primers specific for different insect taxa, and then used these to amplify prey DNA from regurgitates of predatory beetles, bodies of predatory spiders, and faeces of predatory bats. However, because the intention was to develop a PCR-based diagnostic assay, these amplicons were not sequenced and therefore their resolution for discriminating among species remains unknown. Koester et al. (2013) took

a similar approach, developing and testing a range of taxon-specific primers for both 18S and 28S ribosomal RNA genes to identify freshwater macroinvertebrates from mixed samples. Primers designed using sequence information derived from 120 28S rRNA and 145 18S rRNA gene sequences of different species belonging to the class Insecta and Malacostraca, and the phyla Mollusca and Annelida were tested for 130 taxa belonging to the phylum Arthropoda, Mollusca, Annelida, and Chordata. Taxa specific primers were then identified as suitable for targeting 21 different types of aquatic macroinvertebrate, highlighting the appropriate use of these genes for the selective detection of target taxa.

Due to a globally coordinated effort to use COI for DNA barcoding, there are currently much more sequence data available for comparison for COI than for rRNA genes. For this reason, we suggest that for metabarcoding studies where identification of species is important, the availability of reference data outweighs the lack of universal primers, and COI remains the marker of choice. For more specific purposes it may be necessary to design taxa-specific primers targeted at particular animal taxa (e.g. to exclude the host when working with the faeces of predators, or exclude any overrepresented, non-target DNA). For this purpose, COI is not ideal because 'COI does not contain suitably conserved regions' to develop reliable taxa-specific primers (Deagle et al. 2014, p. 1) and different genes are recommended for different animal taxa. In metabarcoding studies where species' identity is less important than comparing molecular operational taxonomic units (MOTUs) among samples, researchers may prefer to explore rRNA genes as an option, particularly if it is considered important to have complete and relatively unbiased samples of all animal taxa.

The full COI barcoding fragment as amplified with LCO1490 (LCO) and HCO2198 (HCO) is 658 bp (Folmer et al. 1994), which is not compatible with the main high-throughput DNA sequencing platforms currently used for DNA metabarcoding biodiversity assessment (i.e. Illumina Miseq is currently limited to ~550 bp and Ion Torrent to ~400 bp). Studies using 454-pyrosequencing were able to use fragments of this length (e.g. McGee & Eaton 2015) but this sequencing technology has become largely obsolete as it has been replaced with other platforms that sacrifice read length for greater depth of coverage. A number of primers exist to amplify shorter DNA regions (see Table S16) and have been successfully applied with a range of species (more details are available from the International Barcode of Life Consortium<sup>1</sup>).

The mlCOIintF primer developed by Leray et al. (2013) in combination with the reverse primer HCO2198 (Folmer et al. 1994), or a degenerate version of the latter (e.g. jgHCO2198; Geller et al. 2013) amplifies a 313 bp region (corresponding to the 3' end of the full 658 bp COI barcode). This primer pair yielded amplification success of 89% and 91% respectively for highly diverse assemblages of marine animals (14 and 23 phyla respectively). Brandon-Mong et al. (2015) compared the amplification success of five pairs of universal internal primer pairs on a diverse group of insects (80 species, 11 orders) and also concluded that the combination of mlCOIintF with the original Folmer reverse primer HCO2198 provided the highest amplification success of 64–80%.

Combining the above fragment with amplification of the 5' portion of the barcoding region using mlCOIintF and LCO1490,

could provide full coverage of the official DNA barcoding region, thereby allowing full comparison and contribution to the already extensive Barcode of Life DNA library. However, even with a degenerate version of the Folmer forward primer (jgLCO1490), amplification success of the 5' portion of the barcoding region has been very inefficient with at most 64% of amplification success (Leray et al. 2013). For biodiversity purposes, the 3' portion of the barcoding region is sufficient to obtain good taxonomic resolution at the species level.

#### *Recommendation*

For broad coverage of metazoan biodiversity, we recommend using mlCOIintF with HCO2198 or a degenerate version of the latter, such as jgHCO2198. By incorporating more sequence degeneracy, the latter primer is capable of maintaining broader taxonomic utility but has been found to amplify fungi (especially Ascomycota), in addition to metazoans. Thus, further refinements such as the development of a blocking primer to specifically inhibit non-target DNA from amplification, may be beneficial in some cases (Vestheim & Jarman 2008).

#### **Primers**

##### **Forward mlCOIintF:**

5' GGWACWGGWTGAACWGTWTAYCCYCC 3'

##### **Reverse HCO2198:**

5' TAAACTTCAGGGTGACCAAAAAATCA 3'

##### **jgHCO2198:**

5' TAIACYTCIGGRTGICCRAARAAYCA 3'

Note: Only the animal specific primer sequences are shown above. Additional primer sequences including multiplex barcodes and platform-specific sequencing adaptors are required for high-throughput DNA sequencing.

A more detailed description of our recommended protocol can be found in Appendix S17 in Supplementary Material.

#### *Additional notes on the extraction of DNA from vertebrate animals*

In terms of taxonomic diversity, vertebrates are a relatively minor part of biodiversity. Yet, due to their relatively large body mass, they are the part that the general public are perhaps most familiar with, and, therefore, are an important target for end-users wishing to monitor biodiversity using DNA metabarcoding approaches. Moreover, vertebrates play important roles in ecosystems, such as keystone predation, mediating the transfer of nutrients between marine and terrestrial ecosystems, soil scarification, consumption or physical destruction of vegetation, pollination, and seed and spore dispersal (Diaz et al. 2013; Regan et al. 2015; Hämäläinen et al. 2017). No primers exist for the selective identification of only terrestrial vertebrates. Where the full diversity of vertebrates is of interest, use of the universal animal primers (e.g. mlCOIintF / jgHCO2198) may be appropriate, noting that non-vertebrate animals such as molluscs and arthropods will also be detected among the sample DNA, where present. Where specific vertebrate taxa are of interest, primers specific to these organisms may be selected. It is beyond the scope of this document to recommend DNA amplification methods suited to all major divisions of vertebrate taxa. However, it is worth noting that many of the methods proposed for identification of specific vertebrate taxa from DNA currently target mitochondrial 12S rRNA gene regions. Primers developed for targeting different areas of this gene are considered to be broadly applicable for the detection of a range of vertebrate organisms, including

<sup>1</sup> [www.dnabarcodes2011.org/conference/preconference/CCDB-PrimersetssequencesandPCRprogramsforanimals.xls](http://www.dnabarcodes2011.org/conference/preconference/CCDB-PrimersetssequencesandPCRprogramsforanimals.xls)

mammals (Ushio et al. 2016), birds (Oskam et al. 2010) and amphibians (Riaz et al. 2011). Primers used to identify the 12S rRNA gene barcode from terrestrial vertebrates are provided in Table S18 in Supplementary Material.

### Fish

There is extensive experience worldwide in obtaining fish DNA from aquatic samples and a number of protocols have been designed that specifically target the DNA of fish, while excluding wider aquatic diversity. In freshwater, many environmental DNA studies of fish have focused on detecting single species to determine the ranges of pest fish, such as koi carp (D. West, pers. comm.), introduced fish such as brown trout (Banks et al. 2016), or declining native fish such as torrent fish *Cheimarrichthys fosteri* (R. Holmes, pers. comm.), rather than characterising entire fish communities. For example, researchers at New Zealand's Cawthron Institute recently completed surveys for the New Zealand Department of Conservation for the presence of brown trout in the ranges of rare native fish and for koi carp (*Cyprinus carpio*) after the completion of removal programmes. To detect single or a few species, primers that target the species of interest are used, and the presence of organisms, or rather their DNA, is frequently confirmed by observation of appropriately sized DNA products (bands) on electrophoresis gels following PCR, or with the use of fluorescent reporting markers, such as SYBR or hydrolysis probes (Turner et al. 2014). To explore the diversity of fish assemblages from environmental DNA, the use of multiple sets of 'universal' primers is sometimes recommended, to account for multiple interacting mechanisms including PCR and sequencing biases, as well as the lower collection probability of DNA sequences that are less abundant in environmental water samples (Evans et al. 2016).

#### *Current practices for the analysis of fish communities with DNA barcodes*

Many of the studies conducted to date have aimed to detect one or a few fish species using primers targeting a narrow range of species from DNA extracted from a water sample (Banks et al. 2016). More recently, several published studies have characterised entire fish communities from water samples (Kelly et al. 2014; Port et al. 2016), but using a variety of primers to amplify a range of genetic regions with varying degrees of success (Table S19 in Supplementary Material). Frequently, mitochondrial DNA (mtDNA) markers are chosen since the copy number of mtDNA is greater than that of nuclear DNA per cell, thereby increasing the chances of DNA detection when DNA may be present at low concentration, or degraded (Miya et al. 2015). The large number of related sequences in the Barcode of Life Data Systems (BOLD) database mean that the cytochrome *c* oxidase gene (COI) was initially used as the main gene region to estimate species diversity (Ardura et al. 2013).

The cytochrome *b*, or *cyt-b*, gene has more recently been used to assess diversity at both genus and species levels (Min et al. 2004; Kartavtsev 2011). The *cyt-b* gene is considered to be a particularly useful gene for phylogenetic work based on the observation that it contains many conserved and variable regions or domains. Hence *cyt-b* has been used in numerous molecular studies of fish diversity (Farias et al. 2001). However, a large number of highly conserved gene regions may limit its ability to distinguish among certain taxa (Satoh et al. 2016). Alternative phylogenetic markers include the mitochondrial 12S and 16S rRNA genes of fish. The 12S rRNA gene region

can be used to unambiguously identify fishes to family, genus and species level (Miya et al. 2015). To date, studies comparing the efficacy of gene regions including *cyt-b*, 12S and 16S mitochondrial rRNA genes for fish detection from environmental DNA are rare (but see Kochzius et al. 2010). However, Ardura et al. (2013) found 12S rRNA gene sequence analysis yielded clear and unambiguous species identification among common commercial marine and freshwater fish species. Since no single gene region has yet been identified as an optimal target for use in fish biodiversity analysis, the use of multiple gene targets, perhaps including combination of nuclear and mitochondrial sequences is sometimes recommended (Ardura et al. 2013).

To date, most metabarcoding studies exploring the specificity and inclusivity of molecular barcoding primers for species diversity have focused on assemblages of marine fishes. Far fewer have investigated their viable use as an indicator of freshwater diversity. For example, we could not find any published studies using metabarcoding to characterise New Zealand freshwater fish communities. However, the comparatively limited number of freshwater and estuarine fish species in New Zealand means that sequences for four gene regions that have been used to characterise communities elsewhere could be downloaded from GenBank, aligned and the degree of variation among species assessed to guide the choice of gene region and primers to characterise fish communities. The genetic region with the most comprehensive species coverage is cytochrome *b* with sequences available for 57 of the 63 species in New Zealand fresh waters ([www.niwa.co.nz/freshwater-and-estuaries/nzffd/NIWA-fish-atlas/fish-species](http://www.niwa.co.nz/freshwater-and-estuaries/nzffd/NIWA-fish-atlas/fish-species)). In contrast, there are sequences for 38 species for COI, 48 species for the d-loop region and 37 species for the small subunit ribosomal rRNA gene (12S).

#### *Recommendation*

With no consensus yet emerging as to the most appropriate gene target for biodiversity assessments of fish, we remain unable to recommend a standard protocol for this application. Recently, Miya et al. (2015) developed a set of universal 12S rRNA gene primers (MiFISH-U/E) for the metabarcoding of fish environmental DNA from the analysis of 880 species. These primers were compared to the EcoPrimers used in the metabarcoding study of Kelly et al. (2014) to estimate the composition of fish fauna in a large tank at the Monterey Bay Aquarium. *In silico* evaluations of the MiFISH-U/E (~170 bp) and EcoPrimer sequences (~105 bp) by Miya et al. (2015) revealed that the former contains more sequence variation than the latter and appeared to outperform the latter in terms of correct taxonomic assignment. Out of 180 fish contained in four marine tanks, the MiFISH-U/U primers were capable of detecting 168 species from 59 families and 123 genera. The comprehensive assessment of the MiFish primers by Miya et al. (2015) supports their use in future assessments of fish biodiversity. Given the current lack of consensus regarding optimal gene target regions for fish biodiversity assessment, we are unable to recommend the approach of Miya et al. (2015) as standard practice for fish biodiversity monitoring. Further testing of these primers, including for the analysis of diverse freshwater fish assemblages, is desirable. Nevertheless, the primer set identified by Miya et al. (2015) would appear to be a sensible option for future exploratory investigations of fish community composition with DNA metabarcoding, particularly as more 12S rRNA gene data becomes available for target fish species.

A more detailed description of the MiFish primers and protocol for their use can be found in Appendix S20 in Supplementary Material.

## Viruses

In our oceans, microorganisms are proposed to contribute to more than 90% of the living biomass, of which viruses are proposed to kill 20% of this biomass every day (Suttle 2005). Numbers of bacteriophage outnumber microbial cells by an order of magnitude in most aquatic environments (Chibani-Chennoufi et al. 2004) and are likely outnumbered only by prokaryotes in terms of their biomass (Suttle 2005). Viruses are perhaps among the largest underexplored microbial components in our biosphere and in the oceans alone are predicted to contain 200 Mt of carbon. The loss of bacterial standing stock due to viral lysis is understood to be a critical component of nutrient transfer and recycling (Fuhrman 1999), especially carbon and nitrogen, but also micronutrients such as iron (Poorvin et al. 2004). While the majority of studies on bacteriophage abundance and diversity has focused on samples collected from the marine environment, viral genotype diversity is predicted to be 10–1000 fold higher in marine sediments and soils (Srinivasiah et al. 2008), with perhaps as many as one million viral genotypes present per gram of rainforest soil (Fierer et al. 2007). Such studies indicate that soils likely harbour among the largest diversity of viral material in the biosphere. Soil-borne plant viruses in particular pose obvious negative consequences for agricultural crops, whereas viral pathogenesis of soil microorganisms can have positive or deleterious impacts, depending on whether the affected microorganism is beneficial or harmful to plant growth and development (Brussow et al. 2004). For example, the acquisition of discrete double stranded RNA is observed to increase virulence of the plant pathogenic basidiomycete *Rhizoctonia solani* to potato (Jian et al. 1998). For soils in particular, relatively little is known about the viruses present or their ecological role. Modern molecular methods, and particularly metagenomics, continue to unveil novel viral assemblages and provide new insights into the diversity and implications of the environmental ‘virome’ (Fierer et al. 2007).

## Recommendation

Viruses lack universally conserved signature DNA sequences (i.e. they do not share a common gene region such as the 16S rRNA gene in Bacteria and 18S rRNA gene in Eukarya; Rohwer & Edwards 2002). This lack of universal phylogenetic marker severely impedes their detection in the environment using amplicon sequencing. For this reason, standard metabarcoding practices are rarely used to assess the presence and abundance of viral DNA or RNA. Instead, virus-focused metagenomics studies either concentrate virus-genomes via their virions or enrich for double-stranded RNA that is a broadly shared feature of virus replication. The virus-enriched nucleic acids are then sequenced by shotgun metagenomics methods (Bibby 2013) for the detection of viruses in environmental material, including marine water (Schmidt et al. 2014), fresh water (Mohiuddin & Schellhorn 2015), soil (Ballaud et al. 2016), crop plants (Aw et al. 2016), native plants and weed plant species (Blouin et al. 2016). We remain unable to recommend a single molecular approach to comprehensively target all of the diversity of viral material for metabarcoding studies in an unbiased manner, i.e. the wide range of virion shapes and sizes and the variety of genomes that may be comprised of either DNA or RNA in either single or double stranded forms.

Nevertheless, the relatively short genome of viruses has been shown to facilitate the detection of full genomes, particularly following the concentration or isolation of virus nucleic acids from complex sample media (Cotten et al. 2014).

## Comments on the metagenomic analysis of environmental DNA

While out of scope for the many studies investigating community composition of large sample numbers, metagenomics warrants mention as an important current and future consideration in biodiversity monitoring. Metagenomics is the direct sequencing of genome-wide DNA from mixed communities, and enables the biological diversity of a sample to be studied without the need to cultivate and isolate individual organisms (Handelsman et al. 1998) and avoiding the need for taxon-specific primers. Typically, as already discussed, amplicon-based metabarcoding analyses rely on the exploration of the sequence diversity of a single gene. In contrast, in metagenomics, all genes in a given community can be sampled. The most common approach currently for metagenomics is ‘whole genome shotgun sequencing’, whereby the total DNA in a sample is sheared into smaller fragments before being sequenced at random. With this approach, full-length small subunit rRNA genes (16S and 18S/18S-ITS) can be simultaneously recovered from both the eukaryotic and prokaryotic fractions of metagenomes using a reference guided algorithm (Miller et al. 2011), if researchers have access to large memory computational resources, such as those provided by government-supported initiatives, e.g. the New Zealand eScience Infrastructure (NeSI). No amplification step is required, which is seen as a particular advantage as DNA amplification is known to bias DNA sequencing outcomes because some sequences are preferentially amplified over others (Kim & Bae 2011; Ziesemer et al. 2015) and because the formation of chimera sequences artificially inflates measures of community diversity (Lasken & Stockwell 2007). For these reasons, shotgun metagenomic data are now frequently used as the standard method to assess biases in PCR amplification (Tederloo et al. 2015; Tremblay et al. 2015), on the understanding that less biased amplicon-based approaches will mimic the outputs of metagenomic data analysis more closely. However, prior steps, including DNA amplification, may be necessary to increase DNA concentrations from organisms of interest (Cotten et al. 2014), as is often the case for studies targeting the diversity of viral material.

The metagenomics approach is now over 15 years old and has evolved from the early days when time and cost limiting Bacterial Artificial Chromosome (BAC)/ fosmid/ cosmid libraries and capillary-based Sanger sequence were relied upon (Beja et al. 2000). Sequencing now exploits the high output and relative low cost of next-generation technologies, such as provided by the Illumina platform. Uptake of metagenomics in recent years has increased substantially, and is continuing to gain in popularity, buoyed by improvements in sequencing yields and lengths (Mardis 2013), alongside the development of analysis tools and expertise. As technologies improve, we predict the application of metagenomics will be progressively extended to larger and more complex studies (i.e. those with many samples and/or those requiring more DNA sequence data to be generated per sample), and for the recovery of large and complex environmental genomes (Jansson 2011).

A clear advantage of metagenomics over targeted gene methods is that it captures all taxa, meaning the technique



can be used to look at prokaryotic and eukaryotic diversity simultaneously. In addition to supporting holistic biodiversity studies, the recovery of protein coding genes provides a mechanism to evaluate ecosystem function, services, health and pathogenicity (Forsberg et al. 2012; Fierer et al. 2013). In theory then, this single approach can be used to explore the total taxonomic and functional diversity in any sample. For example, if a large number of genes are recovered in the pathway for nitrogen fixation, this would suggest that the environment has limited available nitrogen, thus selecting for nitrogen fixing bacteria. In some cases, knowledge of the presence and diversity of biological functions may be more important than knowledge of the distribution of organisms. For example, the presence and abundance of virulence factors for pathogenic *Escherichia coli* may provide evidence of disease risk, whereas evidence of *E. coli* may not, since many *E. coli* strains are non-pathogenic (Dozois et al. 2003). Results can also generate clues about the lifestyle of candidate phyla that elude cultivation efforts (Wrighton et al. 2012; Kantor et al. 2013; Brown et al. 2015). For example, although detected in anaerobic environments, the metabolic potential of the uncultivated bacteria BD1-5, OP11 and OD1 were previously unclear. Using metagenomics methods, Wrighton et al. (2012) were able to determine the likely role of these organisms in hydrogen production, sulfur cycling and the fermentation of refractory sedimentary carbon by reconstructing near complete genomes of the organisms present in an acetate amended aquifer. Through holistic sequencing, a single dataset may benefit a greater number of end users and both fundamental and applied science goals.

Despite shotgun metagenomics providing exciting opportunities for biodiversity monitoring, there remains a number of factors limiting its more widespread use. In terms of genome recovery, the technique at present favours small and simple prokaryotic and viral genomes rather than large and complex eukaryotic genomes because the likelihood of obtaining sequence data covering the full genome coverage decreases as genome size increases. Studying eukaryotes, and particularly microeukaryotes using DNA sequencing techniques remains challenging as they remain relatively poorly represented in sequence databases (Escobar-Zepeda et al. 2015). Further, much of the increase in genome size in eukaryotes relative to prokaryotes is caused by an abundance of non-coding DNA sequence, which creates additional bioinformatics challenges, particularly if the focus of study is biodiversity assessment since DNA sequence alignment, and hence phylogenetic sequence differentiation is typically achieved by assessment of protein-coding, rather than non-coding regions (Zhang et al. 2012). For biodiversity monitoring, a key limitation of shotgun metagenomics methods arises where the DNA is largely derived from a small number of community members. For example, the DNA of gut or faecal material may be highly enriched with DNA from the host organism. This may be of little interest for the researcher, as may be the plant DNA associated with phyllosphere microbial communities. In some cases, it may be possible to employ pre-sequence enrichment methods, for example, mitochondria can be concentrated by differential centrifugation prior to the extraction and sequencing of DNA where insect diversity is the main focus of the study (Zhou et al. 2013). Alternatively, bioinformatics procedures may be applied to filter the host DNA from the subsequent dataset (Chew & Holmes 2009; Schmieder & Edwards 2011) if sufficient target sequence were obtained. However, the identification and removal of

contaminant sequences can be challenging (Kunin et al. 2008).

A blend of amplicon and metagenomic sequencing provides the most information-rich and tractable approach at present for large-scale biodiversity studies. In such an approach, all samples would be surveyed using amplicons, and a smaller set of select samples would be further probed using metagenomics, with or without genome assembly (e.g. Fierer et al. 2013; Lax et al. 2014). This makes use of the affordability and depth provided by amplicon sequencing to obtain the big picture, and uses shallower metagenomic sequencing to provide a primer-bias-free assessment of the phylogeny and genomic potential (i.e. the functions and interactions of genes present) of the dominant community members. This hybrid approach is simple because the same DNA extraction techniques and standardisation protocols apply to both amplicon and metagenomic studies; DNA from one extraction can be split and used twice or accessed from an archive for follow-up metagenomic sequencing. As metagenomic sequencing becomes more affordable, the ratio of metagenomic to amplicon sequenced samples will inevitably increase.

## Conclusions and recommendations

We present a set of standard protocols for the identification of a broad range of taxa from the amplification of environmental DNA, while providing a basic introduction to the important considerations for selecting appropriate methods given the taxa of interest and the sample media from which the DNA was extracted (Table 2). Approaches outlined in this document are designed to include coverage of both macro-organisms and microbial taxa and include specific protocols for the assessment of fungi, micro-eukaryotes, plants, animals, fish and prokaryotes. Wherever possible, we sought to align our methods with existing protocols, such as the Earth Microbiome Project recommendations for the identification of prokaryotes and micro-eukaryote taxa. In doing this, we hope to maximise opportunities for researchers from disparate research groups to directly compare, or combine data collected from their own study sites. In all cases, the protocols provided in this document differ in at least some aspect from already published protocols. These changes were deemed necessary to (1) offer a standardised approach for the amplification and sequencing of DNA from diverse taxa, and (2) provide two-step DNA barcoding protocols because we feel this approach is the most robust to incorporate primer changes as sequencing technologies continue to advance (e.g. primers can easily be updated to sequence ever larger DNA fragments), without the need to order large numbers of pre-barcoded primers. We identified Illumina sequencing as the preferred platform for DNA metabarcoding studies.

A trade-off exists between primer coverage (i.e. the ability to detect a broad range of taxa in the target phylogenetic group), specificity (i.e. the ability to selectively amplify only the target data) and sensitivity (i.e. the ease with which the amplified DNA sequence can be used to distinguish between closely related taxa). Since we propose a standardised approach to target the broadest possible range of taxa, it is important to note that the primers selected for use here will not capture the total biological diversity of the organisms targeted. In many cases, the primers suggested will not resolve taxa at the species level. For example, bacterial and archaeal 16S rRNA genes are thought to provide good classification at the family and genus level, but to lack resolution at the species level

**Table 2.** Recommended DNA extraction and PCR amplification approaches for a wide range of taxa and sample media.

	Taxa								
	Prokaryotes	Eukaryotes	Protists	Chromista	Plants	Fungi	Glomeromycota	Animals	Fish
<b>Soil, sediment or leaf litter</b>	DNeasy PowerSoil, PowerMax, or phosphate buffer (depending on sample volume)								
<b>Faeces or ejecta</b>	DNeasy PowerSoil or PowerMax <sup>®</sup> (depending on sample volume)								
<b>Water or ice</b>	DNeasy PowerWater <sup>1</sup>							DNeasy Blood & Tissue	
<b>Animal tissue</b>	DNeasy Blood & Tissue								
<b>Plant tissue</b>	DNeasy PowerSoil								
<b>Target gene</b>	16S rRNA gene (V3 & V4)	18S rRNA gene (V7 & V8)	18S rRNA gene (V7 & V8)	ITS1	ITS2	ITS2	18S rRNA gene	COI	12S rRNA gene
<b>Primers</b>	515f & 806RB	#3 & #5RC	#3 & #5RC	ITS6F & ITS7R	S2F & S3R	fITS7 or gITS7 & ITS4	NS31 & AML2	mICOIntF & jgHCO2198	MiFish-U-F & MiFish-U-R
<b>Fragment length (bp)</b>	290	325	325	350–450	160–320	122–245	5602	313	163–185
<b>Reference</b>	EMP <sup>3</sup> protocol, based on primers of Caporaso et al. (2012)	Drummond et al. (2015) based on primers of Machida et al. (2012)	Drummond et al. (2015) based on primers of Machida et al. (2012)	Sapkota and Nicolaisen (2015)	Chen et al. (2010b)	Ihrmark et al. (2012)	Simon et al. (1992) and Lee et al. (2008)	Leray et al. (2013)	Miya et al. (2015)
<b>Sequencing platform</b>	Illumina								
<b>Description of DNA amplification approach</b>	p 21A	p 23A	p 24A	p 25A	p 26A	p 27A	p 27A	p 28A	p 30A
<b>Protocol details</b>	Appendix S6	Appendix S8	Appendix S8	Appendix S10	Appendix S11	Appendix S13	No protocol	Appendix S17	Appendix S20

<sup>1</sup>If general eukaryote primers are used but animals are the intended focus of the study, then use the Qiagen DNeasy Blood and Tissue kit instead of the PowerWater kit.

<sup>2</sup>560 bp is a slightly longer fragment length than is currently catered for by current Illumina sequencing platforms (once sequencing adaptors are added). This causes problems for analyses based on this amplicon, but future increases in read length are likely to resolve this issue.

<sup>3</sup>The Earth Microbiome Project ([www.earthmicrobiome.org](http://www.earthmicrobiome.org)).

(Staley 2006). Even where present, an organism may remain non-detectable from the analysis of DNA if concentrations of the target DNA are low. Where information is desired on the presence of a specific organism, further tests are usually required to confirm the ability of the primer to amplify DNA from the target organism. Hence, where necessary, primer sequences may need to be adapted, or entirely new target DNA regions identified to ensure that the DNA of key organisms can be amplified using the chosen approach. In many cases, accurate identification will not be possible until marker gene regions, or whole genome data from the taxa of interest are banked in searchable DNA databases. The attempted amplification of DNA from important taxa that have no representative sequences, particularly less-well studied groups, can also be used to test primer recommendations and inform on which taxa may be underrepresented in eDNA studies.

The approaches outlined in this review are perhaps most suited to describe  $\beta$ -diversity, or the relative differences in community composition and diversity among samples. We suggest this because sequencing error and the amplification

of contaminant DNA can provide false positive results, which, if not identified as such, have the potential to initiate a costly chain of events, such as an attempt to eliminate a biosecurity incursion of a falsely-detected unwanted organism (Wilson et al. 2016). We emphasise that the presence or lack of DNA sequence from a particular species should not, by itself, be used as absolute evidence of the presence or absence of the species in that sample environment; additional analysis is recommended to confirm the presence of priority species, particularly species of conservation concern, pathogens, or possible new incursions. Despite these perceived limitations, the analysis of DNA metabarcoding provides enormous opportunities for biodiversity monitoring across all taxa, ecosystems and sample types.

Given the large number of samples now collected for high throughput DNA analysis, investigation of alternative approaches for the long-term storage of DNA extracts is urgently required; no centralised national repository for sample DNA currently exists in New Zealand. One approach for DNA archiving that may be appropriate is the room-

temperature storage of freeze-dried DNA, a method that is already used widely for the storage of DNA for medical and forensic investigations.

We propose a standardised set of DNA extraction procedures for a variety of environmental media which include the use of (a) Qiagen DNeasy PowerSoil<sup>®</sup>/PowerMax<sup>®</sup> kits or phosphate buffer extractions (depending on sample volume) (Bienert et al. 2012) for the extraction of DNA from soil, sediment, leaf litter, faeces and ejecta, (b) Qiagen DNeasy PowerSoil<sup>®</sup> kits for the extraction of DNA from plant tissue, (c) Qiagen DNeasy Blood and Tissue kits for the extraction of DNA from animal tissue, (d) Qiagen DNeasy Blood and Tissue kits for the extraction of DNA from macro-organisms in water and ice, and (e) Qiagen DNeasy PowerWater<sup>®</sup> kits for the extraction of DNA from microorganisms in water and ice. We identified gene regions and primers specific to a broad range of taxa, for the analysis of highly multiplexed sample DNA using an Illumina MiSeq sequencing platform.

In preparing this review we hope to help make metabarcoding, and potentially metagenomics, more accessible to the wider community of biodiversity researchers. The inclusion of metabarcoding data within broader ecological studies is already becoming routine across multiple disciplines including ecology, plant pathology, microbiology and invasion biology. We expect that this adoption will only increase as the techniques become cheaper, more routine, more standardised, and more robust. Where pragmatic, the adoption of appropriate standards will allow for comparison of data collected from the broadest possible range of organisms, ecosystems and sample media. Such an approach may ease the transition from PCR-amplicon based assessments of biodiversity to the analysis of taxa and genes using random shotgun sequencing as this more holistic approach becomes more amenable for use by individual researchers into the future. This will result not only in more comprehensive understanding of biological communities, but will expand our potential for more comprehensive and broad-scale understanding through meta-analyses, combining data from multiple studies at national and international scales.

## Acknowledgements

This document contributes to key aims of New Zealand's Biological Heritage National Science Challenge (NSC) under Project 1.3 'A National framework for Biological Heritage Assessment Across Natural and Production Landscapes'. The document was conceived and planned as part of a 4-day workshop, which was entirely funded by the NSC. We gratefully acknowledge the in-kind contribution of the 55 participants of the NSC workshop on DNA extraction, storage, amplification and sequencing. Many of the ideas and protocols shared at this event, and subsequent discussions form the basis of this review. Gavin Lear, Ian Dickie, Robert Holdaway, Thomas Buckley and Andrew Dopheide have benefited from direct financial support from the NSC. NSC research funds support related projects by Hannah Buckley and Syrie Hermans. The authors declare no support, financial or otherwise, related to commercial science companies or their products that may be seen to pose a conflict of interest.

## References

Abu Al-Soud W, Rådström P 1998. Capacity of nine

- thermostable DNA polymerases to mediate DNA amplification in the presence of PCR-inhibiting samples. *Applied and Environmental Microbiology* 64: 3748–3753.
- Acinas SG, Sarma-Rupavtarm R, Klepac-Cerai V, Polz MF 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries from the same sample. *Applied and Environmental Microbiology* 71: 8966–8969.
- Adl SM, Habura A, Eglit Y 2014. Amplification primers of SSU rDNA for soil protists. *Soil Biology and Biochemistry* 69: 328–342.
- Altshuler ML 2006. *PCR troubleshooting: the essential guide*. Norwich, UK, Caister Academic Press. 80 p.
- Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM 2009. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS ONE* 4: e6372.
- Amberg JJ, Grace McCalla S, Monroe E, Lance R, Baerwaldt K, Gaikowski MP 2015. Improving efficiency and reliability of environmental DNA analysis for silver carp. *Journal of Great Lakes Research* 41: 367–373.
- Ancion P-Y, Lear G, Neale M, Roberts K, Lewis GD 2014. Using biofilms as a novel approach to assess stormwater treatment efficacy. *Water Research* 49: 406–415.
- Andersen K, Bird KL, Rasmussen M, Haile J, Breuning-Madsen H, Kjaer KH, Orlando L, Gilbert MTP, Willerslev E 2012. Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology* 21: 1966–1979.
- Ando H, Setsuko S, Horikoshi K, Suzuki H, Umehara S, Inoue-Murayama M, Isagi Y 2013. Diet analysis by next-generation sequencing indicates the frequent consumption of introduced plants by the critically endangered red-headed wood pigeon (*Columba janthina nitens*) in oceanic island habitats. *Ecology and Evolution* 3: 4057–4069.
- Ando H, Setsuko S, Horikoshi K, Suzuki H, Umehara S, Yamasaki M, Hanya G, Inoue-Murayama M, Isagi Y 2016. Seasonal and inter-island variation in the foraging strategy of the critically endangered red-headed wood pigeon *Columba janthina nitens* in disturbed island habitats derived from high-throughput sequencing. *Ibis* 158: 291–304.
- Apprill A, McNally S, Parsons R, Weber L 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology* 75: 129–137.
- Ardura A, Planes S, Garcia-Vazquez E 2013. Applications of DNA barcoding to fish landings: authentication and diversity assessment. *ZooKeys* 365: 49–65.
- Avramenko RW, Redman EM, Lewis R, Yazwinski TA, Wasmuth JD, Gilleard JS 2015. Exploring the gastrointestinal "nema biome": deep amplicon sequencing to quantify the species composition of parasitic nematode communities. *PLoS ONE* 10: e0143559.
- Aw TG, Wengert S, Rose JB 2016. Metagenomic analysis of viruses associated with field-grown and retail lettuce identifies human and animal viruses. *International Journal of Food Microbiology* 223: 50–56.
- Aylagas E, Borja A, Irigoien X, Rodriguez-Ezpeleta N 2016. Benchmarking DNA metabarcoding for biodiversity-based monitoring and assessment. *Frontiers in Marine Science* 3: 1–12.
- Bainard LD, Klironomos JN, Hart MM 2010. Differential effect of sample preservation methods on plant and arbuscular

- mycorrhizal fungal DNA. *Journal of Microbiological Methods* 82: 124–130.
- Baldwin DS, Colloff MJ, Rees GN, Chariton AA, Watson GO, Court LN, Hartley DM, Morgan MJ, King AJ, Wilson JS, Hodda M, Hardy CM 2013. Impacts of inundation and drought on eukaryote biodiversity in semi-arid floodplain soils. *Molecular Ecology* 22: 1746–1758.
- Ballaud F, Dufresne A, Francez AJ, Colombet J, Sime-Ngando T, Quaiser A 2016. Dynamics of viral abundance and diversity in a sphagnum-dominated peatland: temporal fluctuations prevail over habitat. *Frontiers in Microbiology* 6: 1494.
- Banks JC, Demetras NJ, Hogg ID, Knox MA, West DW 2016. Monitoring brown trout (*Salmo trutta*) eradication in a wildlife sanctuary using environmental DNA. *New Zealand Natural Sciences* 41: 1–13.
- Bao Z, Ikonaga Y, Matsushita Y, Morimoto S, Takada-Hoshino Y, Okada H, Oba H, Takemoto S, Niwa S, Ohigashi K, Suzuki C, Nagaoka K, Takenaka M, Urashima Y, Sekiguchi H, Kushida A, Toyota K, Saito M, Tsushima S 2012. Combined analyses of bacterial, fungal and nematode communities in andosolic agricultural soils in Japan. *Microbes Environ* 27: 72–79.
- Barberan A, Casamayor EO 2014. A phylogenetic perspective on species diversity, beta-diversity and biogeography for the microbial world. *Molecular Ecology* 23: 5868–5876.
- Barberan A, Ramirez KS, Leff JW, Bradford MA, Wall DH, Fierer N 2014. Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. *Ecology Letters* 17: 794–802.
- Barberan A, Ladau J, Leff JW, Pollard KS, Menninger HL, Dunn RR, Fierer N 2015. Continental-scale distributions of dust-associated bacteria and fungi. *Proceedings of the National Academy of Science of the United States of America* 112: 5756–5761.
- Barrett T, Clark K, Gevorgyan R, Gorelenkov V, Gribov E, Karsch-Mizrachi I, Kimelman M, Pruitt KD, Resenchuk S, Tatusova T, Yaschenko E, Ostell J 2012. BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata. *Nucleic Acids Research* 40: D57–D63.
- Bates ST, Clemente JC, Flores GE, Walters WA, Parfrey LW, Knight R, Fierer N 2013. Global biogeography of highly diverse protistan communities in soil. *The ISME Journal* 7: 652–659.
- Bazzicalupo AL, Balint M, Schmitt I 2013. Comparison of ITS1 and ITS2 rDNA in 454 sequencing of hyperdiverse fungal communities. *Fungal Ecology* 6: 102–109.
- Beja O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich S, Gates CM, Feldman RA, Spudich JL, Spudich EN, DeLong EF 2000. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* 289: 1902–1906.
- Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kausarud H 2010. ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC Microbiology* 10: 189.
- Bellvert J, Crombie K, Horgan FG 2008. Comparative efficiency of the Fenwick Can and Schuiling Centrifuge in extracting nematode cysts from different soil types. *Journal of Nematology* 40: 30–34.
- Bhadury P, Austen MC 2010. Barcoding marine nematodes: an improved set of nematode 18S rRNA primers to overcome eukaryotic co-interference. *Hydrobiologia* 641: 245–251.
- Bibby K 2013. Metagenomic identification of viral pathogens. *Trends in Biotechnology* 31: 11–15.
- Bienert F, De Danieli S, Miquel C, Coissac E, Poillot C, Brun J-J, Taberlet P 2012. Tracking earthworm communities from soil DNA. *Molecular Ecology* 21: 2017–2030.
- Biggs J, Ewald N, Valentini A, Gaboriaud C, Dejean T, Griffiths RA, Foster J, Wilkinson JW, Arnella, Brotherton P, Williams P, Dunn F 2015. Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (*Triturus cristatus*). *Biological Conservation* 183: 19–28.
- Blouin AG, Ross HA, Hobson-Peters J, O'Brien CA, Warren B, MacDiarmid R 2016. A new virus discovered by immunocapture of double-stranded RNA, a rapid method for virus enrichment in metagenomic studies. *Molecular Ecology Resources* 16: 1255–1263.
- Boessenkool S, Epp LS, Haile J, Bellemain E, Edwards M, Coissac E, Willerslev E, Brochmann C 2012. Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. *Molecular Ecology* 21: 1806–1815.
- Bonito G, Reynolds H, Robeson MS, Nelson J, Hodkinson BP, Tuskan G, Schadt CW, Vilgalys R 2014. Plant host and soil origin influence fungal and bacterial assemblages in the roots of woody plants. *Molecular Ecology* 23: 3356–3370.
- Bothwell ML, Taylor BW, Kilroy C 2014. The didymo story: the role of low dissolved phosphorus in the formation of *Didymosphenia geminata* blooms. *Diatom Research* 29: 229–236.
- Boyer S, Brown SDJ, Collins RA, Cruickshank RH, Lefort M-C, Malumbres-Olarte J, Wratten SD 2012. Sliding window analyses for optimal selection of mini-barcodes, and application to 454-pyrosequencing for specimen identification from degraded DNA. *PLoS ONE* 7: e0038215.
- Boyer S, Wratten SD, Holyoake A, Abdelkrim J, Cruickshank RH 2013. Using next-generation sequencing to analyse the diet of a highly endangered land snail *Powelliphanta augusta* feeding on endemic earthworms. *PLoS ONE* 8: e75962.
- Boyer S, Cruickshank RH, Wratten SD 2015. Faeces of generalist predators as ‘biodiversity capsules’: a new tool for biodiversity assessment in remote and inaccessible habitats. *Food Webs* 3: 1–6.
- Bradford TM, Morgan MJ, Lorenz Z, Hartley DM, Hardy CM, Oliver RL 2013. Microeukaryote community composition assessed by pyrosequencing is associated with light availability and phytoplankton primary production along a lowland river. *Freshwater Biology* 58: 2401–2413.
- Bragg LM, Stone G, Butler MK, Hugenholtz P, Tyson GW 2013. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS Computational Biology* 9: e1003031.
- Brandariz-Fontes C, Camacho-Sanchez M, Vila C, Vega-Pla JL, Rico C, Leonard JA 2015. Effect of the enzyme and PCR conditions on the quality of high-throughput DNA sequencing results. *Scientific Reports*: article 8056.
- Brandon-Mong G-J, Gan H-M, Sing K-W, Lee P-S, Lim P-E, Wilson J-J 2015. DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. *Bulletin of Entomological Research* 105: 717–727.
- Brown BL, Watson M, Minot SS, Rivera MC, Franklin RB 2017. MinION™ nanopore sequencing of environmental metagenomes: a synthetic approach. *Gigascience* 6: 1–10.

- Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, Wilkins MJ, Wrighton KC, Williams KH, Banfield JF 2015. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* 523: 208–U173.
- Brown SDJ, Collins RA, Boyer S, Lefort M-C, Malumbres-Olarte J, Vink CJ, Cruickshank RH 2012. Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources* 12: 562–565.
- Brown SP, Callahan MA, Oliver AK Jr, Jumpponen A 2013. Deep Ion Torrent sequencing identifies soil fungal community shifts after frequent prescribed fires in a southeastern US forest ecosystem. *FEMS Microbiol Ecol* 86: 557–566.
- Brundrett MC 2009. Mycorrhizal associations and other means of nutrition of vascular plants: understanding the global diversity of host plants by resolving conflicting information and developing reliable means of diagnosis. *Plant and Soil* 320: 37–77.
- Brussow H, Canchaya C, Hardt WD 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiology and Molecular Reviews* 68: 560–602.
- Buchanan PK, Beever RE, Galloway DJ, Glare TR, Johnston PR, McKenzie EHC, di Menna ME, Pennycook SR, Ridley GS, Smith JMB, Stephenson SL 2004. The fungi of New Zealand - an introduction. In: McKenzie EHC ed. *Introduction to fungi of New Zealand*. Lincoln, Manaaki Whenua Press. Pp. 1–48.
- Buermans HPJ, den Dunnen JT 2014. Next generation sequencing technology: advances and applications. *Biochimica et Biophysica Acta (BBA) – Molecular Basis of Disease* 1842: 1932–1941.
- Bulat SA, Lübeck M, Mironenko N, Jensen DF, Lübeck PS 1998. UP-PCR analysis and ITS1 ribotyping of strains of *Trichoderma* and *Gliocladium*. *Mycological Research* 102: 933–943.
- Bybee SM, Bracken-Grissom H, Haynes BD, Hermansen RA, Byers RL, Clement MJ, Udall JA, Wilcox ER, Crandall KA 2011. Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. *Genome Biology and Evolution* 3: 1312–1323.
- Cai P, Huang Q, Zhang X, Chen H 2006. Adsorption of DNA on clay minerals and various colloidal particles from an Alfisol. *Soil Biology and Biochemistry* 38: 471–476.
- Caley MJ, Fisher R, Mengersen K 2014. Global species richness estimates have not converged. *Trends in Ecology and Evolution* 29: 187–188.
- Callejas C, Gill PR, Catalan AI, Azziz G, Castro-Sowinski S, Batista S 2011. Phylotype diversity in a benthic cyanobacterial mat community on King George Island, maritime Antarctica. *World Journal of Microbiology and Biotechnology* 27: 1507–1512.
- Calvignac-Spencer S, Merkel K, Kutzer N, Kuhl H, Boesch C, Kappeler PM, Metzger S, Schubert G, Leendertz FH 2013. Carrion fly-derived DNA as a tool for comprehensive and cost effective assessment of mammalian biodiversity. *Molecular Ecology Resources* 22: 915–924.
- Campos PF, Gilbert TMP 2012. DNA extraction from keratin and chitin. In: Shapiro B, Hofreiter M eds. *Ancient DNA: methods and protocols*. Totowa, NJ, Humana Press. Pp. 43–49.
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Luzupone CA, Turnbaugh PJ, Fierer N, Knight R 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Science of the United States of America* 2011: 4516–4522.
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal* 6: 1621–1624.
- Carroll IM, Ringel-Kulka T, Siddle JP, Klaenhammer TR, Ringel Y 2012. Characterization of the fecal microbiota using high-throughput sequencing reveals a stable microbial community during storage. *PLoS ONE* 7: e46953.
- Carroll NM, Adamson P, Okhravi N 1999. Elimination of bacterial DNA from Taq DNA polymerases by restriction endonuclease digestion. *Journal of Clinical Microbiology* 37: 3402–3404.
- Cary SC, Fierer N 2014. The importance of sample archiving in microbial ecology. *Nature Reviews Microbiology* 12: 790.
- Cavalier-Smith T 2010. Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. *Biology Letters* 6: 342–345.
- Chariton AA, Court LN, Hartley DM, Colloff MJ, Hardy CM 2010. Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. *Frontiers in Ecology and the Environment* 8: 233–238.
- Chen CW, Huang YM, Kuo LY, Nguyen QD, Luu HT, Callado JR, Farrar DR, Chiou WL 2013. trnL-F is a powerful marker for DNA identification of field vittarioid gemetophytes (Pteridaceae). *Annals of Botany* 111: 663–673.
- Chen H, Rangasamy M, Tan SY, Wang H, Siegfried BD 2010a. Evaluation of five methods for total DNA extraction from western corn rootworm beetles. *PLoS ONE* 5: e11963.
- Chen SL, Yao H, Han JP, Liu C, Song JY, Shi LC, Zhu YJ, Ma XY, Gao T, Pang XH, Luo K, Li Y, Li XW, Jia XC, Lin YL, Leon C 2010b. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE* 5: e8613.
- Chew YV, Holmes AJ 2009. Suppression subtractive hybridisation allows selective sampling of metagenomic subsets of interest. *Journal of Microbiological Methods* 78: 136–143.
- Chibani-Chennoufi S, Bruttin A, Dillmann ML, Brussow H 2004. Phage-host interaction: an ecological perspective. *Journal of Bacteriology* 186: 3677–3686.
- Chomczynski P 1992. Solubilization in formamide protects RNA from degradation. *Nucleic Acids Research* 20: 3791–3792.
- Chu H, Wang CY, Wang HH, Chen H, Tang M 2016. Pine wilt disease alters soil properties and root associated communities in *Pinus tabulaeformis* forest. *Plant and Soil* 404: 237–247.
- Clare EL, Symondson WOC, Fenton MB 2014. An inordinate fondness for beetles? Variation in seasonal dietary preferences of night-roosting big brown bats (*Eptesicus fuscus*). *Molecular Ecology* 23: 3633–3647.
- Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW 2016. GenBank. *Nucleic Acids Research* 44: D67–D72.
- Clarke LJ, Czechowski P, Soubrier J, Stevens MI, Cooper A 2014. Modular tagging of amplicons using a single PCR for high-throughput sequencing. *Molecular Ecology Resources* 14: 117–121.

- Coince A, Cael O, Bach C, Lengelle J, Cruaud C, Gavory F, Morin E, Murat C, Marcais B, Buee M 2013. Below-ground fine-scale distribution and soil versus fine root detection of fungal and soil oomycete communities in a French beech forest. *Fungal Ecology* 6: 223–235.
- Collins RA, Armstrong KF, Holyoake AJ, Keeling S 2013. Something in the water: biosecurity monitoring of ornamental fish imports using environmental DNA. *Biological Invasions* 15: 1209–1215.
- Cone RW, Fairfax MR 1993. Protocol for ultraviolet irradiation of surfaces to reduce PCR contamination. *Genome Research* 3.3: S15–S17.
- Cooper A, Poiner HN 2000. Ancient DNA: do it right or not at all. *Science* 289: 1139.
- Corliss JO 2004. Why the world needs protists! *Journal of Eukaryotic Microbiology* 51: 8–22.
- Cornelisen C, Kirs M, Gilpin B, Scholes P 2012. Microbial source tracking (MST) tools for water quality monitoring. Report Number 2047, Regional Councils and the Coastal Special Interest Group. Nelson, New Zealand, Cawthron Institute. 28 p.
- Costa PS, Reis MP, Avila MP, Leite LR, de Araujo FMG, Salim ACM, Oliveira G, Barbosa F, Chartone-Souza E, Nascimento AMA 2015. Metagenome of a microbial community inhabiting a metal-rich tropical stream sediment. *PLoS ONE* 10: e0119465.
- Cotten M, Oude Munnink B, Canuti M, Deijs M, Watson SJ, Kellham P, van der Hoek L 2014. Full genome virus detection in fecal samples using sensitive nucleic acid preparation, deep sequencing, and a novel iterative sequence classification algorithm. *PLoS ONE* 9: 93269.
- Cottrell MT, Kirchman DL 2012. Virus genes in Arctic marine bacteria identified by metagenomic analysis. *Aquatic Microbial Ecology* 66: 107–116.
- Cowart DA, Pinheiro M, Mouchel O, Maguer M, Grall J, Mine J, Arnaud-Haond S 2015. Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. *PLoS ONE* 10: e0117562.
- Creer S, Fonseca VG, Porazinska DL, Giblin-Davis RM, Sung W, Power DM, Packer M, Carvalho GR, Blaxter ML, Lamshead PJD, Thomas WK 2010. Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Molecular Ecology* 19: 4–20.
- Creer S, Deiner K, Frey S, Porazinska D, Taberlet P, Thomas WK, Potter C, Bik HM 2016. The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution* 7: 1008–1018.
- Cui X, Hu J, Wang J, Yang J, Lin X 2016. Reclamation negatively influences arbuscular mycorrhizal fungal community structure and diversity in coastal saline-alkaline land in Eastern China as revealed by Illumina sequencing. *Applied Soil Ecology* 98: 140–149.
- Davison J, Moora M, Oepik M, Adholeya A, Ainsaar L, Ba A, Burla S, Diedhiou AG, Hiiesalu I, Jairus T, Johnson NC, Kane A, Koorem K, Kochar M, Ndiaye C, Paertel M, Reier U, Saks U, Singh R, Vasar M, Zobel M 2015. Fungal symbionts: global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science* 349: 970–973.
- Deagle BE, Kirkwood R, Jarman SN 2009. Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Molecular Ecology* 18: 2022–2038.
- Deagle BE, Thomas AC, Shaffer AK, Trites AW, Jarman SN 2013. Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: which counts count? *Molecular Ecology Resources* 13: 620–633.
- Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters* 10: article20140562.
- Deiner K, Walser J-C, Maechler E, Altermatt F 2015. Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation* 183: 53–63.
- DeLeon-Rodriguez N, Latham TL, Rodriguez-RLM, Barazesh JM, Anderson BE, Beyersdorf AJ, Ziemba LD, Bergin M, Nenes A, Konstantinidia KT 2012. Microbiome of the upper trophosphere: species composition and prevalence, effects of tropical storms, and atmospheric implications. *Proceedings of the National Academy of Science of the United States of America* 110: 2575–2580.
- Delmont TO, Robe P, Clark I, Simonet P, Vogel TM 2011. Metagenomic comparison of direct and indirect soil DNA extraction approaches. *Journal of Microbiological Methods* 86: 397–400.
- Deng C, Daley T, Smith AD 2015. Applications of species accumulation curves in large-scale biological data analysis. *Quantitative biology (Beijing, China)* 3: 135–144.
- Deslippe JR, Hartmann M, Grayston SJ, Simard SW, Mohn WW 2016. Stable isotope probing implicates a species of *Cortinarius* in carbon transfer through ectomycorrhizal fungal mycelial networks in Arctic tundra. *New Phytologist* 210: 383–390.
- Diaz S, Purvis A, Cornelissen JHC, Mace GM, Donoghue MJ, Wewers RM, Jordano P, Pearse WD 2013. Functional traits, the phylogeny of function, and ecosystem service vulnerability. *Ecology and Evolution* 3: 2958–2975.
- Dichosa AEK, Daughton AR, Reitenga KG, Fitzsimons MS, Han CS 2014. Capturing and cultivating single bacterial cells in gel microdroplets to obtain near-complete genomes. *Nature Protocols* 9: 608–621.
- Dickie IA, St John MG 2016. Second generation molecular understanding of mycorrhizas in soil ecosystems. In: Martin F ed. *Molecular mycorrhizal symbiosis*. Hoboken, New Jersey, USA, Wiley-Blackwell. Pp. 473–493.
- Dineen SM, Aranda R, Anders DL, Robertson JM 2010. An evaluation of commercial DNA extraction kits for the isolation of bacterial spore DNA from soil. *Journal of Applied Microbiology* 109: 1886–1896.
- Dong C, Bai X, Sheng H, Jiao L, Zhou H, Shao Z 2015. Distribution of PAHs and the PAH-degrading bacteria in the deep-sea sediments of the high-latitude Arctic Ocean. *Biogeosciences* 12: 2163–2177.
- Dong D, Yan A, Liu H, Zhang X, Xu Y 2006. Removal of humic substances from soil DNA using aluminium sulfate. *Journal of Microbiological Methods* 66: 217–222.
- Dozois CM, Daigle F, Curtiss III R 2003. Identification of pathogen-specific and conserved genes expressed in vivo by an avian pathogenic *Escherichia coli* strain. *Proceedings of the National Academy of Science of the United States of America* 100: 247–252.
- Drummond AJ, Newcomb RD, Buckley TR, Xie D, Dopheide A, Potter BCM, Heled J, Ross HA, Tooman L, Grosser S, Park D, Demetras NJ, Stevens MI, Russell JC, Anderson SH, Carter A, Nelson N 2015. Evaluating a multigene environmental DNA approach for biodiversity assessment. *Gigascience* 4: 46.

- Dumbrell AJ, Gu F, Ashton PD, Aziz N, Nelson M, Dytham C, Fitter AH, Hargason T 2011. Distinct seasonal assemblages of arbuscular mycorrhizal fungi revealed by massively parallel pyrosequencing. *New Phytologist* 190: 794–804.
- Eichmiller JJ, Bajer PG, Sorensen PW 2014. The relationship between the distribution of common carp and their environmental DNA in a small lake. *PLoS ONE* 9.
- Eichmiller JJ, Miller LM, Sorensen PW 2016. Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. *Molecular Ecology Resources* 16: 56–68.
- Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, deWinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323: 133–138.
- Elbrecht V, Leese F 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. *PLoS ONE* 10: e0130324.
- Epp L, Boessenkool S, Bellemain EP, Haile J, Esposito A, Riaz T, Erseus C, Gusarov VI, Edwards ME, Johnsen A, Stenoien HK, Hassel K, Kausarud H, Yoccoz NG, Brathan KA, Willerslev E, Taberlet P, Coissac E, Brochmann C 2012. New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Molecular Ecology* 21: 1821–1833.
- Escobar-Zepeda A, Vera-Ponce de León A, Sanchez-Flores A 2015. The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics. *Frontiers in Genetics* 6: 348.
- Esling P, Lejzerowicz F, Pawlowski J 2015. Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Research* 43: 2513–2524.
- Evans NT, Olds BP, Renshaw MA, Turner CR, Li Y, Jerde CL, Mahon AR, Pfrender ME, Lamberti GA, Lodge DM 2016. Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources* 16: 29–41.
- Faegri A, Torsvik VL, Goksoyr J 1977. Bacterial and fungal activities in soil: separation of bacteria and fungi by a rapid fractionated centrifugation technique. *Soil Biology and Biochemistry* 9: 105–112.
- Fahner NA, Shokralla S, Baird DJ, Hajibabaei M 2016. Large-scale monitoring of plants through environmental DNA metabarcoding of soil: recovery, resolution and annotation of four DNA markers. *PLoS ONE*: e0157505.
- Farias IP, Orti G, Sampaio I, Schneider H, Meyer A 2001. The cytochrome b gene as a phylogenetic marker: the limits of resolution for analysing relationships among cichlid fishes. *Journal of Molecular Evolution* 53: 89–103.
- Farrington HL, Edwards CE, Guan X, Carr MR, Baerwaldt K, Lance RF 2015. Mitochondrial genome sequencing and development of genetic markers for the detection of DNA of invasive bighead and silver carp (*Hypophthalmichthys nobilis* and *H. molitrix*) in environmental water samples from the United States. *PLoS ONE* 10.
- Ficotela GE, Pansu J, Bonin A, Coissac E, Giguët-Covex C, De Barba M, Gielly L, Lopes CM, Boyer F, Raye FPG, Taberlet P 2015. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources* 15: 543–556.
- Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R, Robeson M, Edwards RA, Felts B, Rayhawk S, Knight R, Rohwer F, Jackson RB 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi and viruses in soil. *Applied and Environmental Microbiology* 73: 7059–7066.
- Fierer N, Liu Z, Rodriguez-Hernandez M, Knight R, Henn M, Hernandez MT 2008. Short-term temporal variability in airborne bacterial and fungal populations. *Applied and Environmental Microbiology* 74: 200–207.
- Fierer N, Ladau J, Clemente JC, Leff JW, Owens SM, Pollard KS, Knight R, Gilbert JA, McCulley RL 2013. Reconstructing the microbial diversity and function of pre-agricultural tallgrass soils in the United States. *Science* 342: 621–624.
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* 3: 294–299.
- Fonseca VG, Carvalho GR, Sung W, Johnson HF, Power DM, Neill SP, Packer M, Blaxter ML, Lamshead PJD, Thomas WK, Creer S 2010. Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature Communications* 1: 98.
- Forsberg KJ, Reyes A, Wang B, Selleck EM, Sommer MOA, Dantas G 2012. The shared antibiotic resistome of soil bacteria and human pathogens. *Science* 337: 1107–1111.
- Frostegard A, Courtois S, Ramišse V, Clerc S, Bernillon D, Le Gall F, Jeannin P, Nesme X, Simonet P 1999. Quantification of bias related to the extraction of DNA directly from soils. *Applied and Environmental Microbiology* 65: 5409–5420.
- Fuhrman JA 1999. Marine viruses and their biogeochemical and ecological effects. *Nature* 399: 541–548.
- Fukumoto S, Ushimaru A, Minamoto T 2015. A basin-scale application of environmental DNA assessment for rare endemic species and closely related exotic species in rivers: a case study of giant salamanders in Japan. *Journal of Applied Ecology* 52: 358–365.
- Furlan EM, Gleeson D, Hardy CM, Duncan RP 2016. A framework for estimating the sensitivity of eDNA surveys. *Molecular Ecology Resources* 16: 641–654.
- Gardes M, Bruns TD 1993. ITS primers with enhanced specificity for basidiomycetes – application to the identification of mycorrhizae and rusts. *Molecular Ecology* 2: 113–118.
- Garrett TR, Bhakoo M, Zhang Z 2008. Bacterial adhesion and biofilms on surfaces. *Progress in Natural Science* 18: 1049–1056.
- Gasser RB, Chilton NB, Hoste H, Beveridge I 1993. Rapid sequencing of rDNA from single worms and eggs of parasitic helminths. *Nucleic Acids Research* 21: 2525–2526.
- Geisen S, Tveit AT, Clark IM, Richter A, Svenning MM, Bonkowski M, Ulrich T 2015. Metatranscriptomic census of active protists in soils. *The ISME Journal* 9: 2178–2190.
- Geller J, Meyer C, Parker M, Hawk H 2013. Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I

- for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources* 13: 851–861.
- Geurs M, Bongers J, Brussard L 1991. Improvements to the heptane flotation method for collecting microarthropods from silt loam soil. *Agriculture, Ecosystems & Environment* 34: 213–221.
- Gianinazzi S, Gollotte A, Binet M-N, van Tuinen D, Redecker D, Wipf D 2010. Agroecology: the key role of arbuscular mycorrhizas in ecosystem services. *Mycorrhiza* 20: 519–530.
- Gilbert JA, Jansson JK, Knight R 2014. The Earth Microbiome Project: successes and aspirations. *BMC Biology* 12: 69.
- Gilpin BJ, Devane M, Nourozi F, Robson B, Scholes P, Lin S 2013. Recommendations for the processing and storage of water samples before polymerase chain reaction (PCR) analysis. *New Zealand Journal of Marine and Freshwater Research* 47: 582–586.
- Glenn TC 2011. Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* 11: 759–769.
- 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, Sepulveda A, Ray A, Baumgardt J, Waits LP 2013. Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*). *Freshwater Science* 32: 792–800.
- Goldberg CS, Turner CR, Deiner K, Klymus KE, Thomsen PF, Murphy MA, Spear SF, McKee A, Oyler-McCance SJ, Cornman RS, Laramie MB, Mahon AR, Lance RF, Pilliod DS, Strickler KM, Waits LP, Fremier AK, Takahara T, Herder JE, Taberlet P 2016. Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution* 7: 1299–1307.
- Goodwin S, McPherson JD, McCombie WR 2016. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* 17: 333–351.
- Griffin MJ, Goodwin AE, Merry GE, Liles MR, Williams MA, Ware C, Waldbieser GC 2013. Rapid quantitative detection of *Aeromonas hydrophila* strains associated with disease outbreaks in catfish aquaculture. *Journal of Veterinary Diagnostic Investigation* 25: 473–481.
- Griffiths RI, Thomson BC, James P, Bell T, Bailey MJ, Whiteley AS 2011. The bacterial biogeography of British soils. *Environmental Microbiology* 13: 1642–1654.
- Guo LD, Hyde KD, Liew ECY 2001. Detection and taxonomic placement of endophytic fungi within frond tissues of *Livistona chinensis* based on rDNA sequences. *Molecular Phylogenetics and Evolution* 20: 1–13.
- Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, Methe B, DeSantis TZ, Petrosino JF, Knight R, Birren BW, Human Microbiome C 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research* 21: 494–504.
- Hadziavdic K, Lekang K, Lanzen A, Jonassen I, Thompson EM, Troedsson C 2014. Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PLoS ONE* 9: e87624.
- Haile J 2012. Ancient DNA extraction from soils and sediments. In: Shapiro B, Hofreiter M eds. *Ancient DNA: methods and protocols*. New York, USA. Humana Press. Pp. 57–63.
- Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE* 6: e17497.
- Hämäläinen A, Broadley K, Droghini A, Haines JA, Lamb CT, Butin S, Gilbert S 2017. The ecological significance of secondary seed dispersal by carnivores. *Ecosphere* 82: e01695.
- Hamdan LJ, Gillevet PM, Pohlman JW, Sikaroodi M, Greinert J, Coffin RB 2011. Diversity and biogeochemical structuring of bacterial communities across the Porangahau ridge accretionary prism, New Zealand. *FEMS Microbiology Ecology* 77: 518–532.
- Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry and Biology* 5: R245–249.
- Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, Causey M, Colonell J, Dimeo J, Efcavitch JW, Giladi E, Gill J, Healy J, Jarosz M, Lapen D, Moulton K, Quake SR, Steinmann K, Thayer E, Tyurina A, Ward R, Weiss H, Xie Z 2008. Single-molecule DNA sequencing of a viral genome. *Science* 320: 106–109.
- Hart MM, Aleklett K, Chagnon P-L, Egan C, Ghignone S, Helgason T, Lekberg Y, Oepik M, Pickles BJ, Waller L 2015. Navigating the labyrinth: a guide to sequence-based, community ecology of arbuscular mycorrhizal fungi. *New Phytologist* 207: 235–247.
- Hebert PDN, Gregory TR 2005. The promise of DNA barcoding for taxonomy. *Systematic Biology* 54: 852–859.
- Henry P, Russello MA 2011. Obtaining high-quality DNA from illusive small mammals using low-tech hair snares. *European Journal of Wildlife Research* 57: 429–435.
- Hermans SM, Buckley HL, Case BS, Curran-Cournane F, Taylor M, Lear G 2017. Bacteria as emerging indicators of soil condition. *Applied and Environmental Microbiology* 83: 10.1128/AEM.02826–02816.
- Hibert F, Taberlet P, Chave J, Scotti-Saintagne C, Sabatier D, Richard-Hansen C 2013. Unveiling the diet of elusive rainforest herbivores in next generation sequencing era? The tapir as a case study. *PLoS ONE* 8: e60799.
- Highlander S 2013. Mock community analysis. In: Nelson EK ed. *Encyclopedia of metagenomics*. New York, Springer New York. Pp. 1–7.
- Hiiesalu J, Oepik M, Metsis M, Lilje L, Davison J, Vasar M, Moora M, Zobel M, Wilson SD, Paertel M 2012. Plant species richness belowground: higher richness and new patterns revealed by next-generation sequencing. *Molecular Ecology* 21: 2004–2016.
- Holdaway RJ, Wood JR, Dickie IA, Orwin KH, Bellingham PJ, Richardson SJ, Lyver POB, Timoti P, Buckley TR 2017. Using DNA metabarcoding to assess New Zealand's terrestrial biodiversity. *New Zealand Journal of Ecology* 41: 251–262.
- Hoorfar J, Malorny B, Abdulmawjood A, Cook N, Wagner M, Fach P 2004. Practical considerations in design of internal amplification controls for diagnostic PCR assays. *Journal of Clinical Microbiology* 42: 1863–1868.
- Hu J, Johnson EG, Wang N-Y, Davoglio T, Dewdney MM 2014. qPCR quantification of pathogenic *Guignardia citricarpa* and non-pathogenic *G. mangiferae* in citrus. *Plant Disease* 98: 112–120.
- Huang Y-T, Lowe DJ, Zhang H, Cooper A 2016. A new method to extract and purify DNA from allophanic soils



- and paleosols, and potential for paleoenvironmental reconstruction and other applications. *Geoderma* 274: 114–125.
- Hug K, Maher WA, Stott MB, Krikowa F, Foster S, Moreau JW 2014. Microbial contributions to coupled arsenic and sulfur cycling in the acid-sulfide hot spring Champagne Pool, New Zealand. *Frontiers in Microbiology* 5: 569.
- Hunter ME, Oyler-McCance SJ, Dorazio RM, Fike JA, Smith BJ, Hunter CT, Reed RN, Hart KM 2015. Environmental DNA (eDNA) sampling improves occurrence and detection estimates of invasive Burmese pythons. *PLoS ONE* 10: e0121655.
- Ihrmark K, Bodeker ITM, Cruz-Martinez K, Friberg H, Kubartova A, Schenck J, Strid Y, Stenlid J, Brandstrom-Durling M, Clemmensen KE, Lindahl BD 2012. New primers to amplify the fungal ITS2 region – evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiology Ecology* 82: 666–677.
- Illumina. undated. 16S metagenomic sequencing library preparation: preparing 16S ribosomal RNA gene amplicons for the Illumina MiSeq system. Illumina Technical Document Part. No. 15044223 Rev. B. 28 p.
- Inskip WP, Jay ZJ, Herrgard MJ, Kozubal MA, Rusch DB, Tringe SG, Macur RE, Jennings RD, Boyd ES, Spear JR, Roberto FF 2013. Phylogenetic and functional analysis of metagenome sequence from high-temperature archaeal habitats demonstrate linkages between metabolic potential and geochemistry. *Frontiers in Microbiology* 4: article95.
- Ivanova NV, Kuzmina ML 2013. Protocols for DNA storage and shipment at room temperature. *Molecular Ecology Resources* 14: 890–898.
- Jain M, Olsen HE, Paten B, Akeson M 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology* 17: 239.
- Jangid K, Whitman WB, Condron LM, Turner BL, Williams MA 2013. Soil bacterial community succession during long-term ecosystem development. *Molecular Ecology* 22: 3415–3424.
- Janosik AM, Johnston CE 2015. Environmental DNA as an effective tool for detection of imperiled fishes. *Environmental Biology of Fishes* 98: 1889–1893.
- Jansson J 2011. Towards “Tera-Terra”: terabase sequencing of terrestrial metagenomes. *Microbe* 6: 309–315.
- Jerde CL, Mahon AR, Chadderton WL, Lodge DM 2011. “Sight-unseen” detection of rare aquatic species using environmental DNA. *Conservation Biology* 4: 150–157.
- Jerde CL, Chadderton WL, Mahon AR, Renshaw MA, Corush J, Budny ML, Mysorekar S, Lodge DM 2013. Detection of Asian carp DNA as part of a Great Lakes basin-wide surveillance program. *Canadian Journal of Fisheries and Aquatic Sciences* 70: 522–526.
- Ji Y, Ashton L, Pedley SM, Edwards DP, Tang Y, Nakamura A, Kitching R, Dolman PM, Woodcock P, Edwards FA, Larsen TH, Hsu WW, Benedick S, Hamer KC, Wilcove DS, Bruce C, Wang X, Levi T, Lott M, Emerson BC, Yu DW 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters* 16: 1245–1257.
- Jian JH, Lakshman DK, Tavantzis SM 1998. A virulence-associated, 6.4-kb, double-stranded RNA from *Rhizoctonia solani* is phylogenetically related to plant bromoviruses and electron transport enzymes. *Molecular Plant Microbe Interactions* 11: 601–609.
- Jiang J, Alderisio KA, Singh A, Xiao L 2005. Development of procedures for direct extraction of *Cryptosporidium* DNA from water concentrates and for relief of PCR inhibitors. *Applied and Environmental Microbiology* 71: 1135–1141.
- Johansen RB, Vestberg M, Burns BR, Park D, Hooker JE, Johnston PR 2015. A coastal sand dune in New Zealand reveals high arbuscular mycorrhizal fungal diversity. *Symbiosis* 66: 111–121.
- Johnston PR 2010. Causes and consequences of changes to New Zealand’s fungal biota. *New Zealand Journal of Ecology* 34: 175–184.
- Kanokratana P, Uengwetwanit T, Rattanachomsri U, Bunternngsook B, Nimchua T, Tangphatsornruang S, Plengvidhya V, Champreda V, Eurwilaichitr L 2011. Insights into the phylogeny and metabolic potential of a primary tropical peat swamp forest microbial community by metagenomic analysis. *Microbial Ecology* 61: 518–528.
- Kantor RS, Wrighton KC, Handley KM, Sharon I, Hug LA, Castelle CJ, Thomas BC, Banfield JF 2013. Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. *Mbio* 4: e00708–00713.
- Kanzaki N, Giblin-Davis RM, Scheffrahn RH, Taki H, Esquivel A, Davies KA, Herre EA 2012. Reverse taxonomy for elucidating diversity of insect-associated nematodes: a case study with termites. *PLoS ONE* 7: e43865.
- Kartavtsev YP 2011. Divergence at Cyt-b and Co-1 mtDNA genes on different taxonomic levels and genetics of speciation in animals. *Mitochondrial DNA* 22: 55–65.
- Kartzinel TR, Pringle RM 2015. Molecular detection of invertebrate prey in vertebrate diets: trophic ecology of Caribbean island lizards. *Molecular Ecology Resources* 15: 903–914.
- Kasahara T, Miyazaki T, Nitta H, Ono A, Miyagishima T, Nagao T, Urushidani T 2006. Evaluation of methods for duration of preservation of RNA quality in rat liver used for transcriptome analysis. *The Journal of Toxicological Sciences* 31: 509–519.
- Kelly RP, Port JA, Yamahara KM, Crowder LB 2014. Using environmental DNA to census marine fishes in a large mesocosm. *PLoS ONE* 9: e86175.
- Keskin E 2014. Detection of invasive freshwater fish species using environmental DNA survey. *Biochemical Systematics and Ecology* 56: 68–74.
- Kim K-H, Bae J-W 2011. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Applied and Environmental Microbiology* 77: 7663–7668.
- Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glockner FO 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research* 41: e1.
- Klymus KE, Marshall NT, Stepien CA 2010. Environmental DNA (eDNA) metabarcoding assays to detect invasive invertebrate species in the Great Lakes. *PLoS ONE* 12: e0177643.
- Knapp M, Clarke AC, Ann Horsburgh K, Matisoo-Smith EA 2012a. Setting the scene – building and working in an ancient DNA laboratory. *Annals of Anatomy* 194: 3–6.
- Knapp M, Stiller M, Meyer M 2012b. Generating barcoded libraries for multiplex high-throughput sequencing. In: Shapiro B, Hofreiter M eds. *Methods in molecular biology*. New York, USA, Humana Press. Pp. 155–170.
- Kochzius M, Seidel C, Antoniou A, Botla SK, Campo D, Cariani A, Vazquez EG, Hauschild J, Hervet C, Hjørleifsdottir

- S, Hreggvidsson G, Kappel K, Landi M, Magoulas A, Marteinson V, Nölte M, Planes S, Tinti F, Turan C, Venugopal MN, Weber H, Blohm D 2010. Identifying fishes through DNA barcodes and microarrays. *PLoS ONE* 5: e12620.
- Koele N, Dickie IA, Blum JD, Gleason JD, de Graaf L 2014. Ecological significance of mineral weathering in ectomycorrhizal and arbuscular mycorrhizal ecosystems from a field-based comparison. *Soil Biology and Biochemistry* 69: 63–70.
- Koester M, Claßen S, Gergs R 2013. Establishment of group-specific PCR primers for the identification of freshwater macroinvertebrates. *Conservation Genetics Resources* 5: 1091–1093.
- Konstantinidis KT, Tiedje JM 2005. Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Science of the United States of America* 102: 2567–2572.
- Kotchoni SO, Gachomo EW, Jimenez-Lopez JC 2011. A plant cocktail amenable for PCR-based genetic analysis in *Arabidopsis thaliana*. *Molecular Biology Reports* 38: 5281–5284.
- Kowallik V, Miller E, Greig D 2015. The interaction of *Saccharomyces paradoxus* with its natural competitors on oak bark. *Molecular Ecology* 24: 1596–1610.
- Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology* 79: 5112–5120.
- Kuch M, Rohland N, Betancourt JL, Latorre C, Stepan S, Poinar HN 2002. Molecular analysis of an 11,700-year-old rodent midden from the Atacama Desert, Chile. *Molecular Ecology* 11: 913–924.
- Kunin V, Copeland A, Lapidus A, Mavromatix K, Hugenholtz P 2008. A bioinformatician's guide to metagenomics. *Microbiology and Molecular Reviews* 72: 557–578.
- Kurose N, Masuda R, Tatara M 2005. Fecal DNA analysis for identifying species and sex of sympatric carnivores: a noninvasive method for conservation on the Tsushima Islands, Japan. *Journal of Heredity* 96: 688–697.
- Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Science of the United States of America* 82: 6955–6959.
- Lang AS, Kruijjer JD, Stech M 2014. DNA barcoding of Arctic bryophytes: an example from the moss genus *Dicranum* (Dicranaceae, Bryophyta). *Polar Biology* 37: 1157–1169.
- Laramie MB, Pilliod DS, Goldberg CS 2015. Characterizing the distribution of an endangered salmonid using environmental DNA analysis. *Biological Conservation* 183: 29–37.
- Laroche O, Wood SA, Tremblay LA, Ellis JI, Lejzerowicz F, Pawlowski J, Lear G, Atalah J, Pochon X 2016. First evaluation of foraminiferal metabarcoding for monitoring environmental impact from an offshore oil drilling site. *Marine Environmental Research* 120: 225–235.
- Lasken RS, Stockwell TB 2007. Mechanism of chimera formation during multiple displacement amplification reaction. *BMC Biotechnology* 7: 19.
- Lau KEM, Washington VJ, Fan V, Neale MW, Lear G, Curran J, Lewis GD 2015. A novel bacterial community index to assess stream ecological health. *Freshwater Biology* 60: 1988–2002.
- Lauber CL, Hamady M, Knight R, Fierer N 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and Environmental Microbiology* 75: 5111–5120.
- Lauber CL, Zhou N, Gordon JI, Knight R, Fierer N 2010. Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiology Letters* 307: 80–86.
- Laver T, Harrison J, O'Neill PA, Moore K, RFarbos A, Paszkiewicz K, Studholme DJ 2015. Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomolecular Detection and Quantification* 31: 1–8.
- Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM, Gibbons SM, Larsen P, Shogan BD, Weiss S, Metcalf JL, Ursell LK, Vázquez-Baeza Y, Van Treuren W, Hasan NA, Gibson MK, Colwell R, Dantas G, Knight R, Gilbert JA 2014. Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science* 345: 1048–1052.
- Lazarus KL, James TY 2015. Surveying the biodiversity of the Cryptomycota using a targeted PCR approach. *Fungal Ecology* 14: 62–70.
- Lear G, Niyogi D, Harding J, Dong Y, Lewis GD 2009. Biofilm bacterial community structure in streams affected by acid mine drainage. *Applied and Environmental Microbiology* 75: 3455–3460.
- Lear G, Washington V, Neale MW, Case B, Buckley HL, Lewis GD 2013. The biogeography of stream bacteria. *Global Ecology and Biogeography* 22: 544–554.
- Lear G, Bellamy J, Case B, Lee JE, Buckley HL 2014. Fine-scale patterns in bacterial community composition and function within freshwater ponds. *The ISME Journal* 8: 1715–1726.
- Lee J, Lee S, Young JPW 2008. Improved PCR primers for the detection and identification of arbuscular mycorrhizal fungi. *FEMS Microbiology Ecology* 65: 339–349.
- Lee SB, Clabaugh KC, Silva B, Odigie KO, Coble MD, Loreille O, Scheible M, Fournay RM, Stevens J, Carmody GR, Parsons TJ, Pozder A, Eisenberg AJ, Budowle B, Ahmad T, Miller RW, Crouse CA 2011. Assessing a novel room temperature DNA storage medium for biological samples. *Forensic Science International: Genetics* 6: 31–40.
- Lejzerowicz F, Esling P, Pawlowski J 2014. Patchiness of deep-sea benthic Foraminifera across the Southern Ocean: insights from high-throughput DNA sequencing. *Deep-Sea Research Part II: Topical Studies in Oceanography* 108: 17–26.
- Lentendu G, Wubet T, Chatzinotas A, Wilhelm C, Buscot F, Schlegel M 2014. Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil: a multiple barcoding approach. *Molecular Ecology* 23: 3341–3355.
- Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology* 10: 34.
- Lin YT, Lin CP, Chaw SM, Whitman WB, Coleman DC, Chiu CY 2010. Bacterial community of very wet and acidic subalpine forest and fire-induced grassland soils. *Plant*

- and Soil 332: 417–427.
- Lindahl BD, Nilsson RH, Tedersoo L, Abarenkov K, Carlsen T, Kjoller R, Koljalg U, Pennanen T, Rosendahl S, Stenlid J, Kauserud H 2013. Fungal community analysis by high-throughput sequencing of amplified markers – a user’s guide. *New Phytologist* 199: 288–299.
- Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M 2012. Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology* 2012: Article 251364
- Liu SL, Li YY, Lu JL, Su X, Tang M, Zhang R, Zhou LL, Zhou CR, Yang Q, Ji YQ, Yu DW, Zhou X 2013. SOAPBarcode: revealing arthropod biodiversity through assembly of Illumina shotgun sequences of PCR amplicons. *Methods in Ecology and Evolution* 4: 1142–1150.
- Lumini E, Orgiazzi A, Borriello R, Bonfante P, Bianciotto V 2010. Disclosing arbuscular mycorrhizal fungal biodiversity in soil through a land-use gradient using a pyrosequencing approach. *Environmental Microbiology* 12: 2165–2179.
- McDonagh S 2003. Equipping and establishing a PCR laboratory. Humana Press, Towata, New Jersey, USA.
- McDonald D, Price MN, Goodrich J, Nawrocki EP, deSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of Bacteria and Archaea. *The ISME Journal* 6: 610–618.
- McGee KM, Eaton WD 2015. A comparison of the wet and dry season DNA-based soil invertebrate community characteristics in large patches of the bromeliad *Bromelia pinguin* in a primary forest in Costa Rica. *Applied Soil Ecology* 87: 99–107.
- Machida RJ, Knowlton N 2012. PCR primers for metazoan nuclear 18S and 28S ribosomal DNA sequences. *PLoS ONE* 7: e46180.
- Machler E, Deiner K, Steinmann P, Altermatt F 2014. Utility of environmental DNA for monitoring rare and indicator macroinvertebrate species. *Freshwater Science* 33: 1174–1183.
- Mahmoudi N, Slater GF, Fulthorpe RR 2011. Comparison of commercial DNA extraction kits for isolation and purification of bacterial and eukaryotic DNA from PAH-contaminated soils. *Canadian Journal of Microbiology* 57: 623–628.
- Manter DK, Weir TL, Vivanco JM 2010. Negative effects of sample pooling on PCR-based estimates of soil microbial richness and community structure. *Applied and Environmental Microbiology* 76: 2086–2090.
- Mao DQ, Luo Y, Mathieu J, Wang Q, Feng L, Mu QH, Feng CY, Alvarez PJJ 2014. Persistence of extracellular DNA in river sediment facilitates antibiotic resistance gene propagation. *Environmental Science & Technology* 48: 71–78.
- Mardis ER 2013. Next-generation sequencing platforms. *Annual Review of Analytical Chemistry* 6: 287–303.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376–380.
- Maropola MK, Ramond JB, Trindade M 2015. Impact of metagenomic DNA extraction procedures on the identifiable endophytic bacterial diversity in *Sorghum bicolor* (L. Moench). *Journal of Microbiological Methods* 112: 104–117.
- Martin-Laurent F, Philippot L, Hallet S, Chaussod R, Germon JC, Soulas G, Catroux G 2001. DNA extraction from soils: old bias for new microbial diversity analysis methods. *Applied and Environmental Microbiology* 67: 2354–2359.
- Martinez-Garcia LB, Richardson SJ, Tyljanakis JM, Peltzer DA, Dickie IA 2015. Host identity is a dominant driver of mycorrhizal fungal community composition during ecosystem development. *New Phytologist* 205: 1565–1576.
- 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: e114639.
- Meadow JF, Zabinski CA 2012. Spatial heterogeneity of eukaryotic microbial communities in an unstudied geothermal diatomaceous biological soil crust: Yellowstone National Park, WY, USA. *FEMS Microbiology Ecology* 82: 182–191.
- Merlin C, Besaury L, Niepceyron M, Mchergui C, Riah W, Bureau F, Gattin I, Bodilis J 2014. Real-time PCR for quantification in soil of glycoside hydrolase family 6 cellulase genes. *Letters in Applied Microbiology* 59: 284–291.
- Meyer M, Stenzel U, Hofreiter M 2008. Parallel tagged sequencing on the 454 platform. *Nature Protocols* 3: 267–278.
- Mikheyev AS, Tin MMY 2014. A first look at the Oxford Nanopore MinION sequencer. *Molecular Ecology Resources* 14: 1097–1102.
- Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF 2011. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biology* 12: R44.
- Miller DN 2001. Evaluation of gel filtration resins for the removal of PCR-inhibitory substances from soils and sediments. *Journal of Microbiological Methods* 44: 49–58.
- Miller DN, Bryant JE, Madsen EL, Ghiorse WC 1999. Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Applied and Environmental Microbiology* 65: 4715–4724.
- Min MS, Okumura H, Jo DJ, An JH, Kim KS, Kim CB, Shin NS, Lee MH, Han CH, Voloshina IV, Lee H 2004. Molecular phylogenetic status of the Korean goral and Japanese serow based on partial sequences of the mitochondrial cytochrome b gene. *Molecules and Cells* 17: 365–372.
- Minamoto T, Yamanaka H, Takahara T, Honjo MN, Kawabata Z 2012. Surveillance of fish species composition using environmental DNA. *Limnology* 13: 193–197.
- Minamoto T, Nake T, Moji K, Maruyama A 2016. Techniques for the practical collection of environmental DNA: filter selection, preservation, and extraction. *Limnology* 17: 23–32.
- Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, Minamoto T, Yamamoto S, Yamanaka H, Araki H, Kondoh M, Iwasaki W 2015. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species.

- Royal Society Open Science 2: 150088.
- Mohiuddin M, Schellhorn HE 2015. Spatial and temporal dynamics of virus occurrence in two freshwater lakes captured through metagenomic analysis. *Frontiers in Microbiology* 6: article 960.
- Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B 2011. How many species are there on earth and in the ocean? *PLoS Biology* 9: e1001127.
- Morrison-Whittle P, Goddard MR 2015. Quantifying the relative roles of selective and neutral processes in defining eukaryotic microbial communities. *ISME J* 9: 2003–2011.
- Moyer GR, Diaz-Ferguson E, Hill JE, Shea C 2014. Assessing environmental DNA detection in controlled lentic systems. *PLoS ONE* 9: e1003767.
- Nagahama T, Takahashi E, Nagano Y, Abdel-Wahab MA, Miyazaki M 2011. Molecular evidence that deep-branching fungi are major fungal components in deep-sea methane cold-seep sediments. *Environmental Microbiology* 13: 2359–2370.
- Nagy ZT 2010. A hands-on overview of tissue preservation methods for molecular genetic analyses. *Organisms Diversity & Evolution* 10: 91–105.
- Nakai R, Abe T, Takeyama H, Naganuma T 2011. Metagenomic analysis of 0.2 µm-passable microorganisms in deep-sea hydrothermal fluid. *Marine Biotechnology* 13: 900–908.
- Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J 2014. Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. *PLoS ONE* 9: e94249.
- Nichols RV, Spong G 2014. Ungulate browsing on conifers during summer as revealed by DNA. *Scandinavian Journal of Forest Research* 29: 650–652.
- Nilsson RH, Ryberg M, Abarenkov K, Sjøkvist E, Kristiansson E 2009. The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. *FEMS Microbiology Letters* 296: 97–101.
- Nolte V, Pandey RV, Jost S, Medinger R, Ottenwaelder B, Boenigk J, Schloetterer C 2010. Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. *Molecular Ecology* 19: 2908–2915.
- Ochoa de Alda JAG, Esteban R, Diago ML, Houmard J 2014. The plastid ancestor originated among one of the major cyanobacterial lineages. *Nature Communications* 5: article 5937.
- Ogram A, Sayler GS, Barkay T 1987. The extraction and purification of microbial DNA from sediments. *Journal of Microbiological Methods* 7: 57–66.
- Olds BP, Jerde CL, Renshaw MA, Li Y, Evans NT, Turner CR, Deiner K, Mahon AR, Brueseke MA, Shirey PD, Pfreder ME, Lodge DM, Lamberti GA 2016. Estimating species richness using environmental DNA. *Ecology and Evolution* 6: 4214–4226.
- Olson ZH, Briggler JT, Williams RN 2012. An eDNA approach to detect eastern hellbenders (*Cryptobranchus a. alleganiensis*) using samples of water. *Wildlife Research* 39: 629–636.
- Opik M, Vanatoa A, Vanatoa E, Moora M, Davison J, Kalwij JM, Reier U, Zobel M 2010. The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). *New Phytologist* 188: 223–241.
- Oskam CL, Haile J, McLay E, Rigby P, Allentoft ME, Olsen ME, Bengtsson C, Miller GH, Schweninger J-L, Jacomb C, Walter R, Baynes A, Dortch J, Parker-Pearson M, Gilbert MTP, Holdaway RN, Willerslev E, Bunce M 2010. Fossil avian eggshell preserves ancient DNA. *Proceedings of the Royal Society B* 277: 1991–2000.
- Ou CY, Moore JL, Schochetman G 1991. Use of UV irradiation to reduce false positivity in polymerase chain reaction. *Biotechniques* 10: 442–446.
- Owen JG, Charlop-Powers Z, Smith AG, Ternei MA, Calle PY, Reddy BVB, Montiel D, Brady SF 2015. Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 112: 4221–4226.
- Pace NR, Stahl DA, Lane DJ, Olsen GJ 1986. The analysis of natural microbial populations by ribosomal RNA sequences. In: Marshall KC ed. *Advances in microbial ecology*. Boston, MA, Springer US. Pp. 1–55.
- Padamsee M, Johansen RB, Stuckey SA, Williams SE, Hooker JE, Burns BR, Bellgard SE 2016. The arbuscular mycorrhizal fungi colonising roots and root nodules of New Zealand kauri *Agathis australis*. *Fungal Biology* 120: 807–817.
- Paknia O, Rajaei Sh. H, Koch A 2015. Lack of well-maintained natural history collections and taxonomists in megadiverse developing countries hampers global biodiversity exploration. *Organisms Diversity & Evolution* 15: 619–629.
- Pal S, Gregory-Eaves I, Pick FR 2015. Temporal trends in cyanobacteria revealed through DNA and pigment analyses of temperate lake sediment cores. *Journal of Paleolimnology* 54: 87–101.
- Pansu J, De Danieli S, Puissant J, Gonzalez J-M, Gielly L, Cordonnier T, Zinger L, Brun J-J, Choler P, Taberlet P, Cécillon L 2015a. Landscape-scale distribution patterns of earthworms inferred from soil DNA. *Soil Biology and Biochemistry* 83: 100–105.
- Pansu J, Giguët-Covex C, Ficetola GF, Gielly L, Boyer F, Zinger L, Arnaud F, Poulenard J, Taberlet P, Choler P 2015b. Reconstructing long-term human impacts on plant communities: an ecological approach based on lake sediment DNA. *Molecular Ecology* 24: 1485–1498.
- Parada AE, Needham DM, Fuhrman JA 2015. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology* 18: 1403–1414.
- Parducci L, Matetovici I, Fontana SL, Bennett KD, Suyama Y, Haile J, Kjaer KH, Larsen NK, Drouzas AD, Willerslev E 2013. Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia. *Molecular Ecology* 22: 3511–3524.
- Patel P, Garson JA, Tettmar KI, Ancliff S, McDonald C, Pitt T, Coelho J, Tedder RS 2012. Development of an ethidium monoazide-enhanced internally controlled universal 16S rDNA real-time polymerase chain reaction assay for detection of bacterial contamination in platelet concentrates. *Transfusion* 52: 1423–1432.
- Pawlowski J, Christen R, Lecroq B, Bachar D, Shahbazkia HR, Amaral-Zettler L, Guillou L 2011. Eukaryotic richness in the abyss: insights from pyrotag sequencing. *PLoS ONE* 6: e18169.
- Pawlowski J, Audic S, Adl S, Bass D, Belbahri L, Berney C, Bowser SS, Cepicka I, Decelle J, Dunthorn M, Fiore-Donno AM, Gile GH, Holzmann M, Jahn R, Jirku M,

- Keeling PJ, Kostka M, Kudryavtsev A, Lara E, Lukes J, Mann DG, Mitchell EAD, Nitsche F, Romeralo M, Saunders GW, Simpson AGB, Smirnov AV, Spouge JL, Stern RF, Stoeck T, Zimmermann J, Schindel D, de Vargas C 2012. CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biology* 10: e1001419.
- Pawlowski J, Lejzerowicz F, Apotheloz-Perret-Gentil L, Visco J, Esling P 2016. Protist metabarcoding and environmental biomonitoring: time for change. *European Journal of Protistology* 55, Part A: 12–25.
- Pedersen MW, Ginolhac A, Orlando L, Olsen J, Andersen K, Holm J, Funder S, Willerslev E, Kjaer KH 2013. A comparative study of ancient environmental DNA to pollen and microfossils from lake sediments reveals taxonomic overlap and additional plant taxa. *Quaternary Science Reviews* 75: 161–168.
- Pedersen MW, Overballe-Petersen S, Ermini L, Sarkissian CD, Haile J, Hellstrom M, Spens J, Thomsen PF, Bohmann K, Cappellini E, Schnell IB, Wales NA, Carøe C, Campos PF, Schmidt AMZ, Gilbert MTP, Hansen AJ, Orlando L, Willerslev E 2015. Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370.
- Philippe E, Franck L, Jan P 2015. Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Research* 43: 2513–2524.
- Pilliod DS, Goldberg CS, Arkle RS, Waits LP 2013. Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences* 70: 1123–1130.
- Pochon X, Bott NJ, Smith KF, Wood SA 2013. Evaluating detection limits of next-generation sequencing for the surveillance and monitoring of international marine pests. *PLoS ONE* 8: e73935.
- Pochon X, Wood SA, Keeley NB, Lejzerowicz F, Esling P, Drew J, Pawlowski J 2015a. Accurate assessment of the impact of salmon farming on benthic sediment enrichment using foraminiferal metabarcoding. *Marine Pollution Bulletin* 100: 370–382.
- Pochon X, Zaiko A, Hopkins GA, Banks JC, Wood SA 2015b. Early detection of eukaryotic communities from marine biofilm using high-throughput sequencing: an assessment of different sampling devices. *Biofouling* 31: 241–251.
- Polz MF, Cavanaugh CM 1998. Bias in template to product ratios in multi-template PCR. *Applied and Environmental Microbiology* 64: 3724–3730.
- Poorvin L, Rinta-Kanto JM, Hutchins DA, Wilhelm SW 2004. Viral release of iron and its bioavailability to marine plankton. *Limnology and Oceanography* 49: 1734–1741.
- Pornon A, Escaravage N, Burrus M, Holota H, Khimoun A, Mariette J, Pellizzari C, Iribar A, Etienne R, Taberlet P, Vidal M, Winterton P, Zinger L, Andalo C 2016. Using metabarcoding to reveal and quantify plant-pollinator interactions. *Scientific Reports* 6.
- Port JA, O'Donnell JL, Romero-Maraccini OC, Leary PR, Litvin SY, Nickols KJ, Yamahara KM, Kelly RP 2016. Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology* 25: 527–541.
- Pruesse E, Quast C, Knittel K, Fuchs B, Ludwig W, Peplies J, Glockner FO 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research* 15: 7188–7196.
- Qiu X, Wu L, Huang H, McDonel PE, Palumbo AV, Tiedje JM, Zhou J 2001. Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. *Applied and Environmental Microbiology* 67: 880–887.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41: D590–596.
- Ramirez KS, Leff JW, Barberan A, Bates ST, Betley J, Crowther TW, Kelly EF, Oldfield EE, Shaw EA, Steenbock C, Bradford MA, Wall DH, Fierer N 2014. Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. *Proceedings of the Royal Society B* 281: 20141988.
- Ramón-Laca A, Gleeson D, Yockney I, Perry M, Nugent G, Forsyth DM 2014. Reliable discrimination of 10 ungulate species using high resolution melting analysis of faecal DNA. *PLoS ONE* 9: e92043.
- Ranjard L, Dequiedt S, Prevost-Boure NC, Thioulouse J, Saby NPA, Lelievre M, Maron PA, Morin FER, Bispo A, Jolivet C, Arrouays D, Lemanceau P 2013. Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity. *Nature Communications* 4: 1434.
- Ranjard L, Lejon DP, Mougél C, Schehrer L, Merdinoglu D, Chaussod R 2003. Sampling strategy in molecular microbial ecology: influence of soil sample size on DNA fingerprinting analysis of fungal and bacterial communities. *Environmental Microbiology* 5: 1111–1120.
- Rao S, Chan YK, Lacap DC, Hyde KD, Pointing SB, Farrell RL 2012. Low-diversity fungal assemblage in an Antarctic dry valleys soil. *Polar Biology* 35: 567–574.
- Ratnasingham S, Hebert PDN 2007. BOLD: the barcode of life data system ([www.barcodinglife.org](http://www.barcodinglife.org)). *Molecular Ecology Notes* 7: 355–364.
- Rawlence NJ, Lowe DJ, Wood JR, Young JM, Chruchman GJ, Huang Y-T, Cooper A 2014. Using paleoenvironmental DNA to reconstruct past environments: progress and prospects. *Journal of Quaternary Science* 29: 610–626.
- Read DJ, Perez-Moreno J 2003. Mycorrhizas and nutrient cycling in ecosystems - a journey towards relevance? *New Phytologist* 157: 475–492.
- Rees HC, Bishop K, Middleditch DJ, Patmore JRM, Maddison BC, Gough KC 2014. The application of eDNA for monitoring of the great crested newt in the UK. *Ecology and Evolution* 4: 4023–4032.
- Regan EC, Santini L, Ingwall-King L, Hoffmann M, Rondinini C, Symes A, Taylor JW, Butchart SHM 2015. Global trends in the status of bird and mammal pollinators. *Conservation Letters* 8: 397–403.
- Reuter JA, Spacek DV, Snyder MP 2015. High-throughput sequencing technologies. *Molecular Cell* 58: 586–597.
- Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E 2011. ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research* 39: e145.
- Ribblett SG, Palmer MA, Wayne Coats D 2005. The importance of bacterivorous protists in the decomposition of stream leaf litter. *Freshwater Biology* 50: 516–526.
- Richter M, Rossello-Mora R 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America* 106: 19126–19131.

- Rodríguez-Marconi S, De la Iglesia R, Díez B, Fonseca CA, Hajdu E, Trefault N 2015. Characterization of bacterial, archaeal and eukaryote symbionts from Antarctic sponges reveals a high diversity at a three-domain level and a particular signature for this ecosystem. *PLoS ONE* 10: e0138837.
- Rohwer F, Edwards R 2002. The phage proteomic tree: a genome-based taxonomy for phage. *Journal of Bacteriology* 184: 4529–4535.
- Roose-Amsaleg CL, Garnier-Sillem E, Harry M 2001. Extraction and purification of microbial DNA from soil and sediment samples. *Applied Soil Ecology* 18: 47–60.
- Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475: 348–352.
- Salipante SJ, Kawashima T, Rosenthal C, Hoogestraat DR, Cummings LA, Sengupta DJ, Harkins TT, Cookson BT, Hoffman NG 2014. Performance comparison of Illumina and Ion Torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Applied and Environmental Microbiology* 80: 7583–7591.
- Sanger F, Nicklen S, Coulson AR 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Science of the United States of America* 74: 5463–5467.
- Santas AJ, Persaud T, Wolfe BA, Bauman JM 2013. Noninvasive method for a statewide survey of eastern hellbenders *Cryptobranchus alleganiensis* using environmental DNA. *International Journal of Zoology: Article ID 174056*.
- Sapkota R, Nicolaisen M 2015. An improved high throughput sequencing method for studying oomycete communities. *Journal of Microbiological Methods* 110: 33–39.
- Satoh TP, Miya M, Mabuchi K, Nishida M 2016. Structure and variation of the mitochondrial genome of fishes. *BMC Genomics* 17: 719.
- Schloss PD, Gevers D, Westcott SL 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE* 6: e27310.
- Schloss PD, Jenior ML, Koumpouras C, Westcott SL, Highlander SK 2016. Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. *PeerJ* 4: e1869.
- Schmidt HF, Sakowski EG, Williamson SJ, Polson SW, Wommack KE 2014. Shotgun metagenomics indicates novel family A DNA polymerases predominate within marine viroplankton. *ISME Journal* 8: 103–114.
- Schmieder R, Edwards R 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27: 863–864.
- Schoch CL, Robbertse B, Robert V, Vu D, Cardinali G, Irinyi L, Meyer W, Nilsson RH, Hughes K, Miller AN, Kirk PM, Abarenkov K, Aime MC, Ariyawansa HA, Bidartondo M, Boekhout T, Buyck B, Cai Q, Chen J, Crespo A, Crous PW, Damm U, De Beer ZW, Dentinger BTM, Divakar PK, Duenas M, Feau N, Fliegerova K, Garcia MA, Ge Z-W, Griffith GW, Groenewald JZ, Groenewald M, Grube M, Gryzenhout M, Gueidan C, Guo L, Hambleton S, Hamelin R, Hansen K, Hofstetter V, Hong S-B, Houbraken J, Hyde KD, Inderbitzin P, Johnston PR, Karunarathna SC, Koljalg U, Kovacs GM, Kraichak E, Krizsan K, Kurtzman CP, Larsson K-H, Leavitt S, Letcher PM, Liimatainen K, Liu J-K, Lodge DJ, Luangsa-ard JJ, Lumbsch HT, Maharachchikumbura SSN, Manamgoda D, Martin MP, Minnis AM, Moncalvo J-M, Mule G, Nakasone KK, Niskanen T, Olariaga I, Papp T, Petkovits T, Pino-Bodas R, Powell MJ, Raja HA, Redecker D, Sarmiento-Ramirez JM, Seifert KA, Shrestha B, Stenroos S, Stielow B, Suh S-O, Tanaka K, Tedersoo L, Telleria MT, Udayanga D, Untereiner WA, Dieguez Uribeondo J, Subbarao KV, Vagvolgyi C, Visagie C, Voigt K, Walker DM, Weir BS, Wei Ss M, Wijayawardene NN, Wingfield MJ, Xu JP, Yang ZL, Zhang N, Zhuang W-Y, Federhen S 2014. Finding needles in haystacks: linking scientific names, reference specimens and molecular data for fungi. *Database: the Journal of Biological Databases and Curation* 2014: article bau061.
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Bolchacova E, Voigt K, Crous PW, Miller AN, Wingfield MJ, Aime MC, An KD, Bai FY, Barreto RW, Begerow D, Bergeron MJ, Blackwell M, Boekhout T, Bogale M, Boonyuen N, Burgaz AR, Buyck B, Cai L, Cai Q, Cardinali G, Chaverri P, Coppins BJ, Crespo A, Cubas PP, Cummings C, Damm U, de Beer ZW, de Hoog GS, Del-Prado R, Dentinger B, Dieguez-Uribeondo J, Divakar PK, Douglas B, Duenas M, Duong TA, Eberhardt U, Edwards JE, Elshahed MS, Fliegerova K, Furtado M, Garcia MA, Ge ZW, Griffith GW, Griffiths K, Groenewald JZ, Groenewald M, Grube M, Gryzenhout M, Guo LD, Hagen F, Hambleton S, Hamelin RC, Hansen K, Harrold P, Heller G, Herrera G, Hirayama K, Hirooka Y, Ho HM, Hoffmann K, Hofstetter V, Hognabba F, Hollingsworth PM, Hong SB, Hosaka K, Houbraken J, Hughes K, Huhtinen S, Hyde KD, James T, Johnson EM, Johnson JE, Johnston PR, Jones EB, Kelly LJ, Kirk PM, Knapp DG, Koljalg U, Kovacs GM, Kurtzman CP, Landvik S, Leavitt SD, Liggenstoffer AS, Liimatainen K, Lombard L, Luangsa-Ard JJ, Lumbsch HT, Maganti H, Maharachchikumbura SS, Martin MP, May TW, McTaggart AR, Methven AS, Meyer W, Moncalvo JM, Mongkolsamrit S, Nagy LG, Nilsson RH, Niskanen T, Nyilasi I, Okada G, Okane I, Olariaga I, Otte J, Papp T, Park D, Petkovits T, Pino-Bodas R, Quaedvlieg W, Raja HA, Redecker D, Rintoul T, Ruibal C, Sarmiento-Ramirez JM, Schmitt I, Schussler A, Shearer C, Sotome K, Stefani FO, Stenroos S, Stielow B, Stockinger H, Suetrong S, Suh SO, Sung GH, Suzuki M, Tanaka K, Tedersoo L, Telleria MT, Tretter E, Untereiner WA, Urbina H, Vagvolgyi C, Vialle A, Vu TD, Walther G, Wang QM, Wang Y, Weir BS, Weiss M, White MM, Xu J, Yahr R, Yang ZL, Yurkov A, Zamora JC, Zhang N, Zhuang WY, Schindel D, Fungal Barcoding C 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proceedings of the National Academy of Sciences of the United States of America* 109: 6241–6246.
- Sharp CE, Smirnova AV, Graham JM, Stott MB, Khadka R, Moore TR, Grasby SE, Strack M, Dunfield PF 2014. Distribution and diversity of Verrucomicrobia methanotrophs in geothermal and acidic environments. *Environ Microbiol* 16: 1867–1878.
- Shaw JLA, 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. *Biological Conservation* 197: 131–138.
- Shehzad W, Riaz T, Nawaz MA, Miquel C, Poillot C, Shah SA, Pompanon F, Coissac E, Taberlet P 2012. Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. *Molecular Ecology* 21: 1951–1965.
- Sherr EB, Sherr BF 2002. Significance of predation by protists in aquatic microbial food webs. *Antonie Van Leeuwenhoek* 81: 293–308.
- Shokralla S, Spall JL, Gibson JF, Hajibabaei M 2012. Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology* 21: 1794–1805.
- Shokralla S, Porter TM, Gibson JF, Dobosz R, Janzen DH, Hallwachs W, Golding GB, Hajibabaei M 2015. Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Scientific Reports* 5: 9687.
- Simon L, Lalonde M, Bruns TD 1992. Specific amplification of 18S fungal ribosomal genes from vesicular-arbuscular endomycorrhizal fungi colonizing roots. *Applied and Environmental Microbiology* 58: 291–295.
- Singer E, Bushnell B, Coleman-Derr D, Bowman B, Bowers RM, Levy A, Gies EA, Cheng J-F, Copeland A, Klenk H-P, Hallam SJ, Hugenholtz P, Tringe SG, Woyke T 2016. High-resolution phylogenetic microbial community profiling. *The ISME Journal* 10: 2020–2032.
- Singh P, Raghukumar C, Verma P, Shouche Y 2012. Assessment of fungal diversity in deep-sea sediments by multiple primer approach. *World Journal of Microbiology and Biotechnology* 28: 659–667.
- Sint D, Niederklapfer B, Kaufmann R, Traugott M 2014. Group-specific multiplex PCR detection systems for the identification of flying insect prey. *PLoS ONE* 9: e115501.
- Sirisena KA, Daughney CJ, Moreau-Fournier M, Ryan KG, Chambers GK 2013. National survey of molecular bacterial diversity of New Zealand groundwater: relationships between biodiversity, groundwater chemistry and aquifer characteristics. *FEMS Microbiology Ecology* 86: 490–504.
- Smitsen RD, Breitwieser I, Ward JM 2004. Phylogenetic implications of trans-specific chloroplast DNA sequence polymorphism in New Zealand Gnaphalieae (Asteraceae). *Plant Systematics and Evolution* 249: 37–53.
- Soergel DAW, Dey N, Knight R, Brenner SE 2012. Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *The ISME Journal* 6: 1440–1444.
- Song ZW, Schlatter D, Kennedy P, Kinkel LL, Kistler HC, Nguyen N, Bates ST 2015. Effort versus reward: preparing samples for fungal community characterization in high-throughput sequencing surveys of soils. *PLoS ONE* 10: e0127234.
- Spear SF, Groves JD, Williams LA, Waits LP 2015. Using environmental DNA methods to improve detectability in a hellbender (*Cryptobranchus alleganiensis*) monitoring program. *Biological Conservation* 183: 38–45.
- Srinivasiah S, Bhavsar J, Thapar K, Liles M, Schoenfeld T, Wommack KE 2008. Phages across the biosphere: contrasts of viruses in soil and aquatic environments. *Research in Microbiology* 159: 349–357.
- Staley JT 2006. The bacterial species dilemma and the genomic-phylogenetic species concept. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 361: 1899–1909.
- Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner HW, Richards TA 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology* 19: 21–31.
- Stoof-Leichsenring KR, Epp LS, Trauth MH, Tiedemann R 2012. Hidden diversity in diatoms of Kenyan Lake Naivasha: a genetic approach detects temporal variation. *Molecular Ecology* 21: 1918–1930.
- Suttle CA 2005. Viruses in the sea. *Nature* 437: 356–361.
- Taberlet P, Gielly L, Pautou G, Bouvet J 1991. Universal primers for amplification of 3 noncoding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105–1109.
- Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, Willerslev E 2006. Power and limitations of the chloroplast trn L (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* 35: e14–e14.
- Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, Willerslev E 2007. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* 35: e14.
- Taberlet P, Prud'Homme SM, Campione E, Roy J, Miquel C, Shehzad W, Gielly L, Rioux D, Choler P, Clement JC, Melodelima C, Pompanon F, Coissac E 2012. Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Molecular Ecology* 21: 1816–1820.
- Takahara T, Minimoto H 2013. Using environmental DNA to estimate the distribution of an invasive fish species in ponds. *PLoS ONE* 8: e56584.
- Takahara T, Minamoto T, Doi H 2015. Effects of sample processing on the detection rate of environmental DNA from the common carp (*Cyprinus carpio*). *Biological Conservation* 183: 64–69.
- Taylor DL, Hollingsworth TN, McFarland JW, Lennon NJ, Nusbaum C, Ruess RW 2014. A first comprehensive census of fungi in soil reveals both hyperdiversity and fine-scale niche partitioning. *Ecological Monographs* 84: 3–20.
- Teasdale SE, Beulke AK, Guy PL, Orlovich DA 2013. Environmental barcoding of the ectomycorrhizal fungal genus *Cortinari*. *Fungal Diversity* 58: 299–310.
- Tedersoo L, Nilsson RH, Abarenkov K, Jairus T, Sadam A, Saar I, Bahram M, Bechem E, Chuyong G, Koljalg U 2010. 454 pyrosequencing and sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytologist* 188: 291–301.
- Tedersoo L, Bahram M, Pölme S, Kõljalg U, Yorou NS, Wijesundera R, Ruiz LV, Vasco-Palacios AM, Thu PQ, Suija A, Smith ME, Sharp C, Saluveer E, Saitta A, Rosas M, Riit T, Ratkowsky D, Pritsch K, Põldmaa K, Piepenbring M, Phosri C, Peterson M, Parts K, Pärtel K, Otting E, Nouhra E, Njouonkou AL, Nilsson RH, Morgado LN, Mayor J, May TW, Majuakim L, Lodge DJ, Lee SS, Larsson K-H, Kohout P, Hosaka K, Hiiesalu I, Henkel TW, Harend H, Guo L-d, Greslebin A, Grelet G, Geml J, Gates G, Dunstan W, Dunk C, Drenkhan R, Dearnaley J, De Kesel A, Dang T, Chen X, Buegger F, Brearley FQ, Bonito G, Anslan S, Abell S, Abarenkov K 2014. Global diversity and geography of soil fungi. *Science* 346: 1256688–1256681/1256610.

- Tedersoo L, Anslan S, Bahram M, Polme S, Riit T, Liiv I, Koljalg U, Kisand V, Nilsson H, Hildebrand F, Bork P, Abarenkov K 2015. Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycKeys* 10: 1–43.
- The N. I. H. H. M. P. Working Group, Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C, Baker CC, Di Francesco V, Howcroft TK, Karp RW, Lunsford RD, Wellington CR, Belachew T, Wright M, Giblin C, David H, Mills M, Salomon R, Mullins C, Akolkar B, Begg L, Davis C, Grandison L, Humble M, Khalsa J, Little AR, Peavy H, Pontzer C, Portnoy M, Sayre MH, Starke-Reed P, Zakhari S, Read J, Watson B, Guyer M 2009. The NIH human microbiome project. *Genome Research* 19: 2317–2323.
- Thomsen PF, Kielgast J, Iversen LL, Møller PR, Rasmussen M, Willerslev E 2012a. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS ONE* 7: e41732.
- Thomsen PF, Kielgast J, Iversen LL, Wiuf C, Rasmussen M, Gilbert MT, Orlando L, Willerslev E 2012b. Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology* 21: 2565–2573.
- Tien CC, Chao CC, Chao WL 1999. Methods for DNA extraction from various soils: a comparison. *Journal of Applied Microbiology* 86: 937–943.
- Tiirik K, Nolvak H, Oopkaup K, Truu M, Preem JK, Heinaru A, Truu J 2014. Characterization of the bacterioplankton community and its antibiotic resistance genes in the Baltic Sea. *Biotechnology and Applied Biochemistry* 61: 23–32.
- Torsvik VL 1980. Isolation of bacterial DNA from soil. *Soil Biology and Biochemistry* 12: 15–21.
- Torsvik V, Goksøyr J, Daae FL 1990. High diversity in DNA of soil bacteria. *Applied and Environmental Microbiology* 56: 782–787.
- Tremblay J, Singh K, Fern A, Kirton ES, He S, Woyke T, Lee J, Chen F, Dangel JL, Tringe SG 2015. Primer and platform effects on 16S rRNA tag sequencing. *Frontiers in Microbiology* 6: article 771.
- Trewick SA, Morgan-Richards M, Russell SJ, Henderson S, Rumsey FJ, Pinter I, Barrett JA, Gibby M, Vogel JC 2002. Polyploidy, phylogeography and Pleistocene refugia of the rockfern *Asplenium ceterach*: evidence from chloroplast DNA. *Molecular Ecology* 11: 2003–2012.
- Tringe SG, Rubin EM 2005. Metagenomics: DNA sequencing of environmental samples. *Nature Reviews Genetics* 6: 805–814.
- Tsai YL, Olson BH 1992a. Detection of low numbers of bacterial cells in soils and sediments by polymerase chain reaction. *Applied and Environmental Microbiology* 58: 754–757.
- Tsai YL, Olson BH 1992b. Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. *Applied and Environmental Microbiology* 58: 2292–2295.
- Turner CR, Miller DJ, Coyne KJ, Corush J 2014. Improved methods for capture, extraction, and quantitative assay of environmental DNA from Asian bigheaded carp (*Hypophthalmichthys* spp.). *PLoS ONE* 9: e114329.
- Turner CR, Uy KL, Everhart RC 2015. Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biological Conservation* 183: 93–102.
- Urakawa H, Martens-Habbena W, Stahl DA 2010. High abundance of ammonia oxidizing Archaea in coastal waters, determined using a modified DNA extraction method. *Applied and Environmental Microbiology* 76: 2129–2135.
- Ushio M, Fukuda H, Inoue T, Makoto K, Kishida O, Sato K, Murata K, Nikaido M, Sado T, Sato Y, Takeshita M, Iwasaki W, Yamanaka H, Kondoh M, Miya M 2016. Environmental DNA enables detection of terrestrial mammals from forest pond water. *bioRxiv*: 068551.
- Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, Bellemain E, Besnard A, Coissac E, Boyer F, Gaboriaud C, Jean P, Poulet N, Roset N, Copp GH, Geniez P, Pont D, Argillier C, Baudoin JM, Peroux T, Crivelli AJ, Olivier A, Acqueberge M, Le Brun M, Moller PR, Willerslev E, Dejean T 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology* 25: 929–942.
- Valiere N, Taberlet P 2000. Urine collected in the field as a source of DNA for species and individual identification. *Molecular Ecology Resources* 9: 2150–2152.
- van Bleijswijk JDL, Begeman L, Witte HJ, IJsseldijk LL, Brasseur SMJM, Grone A, Leopold MF 2014. Detection of grey seal *Halichoerus grypus* DNA in attack wounds on stranded harbour porpoises *Phocoena phocoena*. *Marine Ecology Progress Series* 513: 277–281.
- van der Heijden MG, Bardgett RD, van Straalen NM 2008. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters* 11: 296–310.
- van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C 2014. Ten years of next-generation sequencing technology. *Trends in Genetics* 30: 418–426.
- Van Geel M, Busschaert P, Honnay O, Lievens B 2014. Evaluation of six primer pairs targeting the nuclear rRNA operon for characterization of arbuscular mycorrhizal fungal (AMF) communities using 454 pyrosequencing. *Journal of Microbiological Methods* 106: 93–100.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Francesco VD, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji R-R, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang ZY, Wang A, Wang X, Wang J, Wei M-H, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu SC, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg



- N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers Y-H, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang Y-H, Coyne M, Dahlke C, Mays AD, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X 2001. The sequence of the human genome. *Science* 291: 1304–1351.
- Vestheim H, Jarman SN 2008. Blocking primers to enhance PCR amplification of rare sequences in mixed samples—a case study on prey DNA in Antarctic krill stomachs. *Frontiers in Zoology* 5: 12.
- Voříšková J, Baldrian P 2013. Fungal community on decomposing leaf litter undergoes rapid successional changes. *The ISME Journal* 7: 477–486.
- Wagner Mackenzie B, Waite DW, Taylor MW 2015. Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences. *Frontiers in Microbiology* 6: 130.
- Wagstaff SJ, Garnock-Jones PJ 1998. Evolution and biogeography of the *Hebe* complex (Scrophulariaceae) inferred from ITS sequences. *New Zealand Journal of Botany* 36: 425–437.
- Wang Q, Garrity GM, Tiedje JM, Cole JR 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* 73: 5261–5267.
- Wasaki J, Taguchi H, Senoura T, Akasaka H, Watanabe J, Kawaguchi K, Komata Y, Hanashiro K, Ito S 2015. Identification and distribution of cellobiose 2-epimerase genes by a PCR-based metagenomic approach. *Applied Microbiology and Biotechnology* 99: 4287–4295.
- Watson RA, Carlos GM, Samoilys MA 1995. Bias introduced by the non-random movement of fish in visual transect surveys. *Ecological Modelling* 77: 205–214.
- Wei F, Passey T, Xu X 2016. Amplicon-based metabarcoding reveals temporal response of soil microbial community to fumigation-derived products. *Applied Soil Ecology* 103: 83–92.
- White TJ, Bruns TD, Lee SG, Taylor JW 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ eds. *PCR protocols: a guide to methods and applications*. New York, Academic Press. Pp. 315–322.
- Whitehead AG, Hemming JR 1965. A comparison of some quantitative methods of extracting small vermiform nematodes from soil. *Annals of Applied Biology* 55: 25–38.
- Whiteley AS, Jenkins S, Waite I, Kresoje N, Payne H, Mullan B, Allcock R, O'Donnell A 2012. Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. *Journal of Microbiological Methods* 91: 80–88.
- Wichman SR, Wright SD, Cameron EK, Keeling DJ, Gardner RC 2002. Elevated genetic heterogeneity and Pleistocene climatic instability: inferences from nrDNA in New Zealand coprosma (Rubiaceae). *Journal of Biogeography* 29: 943–954.
- Wilcox TM, McKelvey KS, Young MK, Jane SF, Lowe WH, Whiteley AR, Schwartz MK 2013. Robust detection of rare species using environmental DNA: the importance of primer specificity. *PLoS ONE* 8: e59520.
- Wilcox TM, Schwartz MK, McKelvey KS, Young MK, Lowe WH 2014. A blocking primer increases specificity in environmental DNA detection of bull trout (*Salvelinus confluentus*). *Conservation Genetics Resources* 6: 283–284.
- Wilcox TM, McKelvey KS, Young MK, Sepulveda AJ, Shepard BB, Jane SF, Whiteley AR, Lowe WH, Schwartz MK 2016. Understanding environmental DNA detection probabilities: a case study using a stream-dwelling char *Salvelinus fontinalis*. *Biological Conservation* 194: 209–216.
- Williams BD, Schrank B, Huynh C, Shownkeen R, Waterston RH 1992. A genetic mapping system in *Caenorhabditis elegans* based on polymorphic sequence-tagged sites. *Genetics* 131: 609–624.
- Wilson CC, Wozney KM, Smith CM 2016. Recognizing false positives: synthetic oligonucleotide controls for environmental DNA surveillance. *Methods in Ecology and Evolution* 7: 23–29.
- Winkworth RC, Grau J, Robertson AW, Lockhart PJ 2002. The origins and evolution of the genus *Myosotis* L. (Boraginaceae). *Molecular Phylogenetics and Evolution* 24: 180–193.
- Wood JR, Wilmshurst JM, Worthy TH, Cooper A 2011. Sporormiella as a proxy for non-mammalian herbivores in island ecosystems. *Quaternary Science Reviews* 30: 915–920.
- Wood JR, Crown A, Cole TL, Wilmshurst JM 2016. Microscopic and ancient DNA profiling of Polynesian dog (*kuī*) coprolites from northern New Zealand. *Journal of Archaeological Science* 6: 496–505.
- Wright JJ, Lee S, Zaikova E, Walsh DA, Hallam SJ 2009. DNA extraction from 0.22 μm sterivex filters and cesium chloride density gradient centrifugation. *Journal of Visualized Experiments: JoVE*: 1352.
- Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC, Wilkins MJ, Hettich RL, Lipton MS, Williams KH, Long PE, Banfield JF 2012. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* 337: 1661–1665.
- Wu L, Wen C, Qin Y, Yin H, Tu Q, Van Nostrand JD, Yuan T, Yuan M, Deng Y, Zhou J 2015. Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. *BMC Microbiology* 15: 125.
- Xiao X, Sogge H, Lagesen K, Tooming-Klunderud A, Jakobsen KS, Rohrlack T 2014. Use of high throughput sequencing and light microscopy show contrasting results in a study

- of phytoplankton occurrence in a freshwater environment. *PLoS ONE* 9: e106510.
- Yamanaka H, Motozawa H, Tsuji S, Miyazawa RC, Takahara T, Minamoto T 2016. On-site filtration of water samples for environmental DNA analysis to avoid DNA degradation during transportation. *Ecological Research* 31: 963–937.
- Yang CX, Wang XY, Miller JA, de Blecourt M, Ji YQ, Yang CY, Harrison RD, Yu DW 2014. Using metabarcoding to ask if easily collected soil and leaf-litter samples can be used as a general biodiversity indicator. *Ecological Indicators* 46: 379–389.
- Yang YZ, Zhan AB, Cao L, Meng FJ, Xu WB 2016. Selection of a marker gene to construct a reference library for wetland plants, and the application of metabarcoding to analyze the diet of wintering herbivorous waterbirds. *PeerJ* 4: e2345.
- Yergeau E, Lawrence JR, Sanschagrin S, Waiser MJ, Korber DR, Greer CW 2012. Next-generation sequencing of microbial communities in the Athabasca River and its tributaries in relation to oil sands mining activities. *Applied and Environmental Microbiology* 78: 7626–7637.
- Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, Gilbert JA, Karsch-Mizrachi I, Johnston A, Cochrane G, Vaughan R, Hunter C, Park J, Morrison N, Rocca-Serra P, Sterk P, Arumugam M, Bailey M, Baumgartner L, Birren BW, Blaser MJ, Bonazzi V, Booth T, Bork P, Bushman FD, Buttigieg PL, Chain PSG, Charlson E, Costello EK, Huot-Creasy H, Dawyndt P, DeSantis T, Fierer N, Fuhrman JA, Gallery RE, Gevers D, Gibbs RA, Gil IS, Gonzalez A, Gordon JI, Guralnick R, Hankeln W, Highlander S, Hugenholtz P, Jansson J, Kau AL, Kelley ST, Kennedy J, Knights D, Koren O, Kuczynski J, Kyrpides N, Larsen R, Lauber CL, Legg T, Ley RE, Lozupone CA, Ludwig W, Lyons D, Maguire E, Methe BA, Meyer F, Muegge B, Nakielny S, Nelson KE, Nemergut D, Neufeld JD, Newbold LK, Oliver AE, Pace NR, Palanisamy G, Peplies J, Petrosino J, Proctor L, Pruesse E, Quast C, Raes J, Ratnasingham S, Ravel J, Relman DA, Assunta-Sansone S, Schloss PD, Schriml L, Sinha R, Smith MI, Sodergren E, Spor A, Stombaugh J, Tiedje JM, Ward DV, Weinstock GM, Wendel D, White O, Whiteley A, Wilke A, Wortman JR, Yatsunenko T, Glockner FO 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIXS) specifications. *Nature Biotechnology* 29: 415–420.
- Yoccoz NG, Brathen KA, Gielly L, Haile J, Edwards ME, Goslar T, von Stedingk H, Brysting AK, Coissac E, Pompanon F, Sonstebo JH, Miquel C, Valentini A, de Bello F, Chave J, Thuiller W, Wincker P, Cruaud C, Gavory F, Rasmussen M, Gilbert MTP, Orlando L, Brochmann C, Willerslev E, Taberlet P 2012. DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology* 21: 3647–3655.
- Young JM, Weyrich LS, Cooper A 2014. Forensic soil DNA analysis using high-throughput sequencing: a comparison of four molecular markers. *Forensic Science International-Genetics* 13: 176–184.
- Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* 3: 613–623.
- Zaiko A, Martinez JL, Schmidt-Petersen J, Ribicic D, Samuiloviene A, Garcia-Vazquez E 2015. Metabarcoding approach for the ballast water surveillance – an advantageous solution or an awkward challenge? *Marine Pollution Bulletin* 92: 25–34.
- Zhang A-b, Feng J, Ward RD, Wan P, Gao Q, Wu J, Zhao W-z 2012. A new method for species identification via protein-coding and non-coding DNA barcodes by combining machine learning with bioinformatic methods. *PLoS ONE* 7: e30986.
- Zheng JW, Subbotin SA, He SS, Gu JF, Moens M 2003. Molecular characterisation of some Asian isolates of *Bursaphelenchus xylophilus* and *B. mucronatus* using PCR-RFLPs and sequences of ribosomal DNA. *Russian Journal of Nematology* 11: 17–22.
- Zhou X, Li Y, Liu S, Yang Q, Su X, Zhou L, Tang M, Fu R, Li J, Huang Q 2013. Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *Gigascience* 2: 4.
- Ziesemer KA, Mann AE, Sankaranarayanan K, Schroeder H, Ozga AT, Brandt BW, Zaura E, Waters-Rist A, Hoogland M, Salazar-García DC, Aldenderfer M, Speller C, Hendy J, Weston DA, MacDonald SJ, Thomas GH, Collins MJ, Lewis CM, Hofman C, Warinner C 2015. Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification. *Scientific Reports* 5: 16498.
- Zimmermann J, Glockner G, Jahn R, Enke N, Gemeinholzer B 2015. Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. *Molecular Ecology Resources* 15: 526–542.
- Zinger L, Chave J, Coissac E, Iribar A, Louisanna E, Manzi S, Schilling V, Schimann H, Sommeria-Klein G, Taberlet P 2016. Extracellular DNA extraction is a fast, cheap and reliable alternative for multi-taxa surveys based on soil DNA. *Soil Biology and Biochemistry* 96: 16–19.
- Zulian A, Canello R, Cesana E, Rizzi E, Consolandi C, Severgnini M, Panizzo V, Di Blasio AM, Micheletto G, Invitti C 2016. Adipose tissue microbiota in humans: an open issue. *International Journal of Obesity* 40: 1643–1648.

Editorial board member: George Perry

Received 18 January 2017; accepted 21 September 2017

## Supplementary Material

Additional supporting information may be found in the online version of this article:

**Table S1.** Geographic locations of eDNA studies collated in our review of the literature.

**Table S2.** Numbers of environmental DNA studies out of 167 reviewed papers that targeted species that are invasive, rare, or both.

**Table S3.** Habitats and sample media analysed in 167 research studies on environmental DNA.

**Table S4.** Summary of extraction kits used for the analysis of eDNA from different sample media collated in our review of the literature.

**Table S5.** Primers used to amplify DNA barcodes from prokaryotes.

**Appendix S6.** PCR protocol for prokaryote DNA.

**Table S7.** Primers used to amplify DNA barcodes from eukarya.

**Appendix S8.** PCR protocol for eukaryote DNA.

**Table S9.** Primers used to amplify DNA barcodes from Chromista.

**Appendix S10.** PCR protocol for Chromista DNA.

**Appendix S11.** PCR protocol for plant DNA.

**Table S12.** Primers used to amplify DNA barcodes from fungi.

**Appendix S13.** PCR protocol for fungal DNA.

**Table S14.** Primers used to amplify DNA barcodes from Glomeromycota.

**Table S15.** Primers used to amplify COI DNA barcodes from animals.

**Table S16.** Primers used to amplify mitochondrial 16S rRNA gene barcodes from animals.

**Appendix S17.** PCR protocol for animal DNA.

**Table S18.** Primers used to amplify DNA barcodes from terrestrial vertebrates.

**Table S19.** Primers used to amplify DNA barcodes from fish.

**Appendix S20.** PCR protocol for fish DNA.

The New Zealand Journal of Ecology provides online supporting information supplied by the authors where this may assist readers. Such materials are peer-reviewed and copy-edited but any issues relating to this information (other than missing files) should be addressed to the authors.