Natural England Commissioned Report NECR252

Development of DNA applications in Natural England 2016/2017

First published August 2018

www.gov.uk/natural-england

Foreword

Natural England commission a range of reports from external contractors to provide evidence and advice to assist us in delivering our duties. The views in this report are those of the authors and do not necessarily represent those of Natural England.

Background

DNA based applications have the potential to significantly change how we monitor biodiversity and which species and taxa we monitor. These techniques may provide cheaper alternatives to existing species monitoring, an ability to detect species that we do not currently monitor effectively and the potential to develop new measures of habitat and ecosystem quality.

Natural England has been supporting the development of DNA techniques for a number of years. The use of environmental DNA (eDNA) to determine the presence or absence of great crested newts in ponds is now a standard tool for developers and consultants.

There are still significant limitations to the use of this technology in others areas and in 2016/17 Natural England worked with NatureMetrics on a number of exploratory projects looking at species detection in standing freshwaters, saline lagoons, coastal waters and sediments, terrestrial invertebrate traps, deadwood mould, vegetation and soils. This report presents the results from those projects.

This report should be cited as:

TANG, C.Q., CRAMPTON-PLATT, A., TOWNEND, S., BRUCE, K., BISTA, I. & CREER, S. 2018. Development of DNA applications in Natural England 2016/2017. Natural England Commissioned Reports, Number 252.

Natural England Project Manager – Andy Nisbet, Principal Adviser and Evidence Programme Manager, Natural England Andy.Nisbet@naturalengland.org.uk

Contractor – Kat Bruce, NatureMetrics kat@naturemetrics.co.uk

Keywords – DNA, eDNA, sequencing, metabarcoding, monitoring, invertebrates, fish, vascular plants

Further information
This report can be downloaded from the Natural England Access to Evidence Catalogue: **http://publications.naturalengland.org.uk/** . For information on Natural England publications contact the Natural England Enquiry Service on 0300 060 3900 or e-mail **enquiries@naturalengland.org.uk**.

ISBN 978-1-78354-505-6

Development of DNA applications in Natural England

2016/2017

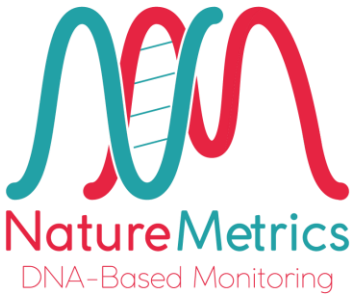


Table of contents

1. Introduction

Natural England is the Government's adviser for the natural environment, providing practical advice, grounded in science, on how best to safeguard England's natural wealth for the benefit of everyone. NatureMetrics is a DNA based biodiversity monitoring company that applies cutting-edge molecular methods to the challenges of tracking nature at large spatial and temporal scales, in remote or difficult ecosystems, or when target species are rare, inconspicuous, or poorly known.

This report details the results of a partnership project in 2016/17 between Natural England and NatureMetrics. The purpose of this work was to share the skills of each partner to test DNA based applications for the detection of species presence in environmental samples. The work herein acts as an initial pilot to guide further development and optimisation, and to identify those areas where the use of DNA based methods have the greatest potential for success.

Natural England has a wide range of needs around species survey and monitoring and needs to improve the efficiency and effectiveness of how we detect species in terrestrial, freshwater and marine environments. This project covers a range of circumstances that Natural England staff encounter as part of their routine operational work and where we believe innovative approaches have a strong role to play. Natural England has a considerable body of expertise around survey, monitoring and the need for species data in reporting, and operational work. This includes access to a substantial resource of semi-natural habitats and the ability to collect samples for analysis.

The partnership benefits both organisations as Natural England benefits from data obtained with cutting edge techniques and technology that could be used to potentially solve some of their taxonomic and ecological problems, while NatureMetrics benefits from the collection of samples by Natural England staff in order to develop real-world molecular applications.

High throughput molecular techniques for ecological management are still developing and the best practices are fluid. Despite being relatively immature compared to traditional techniques, these molecular tools provide an unprecedented view of nature that would otherwise be very difficult to obtain. But because the methods are new and the barriers to entry are high, real application of these techniques are rare. Massive strides are being made in the DNA monitoring world, but still many challenges remain, and this partnership provides an open and collaborative way of tackling some of these.

Conventional biodiversity monitoring: Limitations

Biodiversity survey and monitoring is limited by lack of resources as well as biases and errors in data collected. The collection of these data (depending on the target environment or taxonomic group) can be limited by either a sampling bottleneck and/or a taxonomic bottleneck. Combined, these issues mean that (a) the true distribution and status of most species is unknown, and (b) total community level biodiversity is rarely assessed (Yu et al. 2012; Ji et al. 2013). Instead proxies such as static indicator species are often used but these are not without significant problems (Zettler et al. 2013).

In addition to resource constraints, specific issues limiting survey and monitoring include (Figure 1.1):

1. Low detection rates
2. Low throughput
3. Inconsistency and inaccuracy among taxonomists

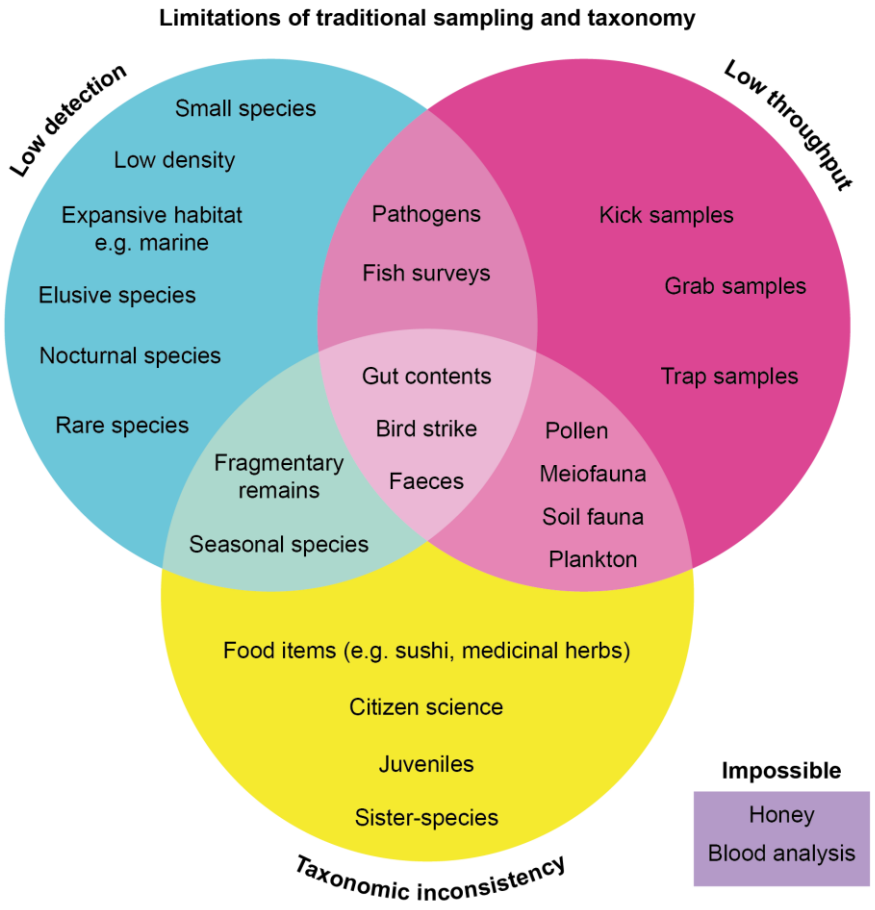


Figure 1.1 Limitations of current biomonitoring methods. These limitations are not mutually exclusive and can compound the difficulty of certain biomonitoring tasks.

Low detection rates

Finding rare, elusive, morphostatic or cryptic taxa is difficult by definition, and these taxa may be particularly important for management decisions (e.g. finding invasives before they establish or conserving rare populations before they become extinct). Low detection rates can result from an inability to distinguish/detect taxa (e.g. taxonomic inconsistency/inaccuracy - see below) or an inability to collect enough information (e.g. cryptic/elusive taxa).

Molecular techniques can detect identifiable DNA even when the organism itself is morphologically indistinguishable from others or is greatly outnumbered by other taxa. A recent large-scale comparative DNA metabarcoding study of macroinvertebrate samples from 18 stream sites across Finland showed that DNA metabarcoding identified more than

twice the number of taxa than the morphology-based protocol and at a high taxonomic resolution (Elbrecht et al. 2017b). Moreover, DNA detection is independent of life stage, which may yield greater opportunities for detection (e.g. from the presence of eggs and juvenile life stages).

Molecular techniques can detect DNA even when the organism itself may not be in the physical sample (eDNA). eDNA analysis has been shown to be more sensitive than traditional survey methods across a wide variety of rare, invasive, and cryptic organisms, including amphibians (Smart et al. 2015), reptiles (Hunter et al. 2015), fish (Jerde et al. 2013), mammals (Schnell et al. 2012), and even invertebrates (Cai et al. 2017).

Low throughput

Traditional taxonomy is not conducive to high throughput, which is necessary for long-term and routine monitoring of species diversity. The scalability of biotic surveys, particularly 'all taxon biodiversity inventories' (ATBI), have traditionally been limited in scale or scope due to the complications surrounding specimen sorting and species identification (Telfer et al. 2015).

Biomonitoring using passive traps and active sampling techniques can capture a huge diversity of organisms, but their extensive deployment is limited by the capacity to taxonomically process the samples. The routine identification of previously identified and curated species, while slow and low throughput with traditional taxonomy, could be achieved in a digital, reliable and rapid fashion using DNA technology (Ji et al. 2013). For example, 134 trap samples analysed using traditional taxonomy took 2,505 person-hours to identify only specimens belonging to seven indicator groups, while the same samples when analysed using DNA metabarcoding took 645-person hours to identify all of the specimens (even non-indicator taxa) (Ji et al. 2013). The standard traditional taxonomy and the DNA taxonomy datasets in this study exhibited statistically correlated α and β diversities and produced the same policy conclusions for conservation. Comparatively, the DNA dataset was produced more rapidly, accounted for a more comprehensive sample of the diversity, was less reliant on taxonomic expertise and was auditable.

Bioblitzes represent a fantastic opportunity to characterise a site's diversity in a short amount of time, however these are not currently routine or sustainable for biomonitoring due to taxonomic limitations. However, combining bioblitzes with DNA taxonomy can generate huge datasets in a rapid time frame. A one-day bioblitz conducted by Telfer et al. (2015) in combination with DNA taxonomy yielded about 3,500 specimens of animals, plants, and lichen. Similarly, a bioblitz of Lizard Island, Australia by 12 people over a 2 week period yielded 983 sequence records from 358 named fish species (Steinke et al. 2017), and these sequences helped to resolve species boundaries that were otherwise difficult to differentiate by traditional taxonomy. In another example a five-person team processed 41,650 museum Lepidoptera specimens (representing 12,699 species) in 14 weeks and sequenced them in the subsequent 6 months (Hebert et al. 2013). This scale is not feasible to replicate with the same level of accuracy with traditional taxonomy.

Unknown accuracy among taxonomists

Traditional taxonomy is difficult to standardise. There is a taxonomic bias in current survey and recording towards easily identifiable or charismatic taxa (birds, butterflies and vascular

plants). Additionally, taxonomists vary in their experience and skill levels, while dichotomous keys also vary and can be based on subjective characters that may be interpreted differently. Consequently, the same query materials can be identified differently by different taxonomists. These inconsistencies may be absent for highly trained taxonomists and the identification of large, easily identifiable species, but trained taxonomists may be an endangered species in themselves (Wägele et al. 2011). Inconsistencies and inaccuracies are a more significant problem for cryptic and neglected taxa such as the meiofauna (Fontaneto et al. 2015). For example, documenting the diversity of marine life is challenging because many species are cryptic, small, and rare, and belong to poorly known groups (Chen et al. 2011).

DNA sequences are digital and can be easily curated and databased. This information is easily auditable and verifiable, a characteristic vital for dispute resolution (Ji et al. 2013; Yu et al. 2012). DNA taxonomy is often used to confirm taxonomist misidentifications (Steinke et al. 2017).

The reduced reliance on expert taxonomists means that DNA taxonomy can be combined with citizen science programs to generate large datasets. For example, great crested newts were successfully monitored by volunteers with limited training in 91.3% of the samples they collected (Biggs et al. 2015). A larger scale citizen science project in California (CALeDNA) aims to characterise aquatic sediment samples in and around California to build up detailed and complex distribution maps.

DNA based monitoring: Limitations

While molecular biodiversity analyses has the potential to solve problems linked with conventional methods, these methods are still developing and thus have associated limitations. Here we discuss:

1. Inaccurate or incomplete databases
2. Imperfect abundance data
3. Lack of standardised methods

Inaccurate or incomplete databases

DNA based technologies referred to in this report return digital DNA sequence information, which needs to be translated into species. This translation requires databases that link DNA sequences to species of known identity. Sequences from unidentified organisms – obtained either by Sanger sequencing or high throughput sequencing – are compared against a reference database to make species identifications. Since the conception of DNA barcoding (Hebert et al. 2003) these databases have been generated with the aim to barcode all of life but there are two key issues with these reference databases: inaccuracy and incompleteness.

Inaccuracies in the reference databases originate from the misidentification of species, sample mix-ups in the laboratory, and contamination (Shen et al. 2013), or less common but unavoidable issues such as the prevalence of pseudogenes (see Song et al. 2007) or introgression (where whole genes can pass unchanged between species; see Whitworth et al. 2007). Reference databases rely heavily on the correct identification of the voucher specimen, mistakes made in the initial identification of the organism or potential

contamination prior to sequencing is cemented in these reference databases. These database errors persist and are compounded after subsequent use by other scientists.

To tackle the proliferation of erroneous sequences, which are not rare in uncurated and open source databases (i.e. Genbank – www.ncbi.nlm.nih.gov), requires sequence scrutiny and database curation (e.g. the Barcode of Life Database – BOLD – www.boldsystems.org; Ratnasingham and Hebert 2007). Increasingly the field is moving towards the use of quality controlled, curated databases (e.g. Somervuo et al. 2016).

Comprehensive DNA reference databases are essential for identification with DNA barcodes (Ekrem et al. 2007) but compiling them is an enormous task. Only a few taxonomic groups localised to specific geographical regions have complete databases (e.g. Romanian butterflies - Dincă et al. 2011; Lower Paraná River freshwater fish - Díaz et al. 2016; European marine fish - Oliviera et al. 2016). The vast majority of taxa and regions have incomplete reference databases. Furthermore, these reference databases typically focus on a few genes (i.e. *cox1* for animals, *matK* and *rbcL* for plants, ITS2 for fungi).

If a query sequence is not in the reference database then its species identity cannot be fully resolved. Nonetheless, it is usually possible to assign higher-level taxonomy (e.g. genus or family) based on the identity of the most similar reference sequences. Even with incomplete reference databases, probabilistic methods exist whereby taxonomic hierarchy can be combined with sequence information to determine probable species identity (e.g. Somervuo et al. 2016). Finally, given the digital nature of the query sequences and the fact that the reference databases are being continually populated, it is possible to reanalyse old data with updated reference databases to retrospectively update past analyses with new data.

Imperfect abundance data

A lot of biodiversity metrics require both diversity and abundance information. While detecting diversity is readily possible, abundance information is very difficult to obtain.

Single-species assays using qPCR have found positive relationships between copy number and species biomass (e.g. Takahara et al. 2012; Pilliod et al. 2013). For DNA metabarcoding studies the number of sequences belonging to individual species has been correlated with biomass/occupancy/rank abundance as a proxy for abundance. The relationship between sequencing output and abundance is not as strong as has been found in single-species assays (e.g. Elbrecht and Leese 2015; Elbrecht et al. 2017a; Hänfling et al. 2016; Klymus et al. 2017). Evans et al. (2016), for example, found “a modest, but positive relationship between species abundance and sequencing read abundance”. There are many confounding factors that can corrupt this relationship: prime among these is that amplification and sequencing efficiency can vary among species, which means that eventual sequence numbers will be depressed for those species where the amplification or sequencing inefficiency is low (e.g. through primer-template mismatches).

An interesting and important avenue of research for quantification via eDNA metabarcoding is the use of internal DNA standards (i.e. quantified short DNA fragments) that could be used to generate sample-specific standard curves, these in turn can be used to convert sequence read numbers into species-specific copy numbers (Ushio et al. 2017). The addition of these standard curves and the subsequent conversions showed “significant positive correlation with those determined by quantitative PCR, suggesting that eDNA metabarcoding with

standard DNA enabled the quantification of eDNA as accurately as quantitative PCR". While this improves the relationship between DNA copy number and metabarcoding output, it is important to remember that the relationship between species abundance and DNA copy number depends on the rate at which DNA is shed from the organisms, and this can vary with behavioural, environmental, ecological and seasonal factors. Thus, even if the metabarcoding output could accurately reflect the amount of each species' DNA in the sample, there may still not be a straightforward link to species abundance. Further research is required to address these issues.

Lack of standardisation

High throughput sequencing, despite being in its infancy, is a very active area of research that has resulted in a multitude of different workflows to sample, amplify, sequence, and analyse samples for biodiversity assessment. Currently no consensus has emerged regarding laboratory pipelines to screen for species diversity and infer species abundances from environmental samples, and many different approaches have been described for each of the following steps:

1. DNA capture
2. DNA preservation
3. DNA extraction
4. Means of dealing with inhibition
5. Choice of gene for analysing particular groups of taxa
6. Design / choice of primers
7. DNA polymerases and associated amplification and library preparation protocols
8. Sequencing platforms and associated parameters
9. Bioinformatic pipelines and taxonomic assignment using reference databases

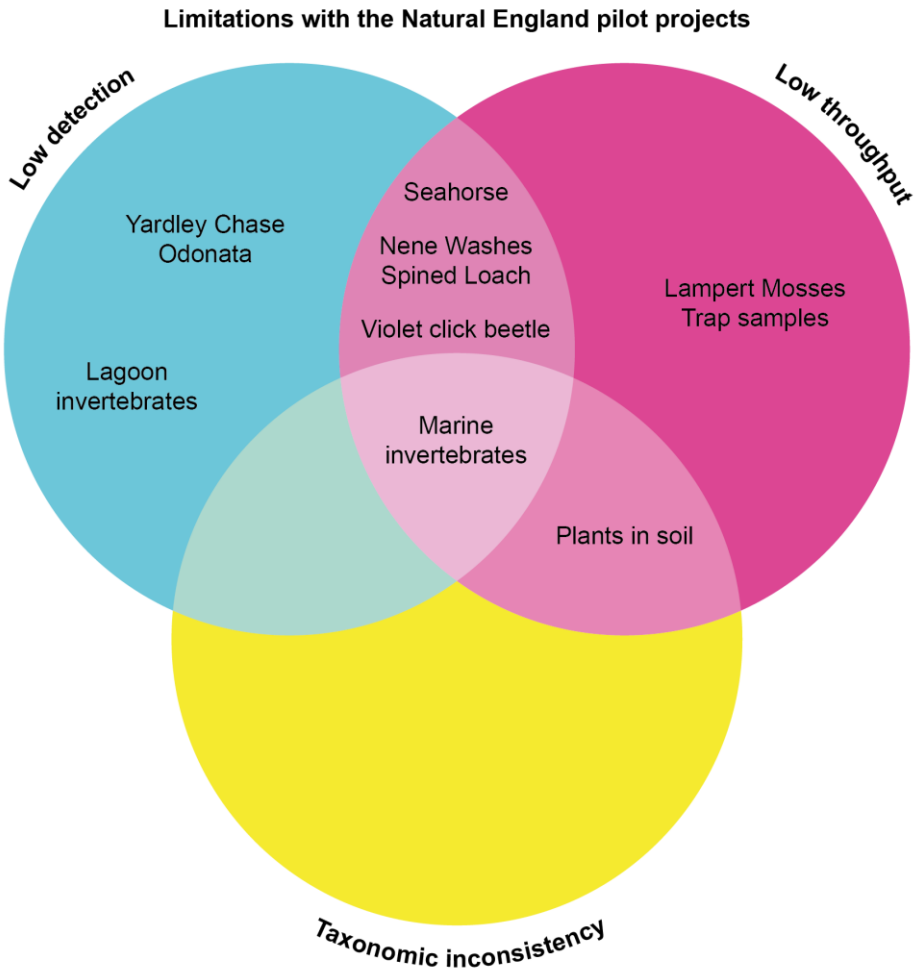
This lack of standardised methods reduces the generality of the findings but has spawned many comparative studies that compare different methods against one another (e.g. Spens et al. 2017), allowing for iterative optimisation of the process. However, it is also notable that in many cases the impact of different approaches is quite minimal in terms of the final metabarcoding output, suggesting that there may be multiple equally valid options at many stages in the metabarcoding process.

It is clear that more comparative studies and more collaboration is needed to determine how repeatable the different methods are and which method (if any one) is the best. It has been clear from our experience of this field that methodologies are in flux and that best practices are not yet settled upon, but this is a very active area of research. For example, there is an EU COST Action called DNAqua-Net, which is a network of active scientific experts, politicians, managers and stakeholders with the aim of developing new genetic tools for bioassessment of aquatic ecosystems in Europe (Leese et al. 2016). Within this network there is a series of working groups developing aspects of the processes listed above, among which is WG3 - Field and Lab protocols (led by Kat Bruce and Emre Keskin). There is currently a strong drive to improve standardisation of DNA based assessment methods while still maintaining flexibility, which could be facilitated by some degree of proficiency testing (as is done with the great crested newts).

Whether DNA technology is seen as a threat to traditional taxonomy or an aspect of its rejuvenation (Mallet & Willmott 2003), there is no denying that well-developed, sensitive and

high throughput DNA methods would at least complement traditional methods. Here we report on a group of pilot projects that have examined the application of DNA based techniques where traditional sampling and identification techniques would otherwise be difficult (Figure 1.2).

Figure 1.2 Limitations of current biomonitoring methods specific to the Natural England pilot projects.



2. Deadwood invertebrates
Violet click beetle - Bredon Hill/Windsor Forest

Introduction

The violet click beetle *Limoniscus violaceus* is an endangered European endemic species (Hyman & Parsons 1994; Nieto & Alexander 2010). The species is widely distributed in Europe but is severely fragmented and rare throughout its range (Nieto & Alexander 2010). In Britain this species is known from only three woods, all of which have been designated as Special Areas of Conservation (SAC): Dixton Wood, Windsor Forest and Great Park, and Bredon Hill. The species is protected by Annex II of the EU Habitats Directive, Wildlife and Countryside Act 1981, and the Natural Environment and Rural Communities Act 2006, but, as with the majority of saproxylic beetles, detailed information on their habitat requirements is still lacking, thus hindering efficient conservation information (Gouix et al. 2015). Expert opinion suggests that the contemporary distribution of this species in Britain is narrower than before owing to poor management of suitable habitat, but there is "insufficient evidence to support" this and "confidence in current assessment is low", indeed all conclusions made for the survival of this species have moderate to low reliability with little actual data (Joint Nature Conservation Committee 2007).

These beetles have the narrowest of ecological niches; they are a saproxylic species that rely on undisturbed, ancient and decaying beech *Fagus sylvatica* or ash *Fraxinus excelsior* trees where the larvae grow in black sodden decaying mulch that forms inside (Whitehead 2003). As part of the saproxylic biosphere these beetles are presumed to play an important role in nutrient cycling and ecosystem function but the extent of this role is unknown because the beetles and their presumed habitats are so rare (Nieto & Alexander 2010). But where it has been more extensively researched, it is considered as an umbrella species for other saproxylic fauna (Gouix et al. 2015).

Current sampling methods for this cryptic species require damage/destruction of habitats that may, but probably won't, contain individuals, as a result direct contact with violet click beetles is minimised (Joint Nature Conservation Committee 2007). Sifting and sorting of the wood mould is time consuming, requires highly expert taxonomists, who are few (Wägele et al. 2011), and is unlikely to be successful due to the low density of the imagoes and the indiscernible body parts and frass. While emergence traps could be a viable means of obtaining meaningful violet click beetle data (Gouix & Brustel 2012), this would require the trapping and disruption of adult and viable individuals, which could be more detrimental to the low density and rare British populations than beneficial.

Objective

To detect key species in deadwood substrates with the initial focus on violet click beetle *Limoniscus violaceus* and wider interest in Coleoptera and Diptera.

Approach

Here we trial a DNA-centric method for sampling traces of violet click beetle (frass, body parts, larvae) left behind in the wood mould. The aim being that presence of violet click beetle could be obtained without too much disruption of the habitat or adults and without the need to sort through wood mould.

Methodology

Field sampling

Key trees in Bredon Hill, Worcestershire and Windsor Forest, Berkshire, two of the known violet click beetle localities in the UK were selected. These trees were those that violet click beetle were thought to occur in. Samples of deadwood 'mould' were taken by plunging a soil corer into the soft tree mould at the hollowed out base of ash and beech trees (Figure 2.1). Upon collection, each individual core was placed in a ziplock bag. All equipment was bleach sterilised in between samples to avoid cross contamination of DNA.

Figure 2.1 Example of the tree hollows sampled for *Limoniscus violaceus*. Samples 1964, 1965, and 1966 were taken from this tree "321", from the front, left hand side, and rear of hollow, respectively. Image by David Heaver, Natural England.

Each sample was preserved in ~150 ml of absolute ethanol at 4 °C until DNA extraction (between 7-9 days).

eDNA metabarcoding

DNA was extracted from both the preservative solution and the wood mould sample itself. The preservative solution was filtered to capture any free cells or DNA (see Annex 1, eDNA sampling with filters) and DNA was then extracted from the filter (see DNA extraction: Filters). For the sediment, each sample was homogenised (see Tissue homogenisation) and DNA was extracted from two subsamples of the homogenate (see DNA extraction: homogenate). Thus we ended up with three DNA extracts from each sample: two from the homogenate and one from the filtered preservative solution. These DNA extracts were

purified to remove PCR inhibitors. (see DNA purification).

PCR amplification of a fragment of the cytochrome oxidase *c* subunit 1 (COI) was carried out using metazoan primers described by Leray et al. (2013) (see Polymerase Chain Reaction). All but one of the DNA extracts were successfully amplified, PCR bands appeared to be high yield and of the expected size. For sample 2010 ('Potential Tree Tag 02628S'), no PCR bands were obtained. This was the sample that yielded the least DNA (Table 2.1).

Table 2.1 Concentration of purified index PCRs

ID	Sample ID	Conc. (ng/µl)
1960	Bredon Hill - Sample 1 34 VCB tree (38 cm depth)	9.24
1961	Bredon Hill - Sample 2 34 VCB tree rear wall (35 cm depth)	9.13
1962	Bredon Hill - Sample 3 34 VCB tree front position (46 cm depth)	3.64
1963	Bredon Hill - Sample 4 34 VCB tree shallow (26 cm depth)	12.2
1964	Bredon Hill - 321 1st Sample front of hollow (18 cm core depth)	0.152
1965	Bredon Hill - 321 sample 2 LHS front rot (36 cm core depth)	11.9
1966	Bredon Hill - 321 sample 3 rear of hollow (20 cm core depth)	6.54
2003	Windsor Forest - VCB1	9.98
2004	Windsor Forest - VCB2	8.36
2005	Windsor Forest - VCB3	11.5
2006	Windsor Forest - VCB4	14
2007	Windsor Forest - VCB6	8.13
2008	Windsor Forest - VCB7	12.4
2009	Windsor Forest - H1 Potential 06983	3.67
2010	Windsor Forest - Potential Tree Tag 02628S	0.092
2011	Windsor Forest - H1 Potential Last one next to bin 6563?	7.32

Successful PCRs were then prepared for sequencing (see Library preparation). All samples were successfully indexed (Table 2.1), PCR bands appeared to be high yield and of the expected size. The concentration of two sequencing libraries (1964 and 2010) were lower than the rest, but this was compensated for by increasing their volume in the final pool. The final pool was sequenced on the 23rd March 2017 using an Illumina MiSeq 2 × 300 kit at 10 pM with a 5% PhiX spike in.

Sequences were processed using a custom bioinformatics pipeline (see Bioinformatics) and an OTU (operational taxonomic units, broadly akin to species) table was produced (Table 2.3).

No reference sequence was available for *Limoniscus violaceus* and efforts to sequence some old elytra were unsuccessful. Therefore, a fresh tissue sample was obtained from Germany. DNA was extracted from the sample and amplified for the COI gene as well as the V4 and V8 hypervariable regions of 18S rDNA. These were sequenced on the MiSeq and processed using the same custom bioinformatics pipeline (see Bioinformatics).

Results and discussion

Sequencing was successful for 15 of the 16 libraries. Sequencing of the library for 2010 failed; very few high quality sequences were generated (523 raw sequences of which there were 68 unique sequences) due to the failure to amplify the DNA in the first place.

Hundreds of thousands of sequences were obtained from the *Limoniscus violaceus* tissue material for each of the three DNA regions (COI, 18S V4 & 18S V8).

The MiSeq paired-end sequencing of the 15 samples yielded 3.75 million reads, of which 3.13 million reads (83.4%) passed our internal quality filter. A total of 1.45 million unique sequences were obtained, which were clustered into 344 OTUs. Of these, 319 were eukaryotic, 23 were bacterial, and 2 were unknown (Figure 2.2A). 189 Eukaryotic OTUs were assigned to Opisthokonta (a group that contains the Metazoa and Fungi), of which 100 were Metazoan and 77 were fungal (Figure 2.2B).

The percentage match for the vast majority of the OTUs to the reference databases was low (average = 86.6%; Figure 2.2C), which is likely due to the high proportion of unsequenced diversity present in wood mould. Typically we do not accept any species-level identifications unless they are at least 97%, however lower matches can still provide higher-level taxonomic information. Of the 100 Metazoan OTUs, only 24 were close matches to anything present in the reference database (>90% identity; Table 2.2); these included an earthworm *Eisenia fetida*, millipede *Proteroiulus fuscus*, woodlouse *Oniscus asellus*, and springtail *Folsomia candida*.

No violet click beetle DNA was found in the environmental samples. This might be because:

1. The samples are true negative samples - violet click beetle DNA was not present in our samples.
2. The assay results are false negatives - VCB DNA was present in the sample but the primers used were not sufficiently sensitive to amplify it. Traces of VCB DNA may have been shrouded by more prevalent sources of DNA, particularly fungi.

The chances of these samples being true negatives is quite high. VCB are rare, and the chances of traces being left behind in a detectable state in the small core used is quite low.

It is also possible that the medium itself is not conducive for eDNA analysis. While soil cores have been used in the past to analyse plant roots for ancient DNA and for more large-scale monitoring of plants (Fahner et al. 2016; and Section 8), no molecular study has assessed the diversity in wood mould. It is likely that the amount of invertebrate DNA left behind in the environment is much too low compared to other more prevalent sources of DNA that are also amplified by the primers used here. This is evidenced by the paucity of identifiable non-fungal or non-bacterial OTUs in the dataset: only a few invertebrate OTUs were found in the whole dataset.

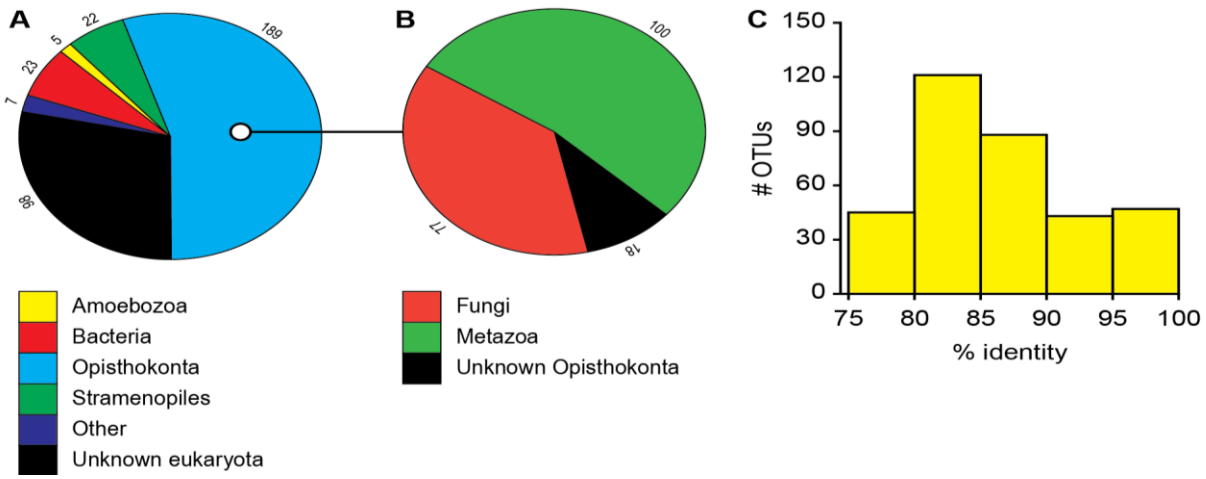


Figure 2.2 A. OTUs matching to broad taxonomic groups according to MEGAN. Numbers around the pie slices are the number of OTUs. 'Other' includes Malawimonadidae, Rhodophyta, Jakobida, and unknowns. B. Opisthokonta OTUs that were fungal, metazoan, or unknown. C. The proportion of high identity matches.

Technological readiness and further developments

It is difficult to gauge the success of this pipeline because the diversity present in the wood mould is dominated by taxa that are not represented on any reference database. Combining the DNA of this black-box of diversity with the likely minute trace amounts released into the environment by a target organism that may or may not be there makes for a difficult task. Here our plan was to use a set of universal and degenerate primers to capture diversity *in toto* and bioinformatically extract the VCB sequences (if present), but an alternative approach would be to design new more specific VCB primers to add another level of specificity from the start.

Since carrying out this pilot project, we have used the sequence data generated from the tissue sample to design some species-specific primers for violet click beetle. These will be tested and optimised henceforth.

Table 2.3 OTU table based on COI. Identifications and percentage identity based on the best match to online reference databases. Higher rank identifications were classified using the RDP database. Only Metazoan taxa that could be classified with more than 90% identity are shown.

Identity	% ID	1960	1961	1962	1963	1964	1965	1966	2003	2004	2005	2006	2007	2008	2009	2011
Amblyomma lepidum	94.6					342										
Arthropoda sp.	99							1710								
Atropacarus sp.	98.5			525												
Bdelloidea sp.	95							3025								
Chamaedrilus chlorophilus	100												1348			18341
Chromadorea sp.	99									824						
Chromadorea sp.	100					2340										
Chromadorea sp.	100					622										
Chromadorea sp.	97		2478													
Chromadorea sp.	100		2156													
Columba palumbus	99.4			3087												
Cryptops hortensis	99.4												2159			
Dendrodrilus rubidus	100								9287							27051
Eisenia fetida	100										89813	130431				
Folsomia candida	100										2061					
Marionina clavata	90.9												3031			
Mesenchytraeus pelicensis	100															11045
Oniscus asellus	100			3180												
Onychiuridae sp.	99.7										907				2548	
Porcellio sp.	100	1852	2213													
Proteroiulus fuscus	100							14733	1328							
Sciaridae sp.	100					532										
Stigmella tityrella	100															2485
Tullbergiidae sp.	99												2885			

3. Invertebrate assemblage
Terrestrial invertebrates - Lampert Mosses SSSI

Introduction

Lampert Mosses is a Site of Special Scientific Interest (SSSI) in Northumberland and part of the Border Mires, Kielder-Butterburn SAC. The site was designated as an SSSI because of its mire habitats and invertebrate assemblage. The invertebrate assemblage is largely restricted to ombrotrophic or soligenous mires and is characterised as having a large number of species from a variety of taxonomic groups but also being generally low productivity (Webb et al. 2017). Lampert Mosses supports more nationally rare and scarce invertebrate species than any other Northumberland peatland (English Nature 1994). The Lampert fauna includes nationally rare flies (e.g. *Spilogona depressiuscula* and *Coenosia paludis*), ground beetles (*Agonum ericeti* and *Carabus nitens*), butterflies (*Coenonympha tullia*), and spiders (*Semljicola caliginosus*).

Conserving these assemblages requires baseline data and monitoring. While invertebrates can be sampled with active trapping (i.e. sweep netting, ground searching, pond netting, etc.), the low density of invertebrates in Lampert Mosses would make detection probabilities low and the sampling effort high. Instead, a more comprehensive sampling of these assemblages would involve passive traps such as malaise or pitfall traps (Webb et al. 2017).

If long-term monitoring using passive traps was adopted to routinely monitor Lampert Mosses, then the sampling bottleneck would turn into a taxonomic bottleneck. Increasing the number of traps would linearly increase the amount of work required of a taxonomist, such that the program would quickly become unscalable. Here DNA metabarcoding of trap samples could act as a suitable alternative to expert taxonomists. Unlike traditional taxonomic work, the average amount of people hours required to identify specimens in a sample reduces with scale. DNA extraction, amplification, purification, and sequencing steps can all be done in multiples and are actually least efficient when done with few samples.

Aim

To test the use of DNA in providing information of terrestrial invertebrate assemblages from trapped samples.

Approach

Here we test the use of DNA metabarcoding on three Lampert Mosses trap samples (two malaise and one pitfall).

Methodology

Field sampling

A malaise trap was deployed and sampled for two consecutive periods: (1st - 17th August 2016 and 17th August - 16th September 2016). A pitfall trap set for August (1st-31st Aug 2016) was sampled once. This was not the ideal timing, but provides a test of what data can

be collected through this approach.

The malaise trap (Figure 3.1) was initially operated using 95% ethanol, but this evaporated partially so the final sample was taken using 100% propylene glycol (propane-1,2-diol). Both 95% ethanol and 100% propylene glycol give excellent DNA preservation in samples. The pitfall traps (Figure 3.1) were operated as a regular 3 x 3 array of 9 pits with mesh covers to prevent small mammal capture and a 'lid' of plywood to prevent dilution from rainfall. Each pit was ¾ filled with 100% propylene glycol. These traps are easily serviced by changing the sample pot (malaise trap) or emptying the pitfalls and refilling with clean propylene glycol. Resultant samples were stored in 95% ethanol before passing on for extraction of DNA and analysis.

Figure 3.1 Malaise and pitfall traps at Lampert Mosses. Images from Natural England

DNA metabarcoding

The trap samples were homogenised to facilitate DNA extraction (see Annex 1, Tissue homogenisation: Trap samples). The pitfall trap sample ('pitfall trap sample 1st-31st Aug 2016') contained a large *Pterostichus nigrita* individual. This was removed from the sample prior to homogenisation, with just a single leg reintroduced, so that this one specimen didn't swamp the genetic signal from the other specimens. Each sample resulted in a single homogenate that was too large for a single DNA extraction. To improve the representativeness of the DNA extraction with respect to the physical sample, multiple DNA extractions were performed on the homogenate and subsequently pooled (see DNA extraction: homogenate). In total, six, nine, and nine DNA extractions were performed for the pitfall and two Malaise traps, respectively. For each of the three samples, equal volumes of these DNA extracts were pooled into a final working DNA extract.

PCR amplification of a fragment of the cytochrome oxidase *c* subunit 1 (COI) was performed using metazoan primers described by Leray et al. (2013) (see Polymerase Chain Reaction). All PCRs for the two Malaise traps were successful. PCR bands appeared to be high yield and of the expected size. No repeat reactions were necessary. For the pitfall trap, the first round of PCRs were unsuccessful owing to suboptimal PCR conditions. Subsequent optimisation of the $MgCl_2$ concentration following Roux (1995) resulted in successful amplification of the target marker.

Successful PCRs were then prepared for sequencing (see Library preparation). All samples

were successfully indexed (Table 3.1), PCR bands appeared to be high yield and of the expected size. The final pooled sequencing library was sequenced on the 6th of March 2017.

Sequences were processed using a custom bioinformatics pipeline (see Bioinformatics). A species by sample OTU table was generated (Table 3.2).

Table 3.1 Concentration of purified index PCRs.

ID	Sample ID	Conc. (ng/µl)
1909	Pitfall trap Sample 1st - 31st Aug 2016	9.76
1910	Malaise Sample 1st - 17th Aug 2016	13.3
1911	Malaise Sample 17th Aug - 16th Sep 2016	13.9

Results and discussion

The MiSeq paired-end sequencing of the 3 samples yielded 324,655 reads, of which 81% passed our internal quality filter. A total of 194,183 unique sequences were obtained, which were clustered into 39 OTUs (excluding bacterial, archeal, and contaminant OTUs). These OTUs belong to 19 families and 30 genera (Figure 3.2). Only five and one OTUs could not be identified to species and genus level, respectively, likely due to the incompleteness of the reference dataset. The vast majority of the malaise trap diversity were Diptera (29/30 OTUs), with only a single Trichopteran species in the second malaise sample. The three samples (1909, 1910, and 1911) each contained 10, 20, and 26 unique OTUs respectively, with a larger proportion of OTUs shared between the two Malaise trap samples; this is unsurprising given that these two samples came from the same trap in a time series (1st - 17th Aug 2016 vs. 17th Aug - 16th Sep 2016). Only three OTUs - *Myrmica ruginodis*, *Molophilus occultus*, and *Pegoplata infirma* - were shared between all three samples.

While this level of diversity is quite low compared with typical pitfall and malaise traps from the British Isles, the sampling location is known to be relatively species-poor so this is not an unexpected result.

Technological readiness and further developments

This pipeline is well developed, although there remain some improvements that can be made. Some research has come to light in recent months suggesting that size sorting of invertebrates is even more important than previously thought in terms of allowing the recovery of small/rare taxa (Elbrecht et al. 2017a), so in future we would carry out this step more thoroughly and consider extracting the DNA separately from each size class to ensure representation of the smallest organisms.

The Qiagen DNeasy DNA extraction kits worked well for extracting DNA from these samples, and no clean-up of the DNA was required prior to PCR. The advantage of the Qiagen extraction kits is that they can be integrated into automated systems where high throughput is required, which is an important consideration if these methods are to be adopted for widespread use.

In future we would be likely to use indexed primers for the initial PCR, as these build in resilience against laboratory contamination (if contamination occurs then it can be traced to

a particular index, and this can be retired from use, allowing contaminant sequences to be screened out from future sequencing runs). This will entail some adaptation of the library preparation methodology.

Table 3.2 OTU table based on COI

OTU ID	Species	PFT_1909	MLT_1910	MLT_1911
OTU14	*Lumbricus rubellus*	7731		
OTU08	*Myrmica ruginodis*	6947	2134	928
OTU24	*Molophilus occultus*	2179	1739	306
OTU132	*Myrmica rubra*	498		
OTU115	*Myrmica scabrinodis*	472		
OTU10	*Tomocerus longicornis*	415		
OTU166	*Pterostichus nigrita*	378		
OTU173	*Rhaphium longicorne*	342		
OTU240	*Pegoplata infirma*	230	228	97
OTU11	*Delia platura*		1160	141
OTU51	*Phaonia* sp.		1143	368
OTU82	*Euphylidorea meigenii*		587	810
OTU136	*Linnaemya vulpine*		553	237
OTU158	*Graphomya transitionis*		400	
OTU167	*Phaonia incana*		308	404
OTU251	*Hilara* sp.		203	
OTU95	*Helophilus pendulus*		155	896
OTU546	Chironomidae		104	
OTU434	*Platycheirus peltatus*		101	
OTU353	*Melanostoma scalare*		89	183
OTU126	*Platycheirus albimanus*		86	505
OTU639	*Hydrotaea irritans*		75	
OTU547	*Coenosia means*		64	
OTU967	*Psychoda phalaenoides*		63	
OTU498	*Tipula oleracea*		59	
OTU122	*Eristalis pertinax*			547
OTU117	*Graphomya maculata*			491
OTU163	*Limnephilus coenosus*			369
OTU203	*Metriocnemus* sp.			358
OTU236	*Delia florilega*			294
OTU320	*Dilophus febrilis*			171
OTU447	*Botanophila fugax*			152
OTU401	*Hydrellia maura*			151
OTU367	*Tricyphona* sp.			146
OTU294	*Camptocladius stercorarius*			128
OTU449	*Tipula melanoceros*			115
OTU552	*Cyphon hilaris*			80
OTU974	*Tipula confusa*			55
OTU675	*Lasiomma* sp.			48

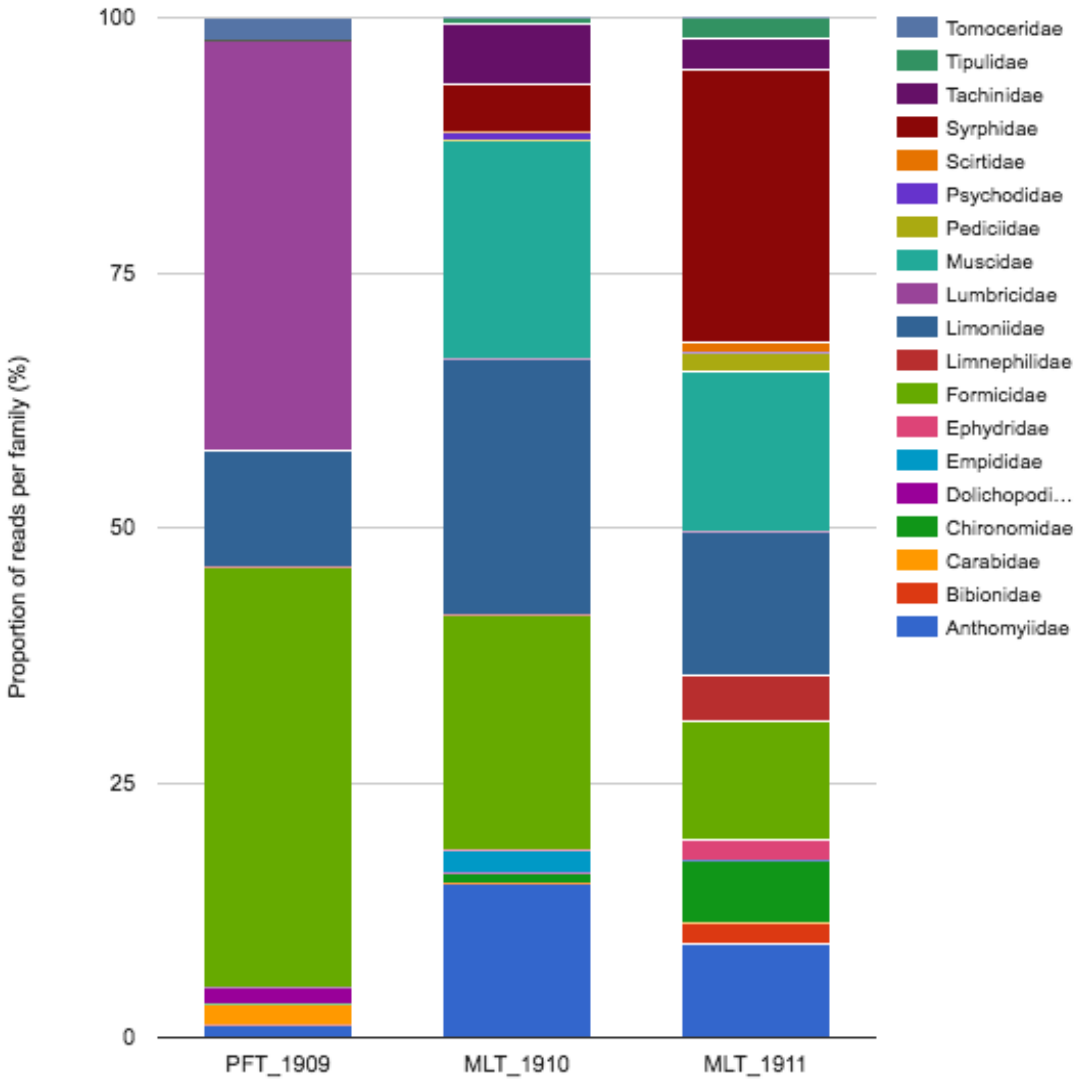


Figure 3.2 Community composition of the pitfall trap (1909) and the two Malaise traps (1910 and 1911) at the family level based on the relative number of reads matching those families. Different profiles are apparent for two trap types.

4. Freshwater vertebrate assemblages Spined loach - Nene Washes

Introduction

Nene Washes is a SSSI, SAC, Special Protection Area (SPA), and Ramsar site in Cambridgeshire. The seasonally flooded wet grassland is an important habitat for breeding and wintering waders and wildfowl, and the plant and animal life associated with its network of ditches. Importantly, the site has been designated an SAC because of its population of spined loach *Cobitis taenia*.

The spined loach is a small bottom feeding fish protected under Appendix 3 of the Bern Convention (Council of Europe 1979) and the Annex II of the EC Directive 93/43/EEC (European Commission 1992; Joint Nature Conservation Committee 2010). The species is restricted to a microhabitat consisting of fine, well oxygenated sediments and macrophytes used for cover and spawning.

The species is known from only five east-flowing river systems in eastern England: the Rivers Trent, Welland, Witham, Great Ouse, and Nene (Perrow & Jowitt 1999). Moreton's Leam, one of the Nene Washes drainage channels, supports the highest density of spined loach in the UK, but the true extent of spined loach distribution among the smaller drainage ditches in Nene Washes is unknown due to difficulty in surveying such ditches for a cryptic species often buried within the sediment. Apart from the fact that the fish don't want to be found, sampling methods are not universally useful: standard seine netting is unsuitable for habitats covered in patches of vegetation, electrofishing is unsuitable in depths over 1 m, bottle traps don't tend to catch bed dwelling spined loach, drag nets are useful in main channels but become blocked by vegetation, and hand nets require a lot of sampling effort (Perrow & Jowitt 1999).

An alternative sampling method is a non-organismal method - eDNA. eDNA may be suitable for sampling elusive and cryptic aquatic species as detectable genetic material is left behind without any obvious signs of the biological source (Thomsen & Willerslev 2014).

Aim

To test the eDNA metabarcoding approach for detection of spined loach and other associated taxa in fenland ditches.

Approach

Here we test an eDNA sampling method using filters to concentrate and capture eDNA and analyse it using DNA metabarcoding to effectively and efficiently characterise the fish community. We also test the reproducibility of various stages in the pipeline to see how consistent sampling and laboratory replicates are.

Methodology

Field sampling

eDNA was sampled at five locations (Figure 4.2) on Nene Washes on the 15th of September 2016 using Sterivex filters (see Annex 1,eDNA sampling with filters). The water sampling was aided by the use of a 2.5 m garden cane with a sterile, disposable plastic cup attached in a wire holder.

eDNA metabarcoding

A separate DNA extraction was performed for each filter (see DNA extraction: Filters). Each site was represented by at least one filter (Table 4.1 - '1st filter'), with additional filters (Table 4.1 - '2nd filter' and '3rd filter') analysed from some sites to assess the repeatability of the sampling process.

PCR amplification of a short region of the 12S rRNA gene (~230 bp) was carried out using the MiFish primers (Miya et al., 2015), which are designed to target fish (see Polymerase Chain Reaction). For each sample, two of the three replicates were successful. PCR bands appeared to be high yield and of the expected size.

Successful PCRs were then prepared for sequencing (see Library preparation). Each DNA extract was prepared as a separate sequencing library (9 libraries in total). Two libraries (1750A and 1751B) were prepared in duplicate to assess the repeatability of the sequencing step. All samples were successfully indexed (Table 4.1), PCR bands appeared to be high yield and of the expected size. The final library was sequenced on the 27th October 2016 using an Illumina MiSeq 2 × 250 kit at 10 pM with a 5% PhiX spike in.

Sequences were processed using a custom bioinformatics pipeline (see Bioinformatics), generating a species-by-sample OTU table (Table 4.2).

Results and discussion

The MiSeq paired-end sequencing of the 11 samples yielded 2.12 million reads, of which 492,312 reads (23.2%) passed our internal quality filter. Despite a low proportion of quality filtered sequences, a total of 27,630 unique sequences were obtained, which were clustered into 16 OTUs (excluding bacterial, archeal, and contaminant OTUs). These included 13 fish taxa, two bird species, and rabbit (Table 4.2; Figure 4.1). Importantly, spined loach was found in two of the five sampling sites: Nene Washes 1 (TL 26679834) and Nene Washes 4 (TL 32209935; Figure 4.2). These sites were both located on the largest ditch. All observed taxa are expected of this area, with the exception of the Atlantic Salmon OTU identified from Nene Washes 4. This is likely an environmental contaminant from wastewater or oils used in fishing bait.

The level of consistency between filters used in the same site is high (Figure 4.3A), with similar read numbers per OTU. Nonetheless there are some cases where OTUs are recovered from one filter but not the other, suggesting that multiple sampling replicates will capture a closer approximation to the true diversity at a given site. Filtering a greater volume of water through a single filter may also have the same effect, particularly if subsamples of water are pooled from different point locations at the sampling site to maximise spatial

representativeness of the filtered sample. It is also possible that the observed variation is partly due to stochasticity during PCR, which can occur when the target DNA is present at very low concentrations. The difference between filters may be due to local sampling differences, for example some filters may become clogged more quickly than others or there may be a higher amount of sediment in one filter compared to the next.

An extremely high level of consistency was observed between library replicates (Figure 4.3B), with any small discrepancies in read numbers attributable to stochastic effects during sequencing.

Technological readiness and further developments

This pipeline is well-developed and already represents an effective tool for determining the fish species present in freshwater habitats. A key factor in detection of fish is the spatial representativeness of the water samples and the amount of water that is filtered, and more work is needed to determine the optimal sampling strategy for different types of water body. The Sterivex filters remain somewhat restrictive because they clog rapidly, which limits the volume of water that can be passed through a single filter. We have conducted tests on some alternative filter units that enable greater volumes of water to be processed, and these will feature in any further work. As with other analyses, we intend to move towards the use of indexed primers for the initial PCR to build into the pipeline an additional layer of resilience to contamination, and this will involve modification of the library preparation protocol.

Table 4.1 Concentration of purified index PCRs. Each site is represented by at least one filter (‘1st filter’), an additional four DNA extracts from different filters (‘2nd filter’ and ‘3rd filter’) were analysed separately to assess the repeatability of the sampling process. A replicate of the library preparation stage (‘rep. 2’) was analysed separately to assess the repeatability of the library preparation process.

ID	Sample ID	Conc. (ng/µl)
1750Ai	Nene Washes Site 3 TL 30639906 - 1st filter - rep. 1	12.5
1750Aii	Nene Washes Site 3 TL 30639906 - 1st filter - rep. 2	18.4
1751A	Nene Washes Site 2 TL 26839852 - 1st filter	16.2
1751Bi	Nene Washes Site 2 TL 26839852 - 2nd filter - rep. 1	16.7
1751Bii	Nene Washes Site 2 TL 26839852 - 2nd filter - rep. 2	20.4
1752A	Nene Washes Site 5 TL 31889947 - 1st filter	14.1
1753A	Nene Washes Site 4 TL 32209935 - 1st filter	18.3
1753B	Nene Washes Site 4 TL 32209935 - 2nd filter	14.8
1754A	Nene Washes Site 1 TL 26679834 - 1st filter	23.5
1754B	Nene Washes Site 1 TL 26679834 - 2nd filter	11.5
1754C	Nene Washes Site 1 TL 26679834 - 3rd filter	12.7

Table 4.2 OTU table based on 12S. The individual filters and replicates are collapsed by site.

Species	Common name	%ID	NW1	NW2	NW3	NW4	NW5
Abramis brama	Common bream	100	3484	0	0	1570	0
Alburnus alburnus	Common bleak	99.5	7209	0	0	0	0
Anguilla anguilla	European eel	100	4392	0	0	0	0
Cobitis taenia	Spined loach	100	2303	0	0	2042	0
Esox lucius	Northern pike	100	2015	0	933	3602	0
Gasterosteus aculeatus	Three-spined stickleback	100	1917	0	0	0	0
Gymnocephalus cernua	Ruffe	99.5	0	0	0	488	0
Leuciscus rutilus	Common roach	100	39417	0	22076	20965	0
Perca fluviatilis	European perch	100	37408	13049	26973	38402	0
Pungitius pungitius	Ninespine stickleback	100	3877	0	0	5696	45833
Salmo salar	Atlantic salmon	100	0	0	0	564	0
Scardinius erythrophthalmus	Common rudd	100	1742	0	0	2397	0
Tinca tinca	Tench	100	6589	39919	0	1272	0
Anas platyrhynchos	Mallard	100	0	18347	0	0	0
Corvus sp.	Corvid	99	0	2623	0	0	0
Nesolagus netscheri	Rabbit	100	0	0	0	0	554

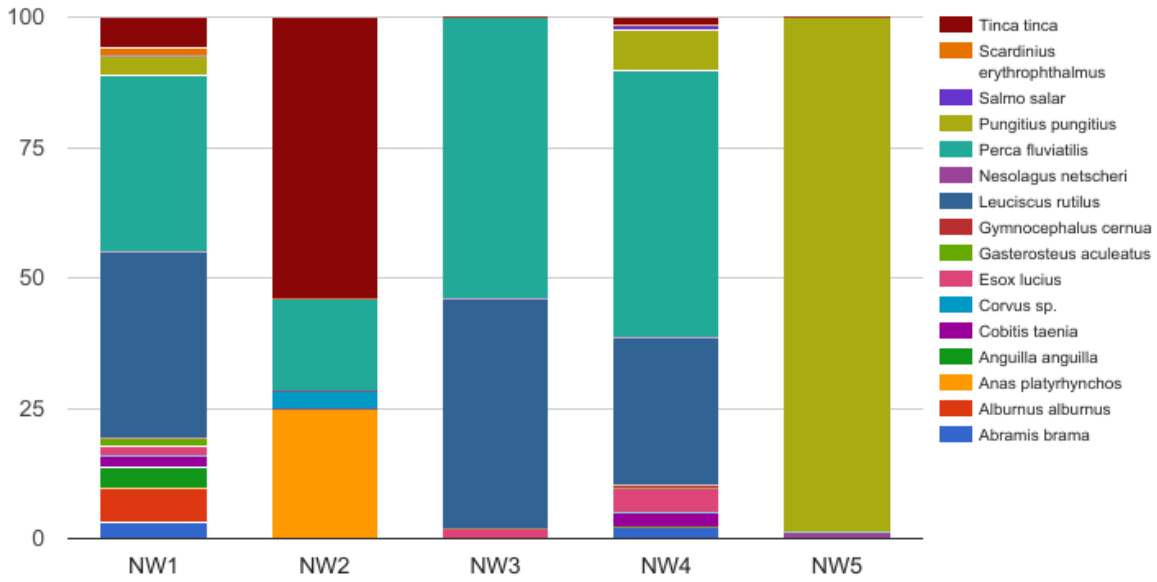
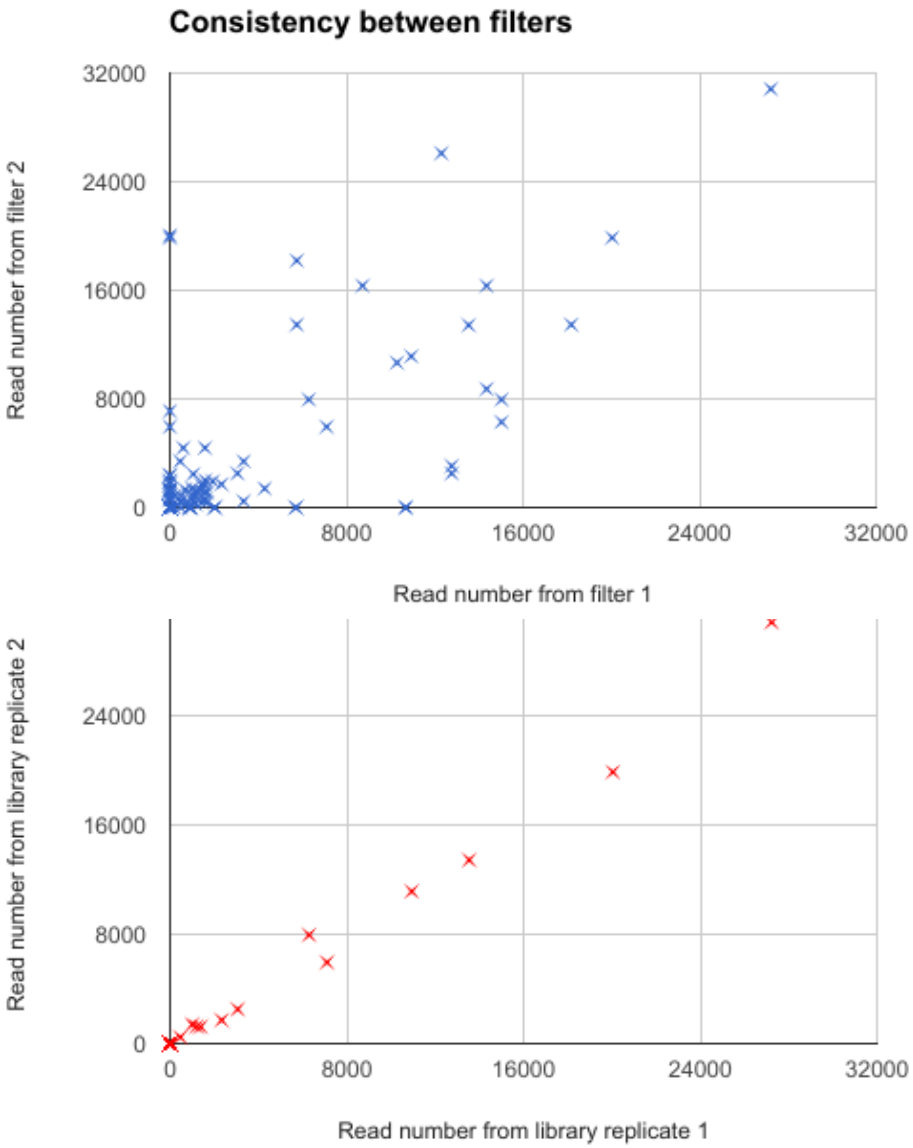


Figure 4.1 Community composition of the five site samples at species level. Bar sizes are based on the relative number of sequence reads matching those species.

Figure 4.2 Locations of the 5 samples. Multiple filters were used as subsamples at each site. Spined loach *Cobitis taenia* was found in both Nene Washes 1 and Nene Washes 4. Imagery ©2018 Google, Map data ©2018, DigitalGlobe, Getmapping plc, Infoterra Ltd & Bluesky.

Figure 4.3 The consistency of the number of reads per OTU compared between (A) filter replicates from the same sample site (blue), and (B) replicate library preparations of the same PCR product (red).



5. Freshwater invertebrate assemblage Odonata - Yardley Chase

Introduction

This project arose from discussion with the Natural England Area Team in relation to problems they were having monitoring SSSI features. A dragonfly (Odonata) assemblage is a feature of several SSSIs in their area and requires multiple visits to record the species of adult dragonflies and damselflies occurring around the ponds or small lakes. This work would either have been commissioned from entomological survey contractors or done by volunteer recorders.

The site chosen was Yardley Chase SSSI (in Buckinghamshire and Northamptonshire) where notable assemblages of Odonata occur in small borrow pits that surround derelict munition storage bunkers on the site (Figure 5.1). The land was used as munitions storage depot by the MOD and is now a training facility for the MOD. The Yardley Chase dragonfly community includes at least 13 breeding species, of which the ruddy darter *Sympetrum sanguineum* is a nationally vulnerable species.

Aim

To test a DNA approach for the detection of Odonata assemblage using water sampled from SSSI lake/pond.

Approach

Here we test an eDNA sampling method coupled with DNA metabarcoding to determine whether dragonflies and damselflies at Yardley Chase can be detected using molecular methods.

Methodology

Field sampling

eDNA was sampled on the 2nd of September 2016 using Sterivex filters (see Annex 1, eDNA sampling with filters). Six 100 ml subsamples of pond water were taken from different points around the pond, with adjacent pairs of subsamples pooled for filtering (i.e. the six subsamples were filtered through three filters).

eDNA metabarcoding

DNA was separately extracted from each of these (see DNA extraction: Filters). PCR amplification of two fragments of the cytochrome oxidase *c* subunit 1 (COI) using Arribas et al. (2016) and Leray et al. (2013) primers was performed to target Metazoa (see Polymerase Chain Reaction). All replicate PCRs were successful for the Arribas primers, PCR bands appeared to be high yield and of the expected size. Amplification using the Leray primers required a 1:5 dilution of the DNA.

Figure 5.1 Sampling site at Yardley Chase SSSI, outlined in red. Imagery ©2018 Google, Map data ©2018 Google.

Successful PCR products were then prepared for sequencing (see Library preparation), with each filter replicate processed as a separate library. All samples were successfully indexed (Table 5.1), PCR bands were of the expected size. The concentration of the final purified libraries was low for both genes, but these were pooled in a higher volume to compensate. The final libraries were sequenced on the 9th of December 2016 for Arribas, and on the 23rd March 2017 for Leray.

Table 5.1 Concentration of purified index PCRs.

ID	Sample ID	Conc. (ng/µl)
1749	Yardley Chase Pond - Arribas	1.36
1749	Yardley Chase Pond - Leray	4.73

Sequences were processed using a custom bioinformatics pipeline (see Bioinformatics). A species by sample OTU table was generated for each gene (not shown).

Results and discussion

The MiSeq paired-end sequencing of the Yardley Chase Pond Arribas sequencing library yielded 169,689 reads, of which 64,986 reads (38.3%) passed our internal quality filter. The

MiSeq paired-end sequencing of the Yardley Chase Pond Leray sequencing library yielded 207,904 reads, of which 165,385 reads (79.5%) passed our internal quality filter. A total of 48,512 and 65,535 unique sequences were obtained from the Arribas and Leray sequencing libraries, respectively, which were clustered into 51 and 36 OTUs (excluding bacterial, archeal, and contaminant OTUs).

Only 7% of these OTUs had a match above 97% identity (Figure 5.1) in the reference database: two species of fungi, a mayfly, a copepod, an ostracod and a midge. This indicates that the vast majority of diversity sequenced in this analysis is unrepresented in genetic databases. The majority of these sequences appear to be prokaryotic.

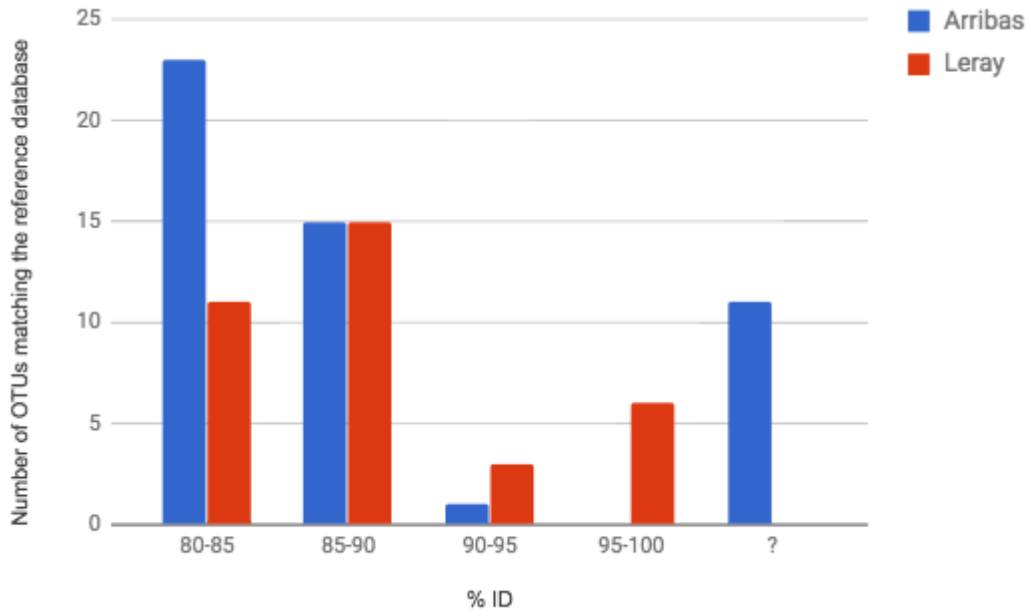


Figure 5.1 The percentage identity of the OTUs matching to GenBank.

Both primer sets suffer from the same key issue, namely amplification of prokaryotic diversity, which is more concentrated in the water column than is environmental DNA from metazoans. Moreover we have found that water sampling tends to concentrate zooplankton, and the metazoan sequence output is dominated by these species.

While target invertebrate DNA may be present in the filter, it is completely obscured by the more prevalent prokaryotic DNA in the sample. This is supported by our experience in testing eDNA samples with different primer sets that yield dramatically different results. For instance, in one experiment (unconnected with the Natural England pilot work) we sequenced the same sample with both 12S MiFish primers (target: fish) and the Leray COI primers (target: metazoans). A diverse fish community was detected with the MiFish primers, confirming the presence of fish DNA in the samples, yet no fish sequences were obtained using the Leray primers.

Besides the Leray and Arribas COI primers, we have also tested the BF / BR primers described by Elbrecht and Leese (2017), which are specifically designed to target freshwater invertebrates. While these have been shown to produce good data from homogenised tissue samples, when applied to eDNA samples they perform similarly to the primer sets tested

here, with the output dominated by prokaryote sequences.

Because of the nature of the COI gene, it does not seem possible so far to create primers that will amplify a broad spectrum of metazoan diversity without also amplifying microbial diversity when this is the dominant source of DNA. This is unfortunate because the COI database is the most comprehensive animal marker database. An alternative method would be to design specific primers for smaller groups, or to use an alternative marker (e.g. 16S rDNA - Klymus et al. 2017).

Technological readiness and further developments

As yet eDNA metabarcoding for freshwater invertebrates in ponds is very much in its infancy. While it can be used to characterise faecal invertebrate communities (e.g. Bohmann et al. 2011), aquatic invertebrates present a bigger challenge because of the prevalence of non-target DNA and organisms concentrated by the filtering process, and the limited amount of ‘eDNA’ shed by invertebrates. There are a few areas for further development in the invertebrate eDNA metabarcoding including improved sampling strategies and more specific primer design.

6. Detecting key species in saline lagoons Lagoon invertebrates - Bembridge/Saltons

Introduction

Brackish lagoons are an ephemeral habitat that support a unique array of flora and fauna. This rare habitat supports highly specialised plants and invertebrates that are adapted to the variably brackish conditions, and often these species are unique to these habitats. Coastal lagoons are a scarce and rare habitat and are protected within Special Areas of Conservation and under Annex 1 of the Habitats Directive.

The Solent and Isle of Wight Lagoons SAC are a series of coastal lagoons with variable salinity and substrates. The lagoons are some of the most important of their kind in Britain with populations of rare species including the lagoon sand shrimp *Gammarus insensibilis* and starlet sea anemone *Nematostella vectensis*. Both species are protected under Schedule 5 of the Wildlife and Countryside Act 1981, and are species of principal importance for the purpose of conservation of biodiversity under the Natural Environment and Rural Communities Act 2006.

The difficulty in monitoring these species is that they are small, elusive, aquatic invertebrates. Complicating sampling further is the fact that they associate with either dense algae or sediment. Lagoon sand shrimp associate with mats of green algae, seaweeds and seagrass making them more elusive. Starlet sea anemones, apart from being a maximum of 15 mm and translucent, typically bury themselves so that only their 1.5 mm crown shows or is completely retracted when not feeding. Both species can be difficult to find in the wild using conventional sampling techniques, and this limits what we know about their ecology and population status.

Aim

To test DNA approach for detection of starlet sea anemone *Nematostella vectensis* and lagoon sand shrimp *Gammarus insensibilis* from water samples.

Approach

Here we test an eDNA sampling method coupled with DNA metabarcoding to determine the presence of lagoon sand shrimp and starlet sea anemone in two brackish lagoons, Bembridge Harbour lagoon (Isle of Wight) and Saltern Lagoon (near Lymington, Hampshire).

Methodology

Field sampling

eDNA was sampled using Sterivex filters (see Annex 1, eDNA sampling with filters). Each sample consisted of 3 separate filter replicates.

eDNA metabarcoding

DNA was separately extracted from each filter (see DNA extraction: Filters) and the DNA extracts from the three filter replicates per sample were then pooled. A whole suite of primers were used to give us the best possible opportunity of detecting the target taxa. These included several that targeted regions of the cytochrome oxidase *c* subunit 1 (COI) (Folmer et al., 1994, Elbrecht et al., 2017, Arribas et al., 2016 and Leray et al, 2013), and two that targeted hypervariable regions of 18S rDNA (V4 and V8; Bradley et al., 2016). Specific details of the reactions can be found in the detailed methods section for Polymerase Chain Reaction. All of the replicate PCRs for each of the six markers were successful. PCR bands appeared to be high yield and of the expected size.

Successful PCR products were then prepared for sequencing (see Library preparation) with a separate sequencing library for each sample and marker combination. All samples were successfully indexed (Table 6.1), PCR bands appeared to be high yield and of the expected size.

The final libraries were sequenced in two parts. COI - Leray and COI - Arribas libraries were sequenced on the 6th March 2017 using an Illumina MiSeq 2 × 300 kit at 12 pM with a 10% PhiX spike in. COI - Folmer, COI - Elbrecht, 18S V4 - Bradley, and 18S V8 - Bradley were sequenced on the 19th June 2017 using an Illumina MiSeq 2 × 300 kit at 15 pM with a 10% PhiX spike in.

Sequences were processed using a custom bioinformatics pipeline (see Bioinformatics). A species-by-sample OTU table was generated for each gene (not shown).

Table 6.1 Concentration of purified index PCRs.

ID	Sample ID	Marker	Conc. ng/µl
1981	Bembridge Harbour Lagoon	COI - Leray	15.3
1982	8 Acre Pond, Saltern Lagoon, Lyminton-Keyhaven	COI - Leray	11.3
1981	Bembridge Harbour Lagoon	COI - Arribas	5.82
1982	8 Acre Pond, Saltern Lagoon, Lyminton-Keyhaven	COI - Arribas	14.55
1981	Bembridge Harbour Lagoon	COI - Folmer	0.484
1982	8 Acre Pond, Saltern Lagoon, Lyminton-Keyhaven	COI - Folmer	1.1
1981	Bembridge Harbour Lagoon	COI - Elbrecht	2.71
1982	8 Acre Pond, Saltern Lagoon, Lyminton-Keyhaven	COI - Elbrecht	2.42
1981	Bembridge Harbour Lagoon	18S V4 - Bradley	3.27
1982	8 Acre Pond, Saltern Lagoon, Lyminton-Keyhaven	18S V4 - Bradley	3.23
1981	Bembridge Harbour Lagoon	18S V8 - Bradley	3.97
1982	8 Acre Pond, Saltern Lagoon, Lyminton-Keyhaven	18S V8 - Bradley	2.73

Results and discussion

The MiSeq paired-end sequencing of the two samples captured a lot of diversity in terms of OTUs (Table 6.2) but the vast majority of these originated from prokaryotic sources, most of which could not be identified or were so distantly related that the identifications were not very

useful. No lagoon sand shrimp *Gammarus insensibilis* or starlet sea anemone *Nematostella vectensis* were detected with any primer set.

COI and 18S OTUs had their own issues. The identifications for the COI OTUs were, for each marker, dominated by low identity hits to the reference database (Figure 6.1). The vast majority of these matches were bacterial, for example the most common identification was a circumstantial low match (<85%) to *Candidatus pelagibacter*, which is a model bacterial organism with a lot of genomic reference data.

Table 6.2 Number of OTUs captured at the two sites using the six different markers.

ID	COI - Leray	COI - Arribas	COI - Folmer	COI - Elbrecht	18S V4 - Bradley	18S V8 - Bradley
1981 - Bembridge	26	34	35	16	25	22
1982 - 8 Acre Pond	36	25	30	19	25	22

Identification of 18S OTUs suffered from a different issue: all matches were close to sequences already deposited on the reference database (e.g. the average 18S V4 and V8 matches were 99.2% and 98.7%, respectively), but the sequences on the database tend to be unidentified environmental or bacterial samples. For example, 67% and 63% of the 18S V4 and V8 OTUs were “Uncultured” alveolate, cercozoa, chlorophyta, eukaryote, marine picoeukaryote, metazoan, or stramenopile. For 18S the difference between species is too low to resolve species (e.g. Tang et al. 2012), and the databases are overrun with ‘dark taxa’ (i.e. species in GenBank that lacked formal scientific names).

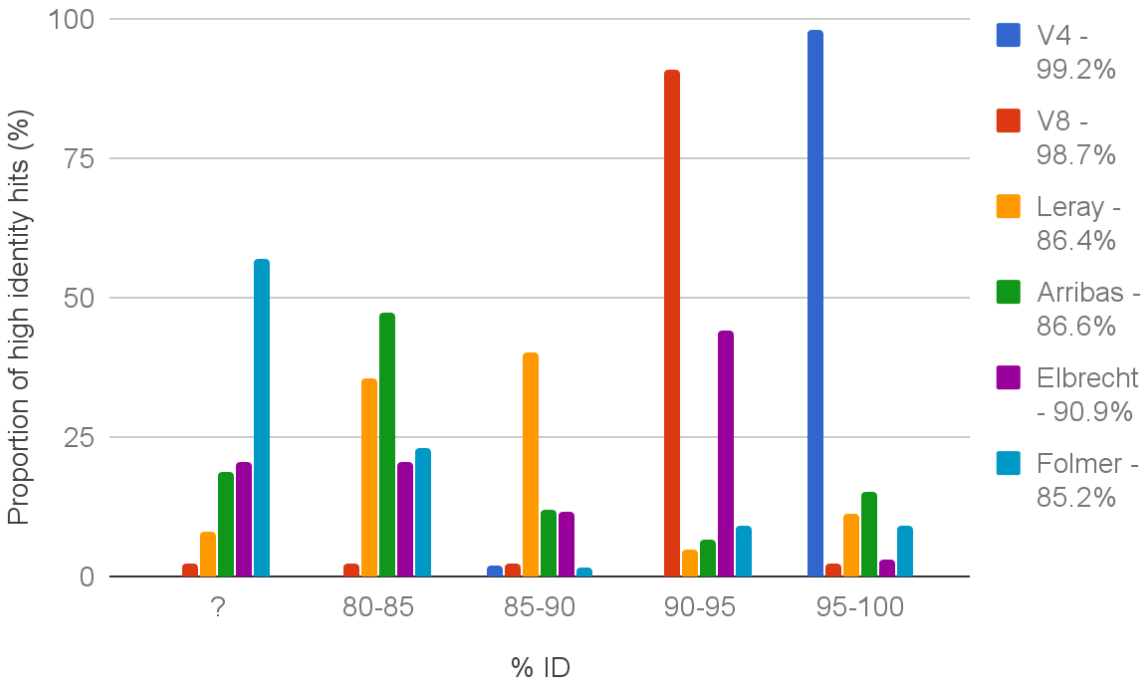


Figure 6.1 The proportion of high identity matches for the six markers grouped among the two sites. The average percentage identity for each gene is shown in the legend.

Technological readiness and further developments

As with the freshwater invertebrate eDNA project, eDNA metabarcoding for lagoon invertebrates is also in its infancy.

Aquatic invertebrate eDNA metabarcoding presents a bigger challenge because of the prevalence of non-target DNA, organisms concentrated by the filtering process, and the limited amount of 'eDNA' shed by invertebrates. There are a few areas for further development in the invertebrate eDNA metabarcoding including better sampling strategies and better primer design. In this case primers specifically targeting the lagoon sand shrimp and starlet sea anemone could be developed.

7. Marine water sample to detect key species Seahorse - Poole harbour

Introduction

There are two species of seahorse in the UK: the short-snouted seahorse *Hippocampus hippocampus* and the spiny seahorse *Hippocampus guttulatus*. Seahorses are elusive and live in shallow, weedy areas, particularly eelgrass beds.

These species are distributed widely around much of the British coastline from the Shetlands westbound to the southern coast with some reported sightings on the east coast and out into the North Sea. Both species have been formally protected in England and Wales under the Wildlife and Countryside Act (1981) (Joint Nature Conservation Committee 1981). Two important seahorse localities include the Poole Harbour SSSI and Studland Bay in Dorset, which support both species.

Seahorse surveys are difficult and inefficient because seahorses live in a low clarity environment among dense eelgrass/seaweed beds, and they are small, cryptic chameleonic animals. All in all there is a high level of specialist expertise and time required to effectively survey seahorses in the conventional way.

Even when the expertise are available, sampling seahorses still requires a lot of effort. Diver surveys are the current standard and can only cover very small areas. For example, over the course of three years (2008-2011), the Seahorse Trust has monitored the Studland Bay population using diver surveys, the effort required for 116 seahorse sightings equated to a total of 346 volunteers, 364 diver days, 612.10 hours of dive time (Garrick-Maidment 2011). While this work is incredibly useful for the continued monitoring and ecological study of the seahorses, it might not be the most efficient and cost-effective way to determine their presence/absence in the first instance.

eDNA sampling could be a useful alternative to seahorse sampling because these elusive creatures still leach plenty of DNA into the environment.

Aim

To detect the presence of seahorse (spiny and short-snouted seahorse) and other ‘difficult to detect’ species, within eelgrass habitat and associated assemblage.

Approach

Here we test an eDNA sampling method coupled with DNA metabarcoding to determine the presence of seahorse in tanks housing them at the Zoological Society of London, and in natural samples in Poole Harbour and Studland Bay.

Methodology

Field sampling

eDNA was sampled using Sterivex filters (see Annex 1, eDNA sampling with filters). Each sample consisted of 3 filter replicates, representing a total of approximately 500 ml of water.

eDNA metabarcoding

DNA was separately extracted from each filter (see DNA extraction: Filters), and DNA extracts from the three filter replicates were then combined for each site. PCR amplification was of a short region of the 12S rRNA gene (~230 bp) using the MiFish primers (Miya et al., 2015) (see Polymerase Chain Reaction). Eight of the nine DNA extracts (six natural samples and three tank samples) were successfully amplified for each of the replicates, but one sample (Poole 4 - Hutchins Buoy) failed to amplify despite dilution of the DNA and the addition of various PCR enhancers. The remaining eight samples resulted in PCR bands that appeared to be high yield and of the expected size.

Successful PCR products were then prepared for sequencing (see Library preparation), with each sample prepared as a separate library (8 libraries). Eight of the nine samples were successfully prepared into libraries, but one sample (Poole 4 - Hutchins Buoy) still failed even after the second round of amplification (Table 7.1).

Table 7.1 Concentration of purified index PCRs.

ID	Sample ID	Conc. (ng/µl)
1954	Poole 1 (SOUTH DEEP)	11.85
1955	Poole 2 (Nr Buoy #36(8))	10.28
1956	Poole 3 (Wareham Channel Buoy)	7.86
1957	Poole 4 (Hutchins Buoy)	0.085
1958	Poole 5 (Poole Bridge)	10.95
1959	Studland Bay	12.3
2024	ZSL *H. hippocampus* = Tank 2	13.1
2025	ZSL *H. hippocampus* = Tank 1	11.8
2026	ZSL *H. guttucatus*	15

All purified index PCRs were pooled into two equimolar libraries, except the sample 1957 (Poole 4 - Hutchins Buoy), which was added to the final *in toto* to compensate for its low concentration. The final libraries were sequenced on the 6th and the 23rd March 2017, respectively, using an Illumina MiSeq 2 × 300 kit at 12 pM with a 10% PhiX spike in.

Sequences were processed using a custom bioinformatics pipeline (see Bioinformatics). A species-by-sample OTU table was generated (Table 7.2).

Results and discussion

The MiSeq paired-end sequencing of the nine samples yielded 1.33 million reads, of which 1.11 million reads (84.3%) passed our internal quality filter. A total of 41,759 unique

sequences were obtained, which were clustered into 20 OTUs (excluding bacterial, archeal, and contaminant OTUs) (Table 7.2). These included 19 fish taxa and 1 bird species (Table 7.2). Importantly, seahorse (*Hippocampus* sp.) was found in two of the three tanks from the ZSL. No seahorse could be found in remaining tank but 105,108 sequences matching Yellow longnose butterflyfish (*Forcipiger flavissimus*) were found, and we hypothesise that the sampling was conducted close to the inflow of water, and that inflow was from a tank housing this species. The two seahorse OTUs most closely matched reference sequences for the lined seahorse *H. erectus* and flat-faced seahorse *H. trimaculatus* since 12S reference sequences for *H. hippocampus* and *H. guttulatus* were not available. Nevertheless, this demonstrated that seahorse DNA can be amplified from water samples using the MiFish primers, and the sequences obtained from the tanks were able to serve as reference sequences henceforth.

No seahorse DNA was found in any of the natural samples. This indicates that either seahorse DNA was absent or the concentration was below the limits of detection in the samples that were taken. The timing and location of the seahorse samples may explain the lack of detections, since it is thought that seahorses migrate to deeper waters during winter or in unfavourable conditions. These samples were taken on the 2nd December 2016 and so seahorses may have migrated to much deeper waters outside of the sampling area or have been in such low abundances to be below the limits of detection by eDNA. It is likely that far greater sampling effort or much more targeted sampling (i.e. of specific eelgrass beds) is necessary to detect such rare species.

16 fish and 1 bird species (Eurasian widgeon *Anas penelope*) were detected in the environmental samples. Most of the species detected are expected in the sampling localities.

The only unresolved findings are:

1. OTU30 - *Chelon labrosus* / *Liza ramada* is either thick-lipped mullet or thin-lipped mullet and maybe incorrectly attributed on the reference database.
2. OTU12 - *Chelidonichthys* sp. matched to two other gurnard species (*Chelidonichthys kumu* and *Chelidonichthys spinosus*). This OTU is likely *Chelidonichthys obscurus*, the only British native gurnard of this genus; this species is not currently represented on the database we use for identification.
3. OTU 10 - *Pomatoschistus* sp. matched to *Pomatoschistus knerii*, but with a low identity score (96.95%). This species is likely a different British goby that is not represented on the database we use for identification.

Technological readiness and further developments

This pipeline is well developed and can effectively identify fish species from eDNA in water samples. The identification of marine fish species is currently limited by the completeness of reference databases, but these continue to be filled. It is possible to revisit and reanalyse these data when the databases are more complete.

One of the key factors that determines success in detection of fish is the amount of water that is filtered, and the Sterivex filters remain somewhat restrictive in this sense as they clog rapidly. We are conducting tests on some alternative filter units that enable greater volumes of water to be processed, and these will feature in any further work. As with other analyses, we intend to move towards the use of indexed primers for the initial PCR to build into the

pipeline an additional layer of resilience to contamination, and this will involve modification of the library preparation protocol.

Table 7.2 OTU table based on 12S.

Common name	Species	ID%	OTU	1954	1955	1956	1958	1959	2026	2025	2024
Seahorse species	*Hippocampus trimaculatus*	96.4	OTU2						195170		
Seahorse species	*Hippocampus erectus*	98.4	OTU1							248600	
Yellow longnose butterfly fish	*Forcipiger flavissimus*	100	OTU3								105108
Allis shad	*Alosa alosa*	100	OTU1				17307				
Stone loach	*Barbatula barbatula*	99.5	OTU3			987					
Gurnard species	*Chelidonichthys* sp.	100	OTU1				11651	20489			
Thick- / thin-lipped mullet	*Chelon labrosus/ Liza ramada*	100	OTU3			1664					
Pacific herring	*Clupea pallasii*	99.5	OTU2		4941						
European bullhead	*Cottus gobio*	99.5	OTU8			36692					
Bass	*Dicentrarchus labrax*	100	OTU1		20106		14642	7014			
Three-spined stickleback	*Gasterosteus aculeatus*	100	OTU2			5464					
Black goby	*Gobius niger*	100	OTU3					1617			
Monkfish	*Lophius piscatorius*	99.5	OTU2				2114				
Minnow	*Phoxinus phoxinus*	98	OTU1	3115		17832		6885			
European flounder	*Platichthys flesus*	99.6	OTU1		6674			14109			
Goby species	*Pomatoschistus knerii*	97	OTU1	25627	20920	8349	7731	6420			
Goby species	*Pomatoschistus* sp.	100	OTU2	9793							
Brown trout	*Salmo trutta*	98	OTU3			14551					
Grayling	*Thymallus thymallus*	100	OTU2			10181					
Eurasian wigeon	*Anas penelope*	100	OTU6	50898				793			

8. Characterising vegetation using DNA in soils Vascular plants - Derbyshire Dales

Introduction

Natural England has a Long Term Monitoring Network (LTMN) on 37 sites in England. On each site a standard set of measurements are made using standard protocols and these include weather, air quality, vegetation, soil attributes, birds and butterflies. Collection of these standardised data allows for comparisons across a wide network, as well as change in communities over time. Understanding these trends and their drivers is necessary for mitigation or adaptive management.

Vegetation surveys are carried out every 4 years in 50 permanently marked plots on each sites. Each plot is 2 m x 2 m and is divided into 25 individual cells. The plant species present in each cell are recorded along with additional information at the cell (vegetation height) or plot (slope angle) scale. These surveys are carried out by either paid contractors or by teams of Natural England staff working with external volunteers. The second approach potentially introduces more inconsistency between surveys and overall error. Moreover, the surveys can only be carried out in certain seasons when the plants can be reliably identified.

Molecular techniques may offer a means of quality assuring traditional field survey or even for monitoring vegetation at different times of the year. Soil cores offer a way of easily sampling plots for their roots. Soils have been used to assess plant diversity before (Yoccoz et al. 2012; Fahner et al. 2016).

Aim

To test if plant species can be detected through DNA analysis of the top 15 cm of soil and compare findings to traditional field survey.

Approach

Here we extract DNA from roots isolated from soil cores and DNA metabarcode them using *rbcL* and ITS2 genes to characterise the plant species present in the sample.

Methodology

Field sampling

Soil cores were taken from plots on Derbyshire Dales National Nature Reserve. These were replicate plots set up using the LTMN methodology (not the actual long-term monitoring plots). Each plot was surveyed by expert field botanists from Natural England's Field Unit using the standard LTMN methodology. In addition the cover of each species was estimated in each cell. Soil cores were then taken from the centre of each cell to sample plant roots.

DNA metabarcoding

Roots were isolated from the soil using a series of sieve and wash steps (see Annex 1, Roots from soil) and for a subset of 20 samples, DNA was extracted using a PowerSoil kit

(see DNA extraction: Homogenate). The PowerSoil kit was chosen ahead of FastDNA, DNeasy Plant Mini kit, and GeneAll because of higher DNA yields and success of subsequent PCR steps. To reduce the effect of plant and soil inhibitors, the DNA was purified post extraction (see DNA purification).

PCR amplification of a *rbcL* cpDNA (~450 bp) and ITS2 (~475 bp) using primers and conditions described by de Vere et al. (2017) and Sickel et al. (2015), respectively, were performed to target plants (see Polymerase Chain Reaction). PCR success for *rbcL* was variable for the 20 samples, but after varying levels of DNA dilution, all three replicate PCRs were successful. All of the ITS2 PCRs were successful. PCR bands appeared to be high yield and of the expected size.

Successful PCR products were then prepared for sequencing (see Library preparation) with each sample prepared as a separate sequencing library (19 libraries). All but one of the samples were successfully indexed (Table 8.1), PCR bands appeared to be high yield and of the expected size. The *rbcL* library for 1801 was not successful, while the final concentrations of two *rbcL* samples were lower than the other 17 libraries (1805 and 1810), but this was compensated for by increasing the final volume of these samples in the final pool. The final pool was sequenced on the 23rd March 2017 using an Illumina MiSeq 2 × 300 kit at 10 pM with a 5% PhiX spike in.

Sequences were processed using a custom bioinformatics pipeline (see Bioinformatics). A species-by-sample OTU table was generated for *rbcL* and ITS2 (combined in Table 8.2).

Results and discussion

MiSeq paired-end sequencing of the 19 libraries, together with another 75 libraries (total number of libraries = 94), yielded 44 million reads, of which 41.4 million reads (94%) passed Illumina's internal filter.

rbcL

A total of 19 samples were sequenced targeting the chloroplast gene *rbcL*. After removing contaminant OTUs and collapsing oversplit OTUs (i.e. those where subtly different sequences were split among two or more different OTUs), a total of 31 unique OTUs were identified (Table 8.2). These OTUs belonged to a total of 18 families, the majority of which were grasses (Poaceae; Figure 8.1).

The main challenge with using *rbcL* is that the gene is not evolving fast enough to universally discriminate between species – i.e. the resolution of the marker is not at the species level for all taxa. This was the case for 14 of the 31 OTUs. For instance, it is not possible to differentiate between *Agrostis canina*, *Agrostis curtisii*, or *Agrostis scabra*. In some cases it was not possible to differentiate between species belonging to different genera (e.g. *Ditrichum heteromallum* and *Pleuridium acuminatum*). These issues may result from the resolution of the marker or from misidentifications in the reference database.

Table 8.1 Concentration of purified index PCRs. Amplification of 1801 for *rbcL* failed.

ID	Sample ID	*rbcL* (ng/µl)	ITS2 (ng/µl)
1799	Derbyshire Dales DNA 1 Cell 1	8.51	19.1
1800	Derbyshire Dales DNA 1 Cell 2	7.42	19.9
1801	Derbyshire Dales DNA 1 Cell 3	N/A*	19
1802	Derbyshire Dales DNA 1 Cell 4	6.96	19.8
1803	Derbyshire Dales DNA 1 Cell 5	6.47	17.7
1804	Derbyshire Dales DNA 1 Cell 6	13.9	19.5
1805	Derbyshire Dales DNA 1 Cell 7	0.865	17.5
1806	Derbyshire Dales DNA 1 Cell 8	11.9	17.2
1807	Derbyshire Dales DNA 1 Cell 9	7.1	18.9
1808	Derbyshire Dales DNA 1 Cell 10	9.86	21.8
1809	Derbyshire Dales DNA 1 Cell 11	3.81	16.5
1810	Derbyshire Dales DNA 1 Cell 12	0.369	2.28
1811	Derbyshire Dales DNA 1 Cell 13	8.88	16.7
1812	Derbyshire Dales DNA 1 Cell 14	11.1	17.7
1813	Derbyshire Dales DNA 1 Cell 15	9.42	13.6
1814	Derbyshire Dales DNA 1 Cell 16	4.44	11.8
1815	Derbyshire Dales DNA 1 Cell 17	12.5	17.2
1816	Derbyshire Dales DNA 1 Cell 18	16.4	18.7
1817	Derbyshire Dales DNA 1 Cell 19	12	17.6
1818	Derbyshire Dales DNA 1 Cell 20	16.3	17.6

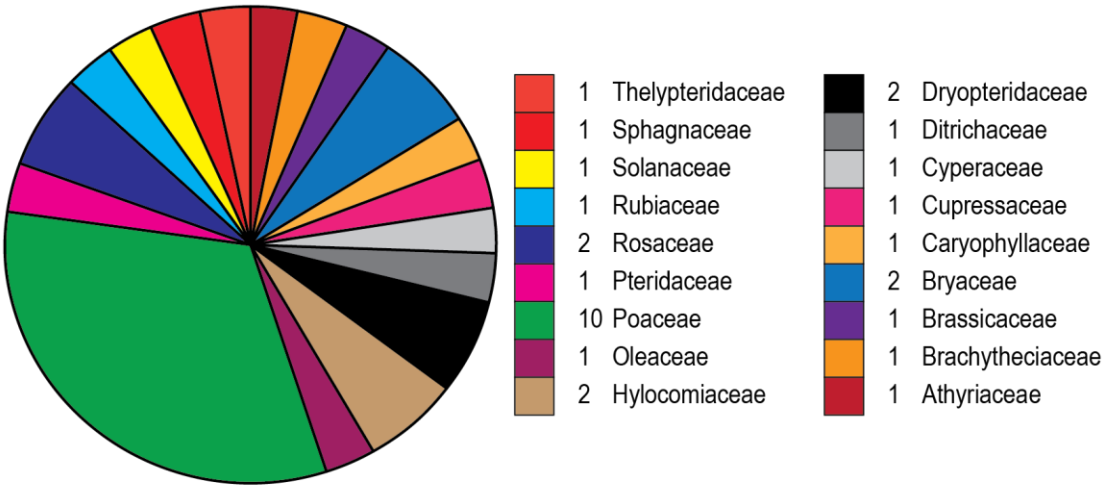


Figure 8.1 Distribution of 31 *rbcL* OTUs among the 18 families.

ITS2

All 20 samples were successfully sequenced for ITS2. While ITS2 can be used to recover plant OTUs, it also recovers fungi, which are highly prevalent in soil. Twenty OTUs remained after removing non-plant OTUs, which were split across 9 families (Figure 8.2). Consistent with the *rbcL* data, the majority of OTUs were grasses (Poaceae; Figure 8.2).

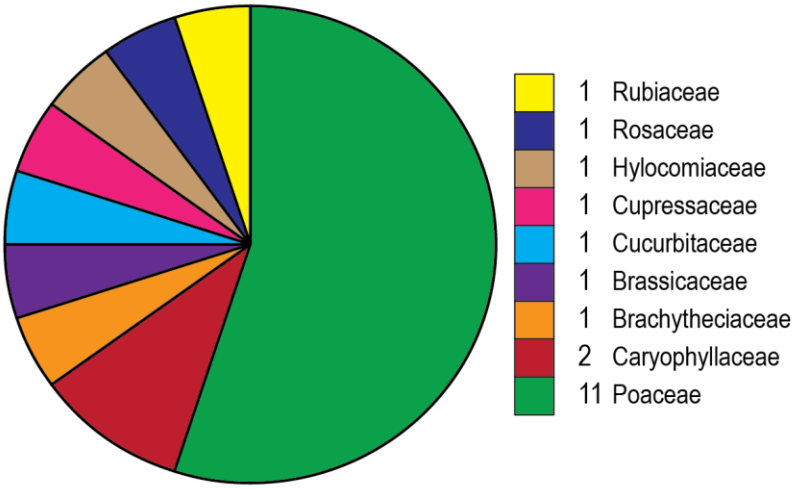


Figure 8.2 Distribution of 20 ITS2 OTUs among the 9 families.

rbcL and ITS2

Table 8.2 combines the two datasets for these plots. The variability of ITS2 is higher than *rbcL* and so there tends to be more species level resolution for that gene, but then the coverage of *rbcL* in the databases is much higher, resulting in more identifications (even if at a lower resolution).

There is some agreement between the two datasets - for example, common mouse-ear *Cerastium fontanum* is only found in one plot (Derbyshire Dales DNA 1 Cell 12) by both markers. But the same DNA extract has often yielded different diversity depending on the gene used. For example 19 OTUs were in Derbyshire Dales DNA 1 Cell 7, of which only 3 were common to both *rbcL* and ITS2 (Figure 8.3). This is likely due to a combination of unreliable reference databases (i.e. misidentifications or missing species) and poor species resolution in *rbcL* (e.g. it is possible that one of either *rbcL* Poaceae sp. 1 or Poaceae sp. 2 are from the same plant as ITS2 *Festuca* sp.).

An initial comparison of the species identified through DNA analysis with those identified by field surveyors indicates that the main species are similar although the DNA results often give several closely related species as possibilities (especially with grasses). Some of the bryophyte and vascular plant species identified in the DNA analysis results seem unlikely to have actually been present (e.g. aquatic species or non-native species occurring as rare casuals in Britain). No ferns were recorded by the field surveyors and grown plant would have been obvious at the time of survey. Some species were recorded in the field survey but not the DNA analysis but the field survey covered the whole 40 cm x 40 cm cell while the soil sample covered a much smaller area in the centre of the cell. A fuller comparison is planned for these data.

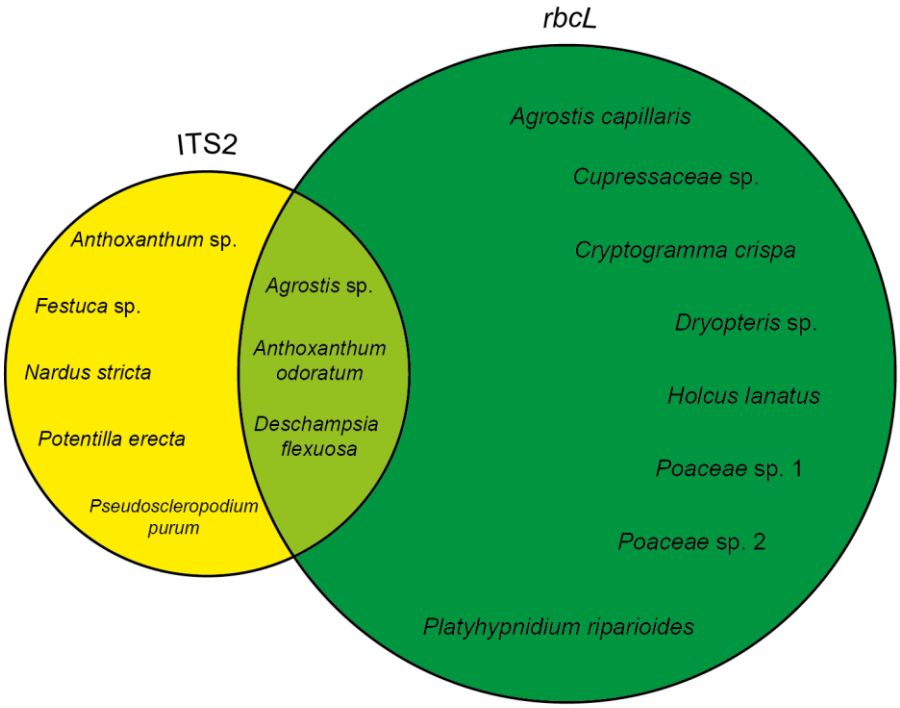


Figure 8.3 Unique and shared OTUs identified from *rbcL* and ITS2 datasets for Derbyshire Dales DNA 1 Cell 7.

Table 8.2 OTU table based on rbcL and ITS2. Multiple identities are returned when multiple equally high matches are present in the reference database. For some of the identifications there are multiple OTUs that have been collapsed into one (# OTUs), which may be associated with a range of identity similarity (ID%). Note that some species may be split into multiple OTUs if it is lumped in with other species, which is either due to the low resolution marker or mistakes in the database. Continued over four pages.

Family	Species	ID%	Marker	#OTU	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Athyriaceae	*Athyrium filix-femina* *Athyrium distentifolium*	100	*rbcL*	1					13117									1094						
Brachytheciaceae	*Platyhypnidium riparioides*	99.3	*rbcL*	1		3679				657	1051						16871		27958	60				
Brachytheciaceae	*Pseudoscleropodium purum*	100	ITS2	1							601													
Brassicaceae	*Armoracia rusticana*	98.4	ITS2	1		1928																		
Brassicaceae	*Eruca vesicaria* *Diplotaxis harra* *Brassica tournefortii*	100	*rbcL*	1															956					
Bryaceae	*Bryum cyclophyllum*	100	*rbcL*	1	8031																			
Bryaceae	*Bryum funkii*	100	*rbcL*	1																			1006	
Caryophyllaceae	*Cerastium fontanum*	100	ITS2	2												22302								
Caryophyllaceae	*Cerastium fontanum*	100	*rbcL*	1												11659								
Caryophyllaceae	*Silene dioica*	100	ITS2	1													225							
Cucurbitaceae	*Cucurbita pepo*	97.2	ITS2	1					553															
Cupressaceae	*Chamaecyparis lawsoniana*	100	ITS2	1																			117	
Cupressaceae	*Chamaecyparis lawsoniana* *Juniperus communis*	100	*rbcL*	1							1624													
Cyperaceae	*Carex pilulifera*	100	*rbcL*	1		10247		14090					1301	1708			24699	19757			3836	5808		
Ditrichaceae	*Ditrichum heteromallum* *Pleuridium acuminatum*	99.6	*rbcL*	1					514														422	
Dryopteridaceae	*Dryopteris carthusiana* *Dryopteris dilatata*	100	*rbcL*	1							5929													
Dryopteridaceae	*Dryopteris filix-mas*	100	*rbcL*	1																3549			260	
Hylocomiaceae	*Hylocomiastrum himalayanum*	98.5	*rbcL*	1															1981					
Hylocomiaceae	*Rhytidiadelphus squarrosus*	100	ITS2	1															707				260	
Hylocomiaceae	*Rhytidiadelphus squarrosus*	100	*rbcL*	1																			64	
Oleaceae	*Fraxinus excelsior*	100	*rbcL*	1	3179																			
Poaceae	*Agrostis canina*	100	*rbcL*	1	38340			1341	1880	772	27174		2292	1257	25736	1221		681			4152			

Family	Species	ID%	Marker	#OTU	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	Agrostis curtisii *Agrostis scabra*																							
Poaceae	*Agrostis capillaris*	100	*rbcL*	1	993	77570		9949	16204		8820		695	1622		2799							1420	
Poaceae	*Agrostis capillaris* *Agrostis castellana*	99.5-100	ITS2	2					385														460	
Poaceae	*Agrostis capillaris* *Agrostis gigantea* *Agrostis scabra*	99.1-100	ITS2	2		671	194		1386		3793						307						794	
Poaceae	*Agrostis stolonifera* *Agrostis canina* *Agrostis vinealis* *Alopecurus geniculatus*	100	ITS2	1					337														214	
Poaceae	*Agrostis vinealis*	99.1-99.5	ITS2	2					1768														416	
Poaceae	*Anthoxanthum odoratum*	100	*rbcL*	1					1078		622		5743	1665										
Poaceae	*Anthoxanthum odoratum*	99.1-100	ITS2	2							340		321											
Poaceae	*Anthoxanthum odoratum* *Anthoxanthum aristatum*	96.2-100	ITS2	3	626		1629	565	375		6474		451	6224		764	227	335	776	2458			545	
Poaceae	*Anthoxanthum odoratum* *Anthoxanthum ovatum*	98.6	ITS2	1										1249			407			198			200	
Poaceae	*Calamagrostis stricta*	99.6	*rbcL*	1									613											
Poaceae	*Deschampsia danthonioides*	98.5-100	*rbcL*	2															702			2041		
Poaceae	*Deschampsia flexuosa*	97.4-100	*rbcL*	4	49117	7268		7376		27922	43577				48214	900	4117	8561	32246				3022	
Poaceae	*Deschampsia flexuosa*	98.8-100	ITS2	4	3023	2684		9444		9334	18420				9441	929	2703		13615	726			290	
Poaceae	*Festuca ovina*	99.5	ITS2	2																1030			149	
Poaceae	*Festuca ovina* *Festuca lemanii* *Festuca filiformis* *Festuca vivipara*	100	ITS2	1				279	617		346									972				
Poaceae	*Festuca rubra* *Avenula pratensis* *Festuca ovina* *Deschampsia flexuosa*	98.9	*rbcL*	1							10139													
Poaceae	*Festuca rubra*	96.3-	*rbcL*	11	2930			3780		3559	8454				6577			551	1483					

Family	Species	ID%	Marker	#OTU	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	Avenula pratensis *Festuca ovina* *Deschampsia flexuosa* *Festuca armoricana* *Festuca longifolia* *Vulpia unilateralis*	98.9																						
Poaceae	*Holcus lanatus*	94.4	*rbcL*	1		1168				1181	3320													
Poaceae	*Nardus stricta*	95.1-100	ITS2	3	16935	15381	4417	12790	29607	32980	1060					17224	17056		16569	1964		8724		
Poaceae	*Nardus stricta*	100	*rbcL*	1		1657			583	1373						236								
Poaceae	*Poa pratensis* *Poa ligularis* *Poa chaixii*	100	*rbcL*	1																1106		3912		
Pteridaceae	*Cryptogramma crispa*	100	*rbcL*	1							11323													
Rosaceae	*Potentilla erecta*	100	ITS2	1	73914	99851	16063	20170	27169	68833	21776	91446	58217	86612	79893	49303	17359	49763	65589	22799	91689	93331	13864	123357
Rosaceae	*Potentilla reptans* *Potentilla erecta*	100	*rbcL*	1		5965				61445		127420	79773	120118			2126	76675	61033	70	98471	91972		131208
Rosaceae	*Potentilla reptans* *Potentilla sterilis* *Potentilla erecta* *Duchesnea indica* *Potentilla canadensis* *Potentilla anglica*	98.1	*rbcL*	1																		757		
Rubiaceae	*Galium saxatile*	97.5-99.4	ITS2	8	4527		137	235				3063							3237	3431	564	685	4007	
Rubiaceae	*Galium saxatile* *Galium sterneri*	98.1-100	*rbcL*	2																		5453		
Solanaceae	*Solanum dulcamara* *Capsicum annuum*	100	*rbcL*	1		4420																		
Sphagnaceae	*Sphagnum palustre* *Sphagnum fallax*	99.3-99.6	*rbcL*	2									462							1834				
Thelypteridaceae	*Oreopteris limbosperma*	100	*rbcL*	1													2052							

Technological readiness and further developments

It is possible to obtain plant identifications from soil cores using both *rbcL* and ITS2, but both have significant technical difficulties in the laboratory and analytical workflows.

1. Root extraction from soil is a very labour intensive process and is not very clean. Finding a way to do this in the lab was a significant challenge and undoubtedly resulted in tissue loss and lower DNA retention owing to the various wash steps. Perhaps an alternative way of separating the roots from the soil could be explored; while contraptions and protocols do exist for this purpose, they are too large scale, with no consideration of sterility and cross contamination. A possible method might be ultrasonic baths in disposable containers and floatation in an appropriate buffer, or freeze-drying.
2. *rbcL* is most prevalent in photosynthetic tissue, and so the copy number of the gene in the roots is lower than in the leaves. Despite this, meaningful data could still be produced. The long length of the marker chosen posed problems with the bioinformatics, and so a future development would be picking a smaller part of the *rbcL* gene so that the ends could easily merge at the bioinformatics stage, which would also improve the overall quality of the reads.
3. The high number of fungal OTUs found using ITS2 is probably due to a combination of difficulty in removing soil from the roots (discussed above), and a lack of specificity of the ITS2 primers used. A more plant-specific set of primers could probably be developed to reduce the amount of fungal sequences recovered.
4. The length variability of ITS2 brings bioinformatics challenges, requiring specific programs and databases to be used.
5. Common to both markers is the need for a highly curated taxonomic database. There are likely errors that could hamper any future development of these techniques. For example, identical ITS2 sequences are identified in the database variously as sheep's-fescue *Festuca ovina*, fine-leaved sheep's-fescue *Festuca filiformis*, viviparous sheep's-fescue *Festuca vivipara*, and confused fescue *Festuca lemanii*. Problems with accurately and consistently differentiating between these species morphologically will be inherited by molecular approaches unless curation of the reference databases can be undertaken.

While *rbcL* and ITS2 have been recommended for these types of soil DNA analyses in the past (Fahner et al. 2016), the true diversity of the samples has never really been analysed. While it is possible to obtain a high number of OTUs from a single soil core, it is not easy (but certainly possible) to determine whether the identity of those OTUs corresponds well to the traditional taxonomy. As it stands, both markers provide an insight into the diversity present in soil roots, but some developments in the sampling, cleaning, and laboratory processes are required before it can be routinely used.

9. Species in marine sediments Marine Invertebrates

Introduction

Intertidal and subtidal marine protected areas are protected under various legislative drivers, and thus require monitoring. Natural England undertake many sediment monitoring projects to improve understanding of marine protected site condition and to identify impacts through research studies.

Currently, analysis of sediment grabs and cores from the marine environment is time-consuming, requires expertise in identification skills and is therefore very expensive. Using DNA analysis could provide a technique to reduce the costs of both operational and investigative monitoring, whilst still providing the detailed level of information required.

Aim

Develop a method for analysing full marine invertebrate communities within intertidal and subtidal core sediment samples.

Approach

Here we test the utility of DNA metabarcoding of a marine sediment sample. Using two DNA markers to increase the breadth of diversity detected.

Methodology

Field sampling

Five sediment core samples were collected using a 0.05m^2 hand held core in December 2016 from the Wash SSSI in Norfolk. The samples were sieved to 0.5 mm in seawater on the shore approximately 2 hours after collection. The sieve was soaked in 10% bleach for a few minutes between each sample and thoroughly rinsed in seawater prior to the next sample to remove all traces of bleach. Sieved samples were placed in 250 ml sterile sample pots and covered in Longmires buffer solution. Sample size ranged from approximately 25 to 100 ml, and were topped up to 225 ml with the buffer. Samples were then sent to Bangor University where they were stored at 4 °C until DNA extraction (approximately 4 weeks).

DNA metabarcoding

The sample processing, DNA extraction, and COI PCR amplification for this project was led by Dr Iliana Bista and Professor Simon Creer at Bangor University, and sequencing libraries were provided to NatureMetrics for sequencing and analysis (see Marine invertebrate metabarcoding for more details).

The sample was separated into five subsamples, which were separately homogenised. In order to facilitate the DNA extraction, large empty shells (e.g. from mussels) were removed manually prior to homogenisation. DNA was extracted from the homogenate using a MoBio

PowerMax Soil kit following a protocol modified from the manufacturer. These DNA extracts were purified to remove PCR inhibitors (see Annex 1, DNA purification).

COI metabarcoding

PCR amplification of a fragment of the cytochrome oxidase *c* subunit 1 (COI) was undertaken using the Leray et al. (2013) primers, which target Metazoa. The primers were modified to include Illumina adaptors necessary for downstream high throughput sequencing. The PCRs prepared into sequencing libraries with a second PCR, which used the Illumina tails from the first PCR as a priming site. The second PCR added Illumina indexes and sequencing adaptors.

All PCRs were successful, with all amplicons being of the expected size. The yield of 2 samples (Samples 2 and 4) were lower than the rest (Table 9.1), but these were repeated and finally pooled in a higher volume to compensate. The final libraries were sequenced on the 23rd March 2017.

Table 9.1 Concentration of purified index PCRs.

Marker	ID	Sample ID	Conc. (ng/µl)
COI	1	Sample 1	4.68
COI	2	Sample 2	0.399
COI	3	Sample 3	5.2
COI	4	Sample 4	0.304
COI	5	Sample 5	7.16
18S	1-5	Combined (1-5)	4.44

Sequences were processed using a custom bioinformatics pipeline (see Bioinformatics). A species-by-sample OTU table was generated for each gene (Table 9.2).

18S metabarcoding

PCR amplification of a fragment of the V4 hypervariable region of 18S rDNA was performed using primers designed by Bradley et al. (2016). The primers were modified to include Illumina adaptors necessary for downstream high throughput sequencing. These PCRs were performed in triplicate. All of the replicate PCRs were successful, PCR bands appeared to be of high yield and of the expected size.

Triplicate PCRs for each of the five subsamples (total = 15) were pooled into a single sample and prepared into a sequencing library (see Library preparation). This library preparation was successful, the PCR band appeared to be high yield (Table 9.1) and of the expected size. The library was sequenced on the 29th September 2017. Sequences were processed using a custom bioinformatics pipeline (see Bioinformatics) and a species-by-sample OTU table was generated for each gene (Table 9.3).

Results and discussion

COI metabarcoding

The MiSeq paired-end sequencing of the five COI libraries yielded 243,045 reads, of which 202,887 reads (83.5%) passed our internal quality filter. A total of 126,963 unique sequences were obtained, which were clustered into 50 Metazoan OTUs (excluding bacterial, archeal, and contaminant OTUs). Only 8 OTUs (16%) remained after removing those OTUs with no match to a reference database higher than 90% and after lumping OTUs with the same identity (Table 9.2). Matches below 90% are typically of microbial origin.

Metazoan diversity detected in the sample consisted of three mollusc species, three copepod species, and two annelid species. Only one metazoan OTUs (the mollusc *Hydrobia ulvae*) was detected in all five subsamples. This heterogeneity among the subsamples indicates that either the samples were not processed equally at some point in the process (homogenisation, DNA extraction, amplification, sequencing, etc.) or that multiple subsamples are required to sequence the entire breadth of the sample. It is likely a combination of these factors.

Table 9.2 OTU table based on COI. For some of the identifications multiple OTUs that have been collapsed into one (# OTUs), which may be associated with a range of identity similarity (ID%).

Taxa	Closest species match	ID%	# OTUs	1	2	3	4	5
Copepod	*Acartia bifilosa*	99.7	1				613	435
Copepod	*Centropages hamatus*	100	1	354		239	127	
Common cockle	*Cerastoderma edule*	100	1		497			
Laver Spire Shell	*Hydrobia ulvae*	98-100	7	27198	21261	31307	9038	30458
Baltic clam	*Macoma balthica*	98-100	2	2781	229			3367
Annelid	*Nephtys hombergii*	100	1		899			
Annelid	*Scoloplos armiger*	95.5	1	713				
Copepod	*Temora longicornis*	100	1				115	

18S metabarcoding

The MiSeq paired-end sequencing of the sequencing library clustered into 82 OTUs including 30 metazoans, 10 plants, 22 protists, and 13 Algae. 88.9% of the sequence reads are Metazoan (Table 9.3). The four most prevalent OTUs account for ~80% of the sequence reads. These were three molluscs (Laver spire shell *Hydrobia ulvae* - 38.5%; Baltic clam *Macoma balthica* - 20.3%; Common cockle *Cerastoderma edule* - 13.6%) and a copepod (*Centropages* sp. - 6.7%). These four OTUs, along with 2 others are represented in both the COI and 18S metazoan datasets. Only the copepod *Temora longicornis* was unique to the COI dataset.

Notably, multiple species of platyhelminths, nematodes and ostracods occur in the 18S dataset but these groups are entirely absent from the COI dataset. This is consistent with other studies (e.g. Bhadury et al. 2006) and our own in-house data from similar samples, which also found that these key groups in the marine benthic environment are much better recovered by 18S primers than by commonly-used COI primers.

Discrepancies between the two datasets occur because:

1. Species are not equally represented by COI and 18S sequences in the reference databases. For example the OTU identified by 18S as *Ceramium rubrum* at 99.7% is likely *Ceramium secundatum*, which was identified by COI with 100% match. This species has no 18S sequence available online.
2. COI has better species-level resolution among closely-related taxa. For example, the copepod species *Centropages hamatus* is a 100% match the reference database with both COI and 18S, but for 18S the sequence is also identical to two other species (*Centropages abdominalis*, *Centropages typicus*).
3. COI and 18S primers detect different subsets of diversity, with 18S detecting groups that are not detected by COI.

The identifications with low percentage sequence matches should be treated with very low confidence and usually represent cases where the true species is absent from the reference database. Ideally, a more sophisticated taxonomic assignment approach should be implemented so that taxa can be reliably assigned to higher-level groups in these cases.

Table 9.3 Metazoan OTU table based on the hypervariable V4 region of 18S.

Taxa	Species	ID%	# Sequence
Annelid species	*Nephtys* sp.	100	301
Annelid: Paddleworm	*Eteone longa*	100	38
Bryozoan species	*Amathia pustulosa*	100	260
Bryozoan species	*Membranipora* sp.	97.1	110
Bryozoan species	*Anguinella palmata*	88	93
Copepod species	*Centropages* sp.	100	4318
Copepod species	*Acartia bifilosa*	100	389
Copepod species	Thalestridae sp.	97.2	305
Hydroid: Sea beard	*Nemertesia antennina*	100	42
Mollusc: Baltic clam	*Macoma balthica*	100	13102
Mollusc: Common cockle	*Cerastoderma edule*	100	8763
Mollusc: Laver Spire Shell	*Hydrobia ulvae*	99.8	24831
Nematode species	*Pellioditis mediterranea*	99.7	23
Nematode species	*Pellioditis marina*	98.9	21
Nematode species	*Viscosia* sp.	100	212
Nematode species	*Viscosia* sp.	100	105
Ostracod species	*Leptocythere lacertosa*	100	1974
Ostracod species	*Limnocythere inopinata*	95.8	342
Ostracod species	*Semicytherura striata*	99.5	274
Ostracod species	*Ilyocypris angulata*	100	41
Platyhelminth species	*Scanorhynchus forcipatus*	96.9	704

Taxa	Species	ID%	# Sequence
Platyhelminth species	*Promesostoma meixneri*	100	411
Platyhelminth species	Schizorhynchidae sp.	92.7	108
Platyhelminth species	*Cilionema hawaiiensis*	97.4	104
Platyhelminth species	*Archilopsis arenaria*	98.2	101
Platyhelminth species	*Macrostomum pusillum*	100	98
Platyhelminth species	*Schizorhynchoides caniculatus*	98.2	54
Platyhelminth species	*Phonorhynchus helgolandicus*	96.1	34
Platyhelminth species	*Cheliplana* cf.	100	33
Platyhelminth species	*Proxenetes* sp.	99.7	21
Uncultured eukaryote	Uncultured metazoan	87.6	51
Uncultured eukaryote	Uncultured metazoan	97.4	48
Uncultured eukaryote	Uncultured metazoan	99.7	24

Technological readiness and further developments

Intertidal and subtidal sediments are highly biodiverse, and obtaining a representative picture of this diversity may require that multiple markers are used. Here we used a combination of short regions of COI and 18S. Both markers suffer from incomplete representation of marine benthic organisms in the reference databases, but it is nevertheless possible to obtain sequences representing a broad spectrum of the metazoan diversity. Here we focus on macrofauna, but a benefit of these analyses is that analysis of meiofaunal and microbial diversity could be carried out in parallel.

More work needs to be carried out to determine how complete these datasets are - for instance by direct comparison of morphological and molecular data. The identification of OTUs is currently limited by the completeness of reference databases, but these continue to be improved and it is possible to revisit and reanalyse these data when the databases are more complete.

We recommend focusing on the 18S gene for future work due to its ability to detect a wider range of metazoan groups.

10. Discussion

Molecular biomonitoring has been promoted as a means to revolutionise the way environmental managers detect and identify species. Here we explored eight monitoring scenarios where there are real challenges in obtaining data.

The pilots cover a wide range of organisms, which can be broadly categorised into fish (freshwater and marine), invertebrates (terrestrial, freshwater, marine) and plants, and each has yielded their own promises and challenges, which we discuss in turn.

Fish

eDNA has proved to be a powerful and non-invasive alternative or complementary method to traditional electrofishing, gillnetting or trawling, which are invasive and have potential size biases. Moreover, with conventional techniques, a high level of sampling effort is required to catch the rarer or more elusive species, or those with low catchability - resulting in a general paucity of data. In comparison, eDNA metabarcoding can yield large amounts of data relatively quickly and cost-effectively. For example, Civade et al. (2016) found that a single eDNA sampling campaign in an Alpine river captured as much data as 20 cumulative years of conventional monitoring. Here we have piloted a full workflow from sampling to data analysis that provides fish community data from freshwater and marine environments. This workflow has built upon methods described by Spens et al. (2017) and Miya et al. (2015).

The aim of these projects was specifically to characterise the fish community composition in Nene Washes, Poole Harbour and Studland Bay, with a particular focus on detecting spined loach (*Cobitis taenia*) and seahorses (*Hippocampus* sp.), which are difficult to detect with traditional methods. Indeed, the conservation status of both British seahorse species is currently given as Data Deficient (IUCN 2017). While previous studies have successfully used eDNA to detect closely related species (weather loach *Misgurnus fossilis* - Thomsen et al. 2012; Bay pipefish *Syngnathus leptorhynchus* - Andruszkiewicz et al. 2017), to our knowledge there has been no published attempt to detect spined loach or seahorses using eDNA monitoring.

Here we have used eDNA metabarcoding to detect a total of 32 different fish taxa across 17 samples from 8 freshwater filter samples, 3 aquarium tanks, and 6 marine filter samples. From the freshwater samples, spined loach was successfully detected in addition to 12 other fish species, all of which are expected in the sampling area, with the exception of the Atlantic Salmon, which is likely an environmental contaminant from wastewater or oils in fishing bait. From the mesocosms, two seahorse species and a longnose butterflyfish (*Forcipiger flavissimus*) were found, which equated to two higher taxonomic level identifications resulting from an incomplete reference database and a false negative, which we hypothesise was due to the sampling being potentially conducted close to the inflow of water from a tank housing Yellow longnose butterflyfish. In the marine samples, seahorses were not among the 16 fish taxa detected, which we hypothesise is due to insufficient sampling and the possible absence of seahorses from the sampled habitat at the time of sampling. We expect that the late sampling of shallow seagrass beds maybe have been inappropriate for seahorse DNA

capture because they tend to migrate to deeper waters during unfavourable conditions.

These fish eDNA metabarcoding surveys were among the first performed with our pipeline and act as important first proofs of concept. We present real data that shows that fish communities can be characterised with eDNA but acknowledge that there are limitations associated with that iteration of the pipeline and have since addressed these to improve the methodology. In the next section we will discuss some of these limitations and proffer ways that we have dealt with them.

Limitations

eDNA surveys are semi-quantitative

The method is only semi-quantitative and abundance data is currently not as accurate as physically catching and counting. While both traditional and eDNA methods have their biases (e.g. different catchabilities - Arreguín-Sánchez 1996), those affecting eDNA methods are less well understood than those affecting physical catch data. Combined, these biases affect the final relationship between the measure of target DNA quantity (whether sequence reads in metabarcoding or quantitation cycle in qPCR) and the measure of abundance (typically biomass or number of individuals), and this is likely to vary substantially among taxa and ecological context (Tillotson and Quinn 2018).

Many factors are thought to affect eDNA abundance measures, including water temperature, water pH, UV exposure, time since DNA release, salinity, inhibition, environmental contamination, spawning, primer bias, proximity to target, sampling method, gene choice, PCR replication, bioinformatic processing, etc. Despite these biases, careful sampling, primer and experimental design can result in data that retains some quantitative information. Indeed a recent trial we conducted at a fish farm in Calverton detected 8 fish species with a tight relationship between relative proportion of fish stock biomass (determined by physical catch of the entire fish population on the same day as eDNA samples were taken) with the relative proportion of sequence reads obtained for each of those detected species. Of course much more work is required to better understand this relationship, and much more work is required to determine how universal this relationship is especially in more open environments (i.e. riverine and marine samples) where other conflicting issues are likely at play (e.g. significantly higher dilution of DNA). This is an active and promising area of eDNA research (e.g. Lacoursière-Roussel et al. 2016; Inui et al. 2017; Chambert et al. 2018; Tillotson and Quinn, 2018) where the use of multivariate models including more abiotic (e.g. temperature, pH, etc.) and biotic (e.g. life history and phenology, etc.) variables are accounting for an increasing amount of the observed variation (Tillotson and Quinn 2018), with the promise that the more we know about the environment the better we will be able to predict species' abundance from eDNA surveys.

An alternative way of interpreting sequence data with relative abundance in mind is to use site occupancy. More abundant species will tend to be detected more frequently across multiple subsamples within the same sample than the less abundant ones and this frequency score can be modelled to give a relative abundance measure. This is the approach taken by the Environment Agency and SEPA (Environment Agency 2017).

eDNA is typically low quantity and quality

Analysing eDNA, compared to DNA derived from tissue samples, is a greater challenge because it is inherently low quantity and low quality. Because of the low quantity of high quality DNA, there is an increased impact of stochasticity; put simply when there isn't much amplifiable DNA it becomes increasingly possible that the target DNA isn't pipetted into the final PCR reaction (Taberlet 1999). As such it is necessary to perform multiple PCR replicates in order to combat the increased stochasticity associated with low concentration DNA. Ficetola et al. (2015) recommend that when detection probabilities are low, at least 8 replicates are required as is recommended with ancient DNA. This issue appears to be more significant in marine samples where the detection probabilities of the target animals are likely lower due to the increased dilution, salinity and tidal effect of marine environments.

eDNA is often inhibited

Extraction of DNA from water often co-extracts compounds that inhibit downstream molecular processes. eDNA samples are often plagued by inhibition especially if taken from water bodies that are stagnant, turbid, full of sediment and overgrown (Williams et al. 2017). For example, environmental water samples often contain degraded plant material and their associated breakdown compounds (e.g. tannins), which can completely inhibit amplification (e.g. Jane et al. 2015). Detecting inhibition is done using internal positive controls which are added to the extracted DNA and are only detectable afterwards in the absence of any inhibition. Careful consideration of the effect of inhibitors should be central to any eDNA project, but also the choice of an appropriate inhibitor-releasing strategy. Dilution of the DNA and the co-extracted inhibitors is very common, but this strategy also dilutes the target DNA and so reduces the detection probability of the target (e.g. McKee et al. 2015). Post-extraction purification columns (e.g. Qiagen's PowerClean or Zymo's One-Step) can also be used to remove inhibitory compounds. PCR enhancers such as bovine serum albumin, Dimethyl sulfoxide, Betaine, inhibitor-resistant polymerases, etc. can also be used to alleviate the effect of inhibitors. As yet the effect of different inhibitor-releasing strategies has not been fully evaluated with eDNA metabarcoding as all of the eDNA inhibitor comparisons have been tested with qPCR techniques.

False positives associated with contamination

There is a high risk of contamination, which can be introduced at any stage from sampling to sequencing. This risk is higher due to the sensitive nature of eDNA techniques; for example, human DNA can be introduced into the sample by human contact with the waterbody, by the sampler, or at any stage in the laboratory process. These contamination issues can be monitored with effective and judicious use of negative controls (in the field and lab) and minimised with substantial preventative measures. Common laboratory practises associated with eDNA include bleach and UV sterilisation, strict decontamination protocols for all equipment and clothing, pre-packaged sterile filter tips for pipettes, double gloves, dedicated eDNA extraction labs, etc.

Despite extensive measures taken to effectively minimise and prevent contamination at all stages of the process, these issues can still influence the final output. The risk of allochthonous DNA (i.e. DNA originating from outside of the system) is still apparent. The sensitivity of these methods means that false positive detections can arise from prey matter deposited by birds in their faeces (Merkes et al. 2014), food waste (Thomsen et al. 2016; Stoeckle et al. 2017), fishing baits, and even contaminants present in laboratory reagents

(Leonard et al. 2007). Dealing with allochthonous DNA detection is a difficult but important limitation that can only be dealt with through the use of controls and a level of truth sensing the data. Notably, care should be taken in interpreting detections of Atlantic Salmon, as this seems to be a common environmental contaminant.

Moving forward

All of the aforementioned issues with eDNA are associated with a burgeoning technique in its infancy and as with any powerful new tool these growing pains are being tackled by a proactive academic field. For example, to fine tune how abundance information is obtained from sequencing, researchers at the Kyoto University, Japan are using additional controls to better model how sequence reads match with fish species abundance and biomass (Ushio et al. 2018), and to address how best to sample and process eDNA for fish, researchers across Europe as part of an EU COST Action are collaborating to determine the best methods to improve the quality and quantity and standardisation of eDNA collection techniques.

Even with these teething issues, we can already produce reliable data at a scale that would be impractical with conventional techniques. For example, a single person can sample multiple water bodies in a single day, which can be analysed simultaneously. Moreover, the sample DNA samples can be analysed for multiple taxonomic groups (i.e. amphibians as well as fish). We have further developed the pipeline used here, with additional steps and tweaks that include a different higher capacity filter, a more rigorous DNA extraction protocol, significantly more PCR replicates, and more automated and thorough bioinformatic pipeline. Our pipeline is improving iteratively as we learn about and overcome limitations of the method, so far we have analysed water samples from ponds, lakes, streams, rivers, estuaries, coasts and the open ocean, in locations ranging from the equator to the North Pole.

So, despite eDNA being typically difficult to work with, there have been significant and recent advances in how to improve the detection probabilities of target organisms from eDNA. We know there are limitations to eDNA techniques, but these continue to be dealt with in a transparent and rigorous way, so that increasingly, eDNA metabarcoding will provide robust biodiversity information from which reliable inferences for ecology and conservation can be made.

Invertebrates

DNA monitoring holds great potential for studying invertebrate biodiversity, which is difficult to monitor using traditional methods that are often limited in taxonomic resolution, accuracy, and throughput (Ji et al. 2013). Combined, these limitations mean that biomonitoring projects are downsized with taxonomically restricted sampling associated with taxa that are easy to monitor and rarely attempt to encompass the full breadth of diversity present.

This pilot work included five invertebrate projects covering different types of samples: terrestrial invertebrates from deadwood mould and traps, freshwater invertebrates from eDNA, and marine invertebrates from eDNA and from sediment samples. Each of the sets of taxa and sample types had their own specific limitations, but the main issues concern the ratio of target organism DNA to non-target co-extracted DNA, and the taxonomic breadth

encompassed by the term ‘invertebrates’.

Invertebrate metabarcoding differs from fish metabarcoding in that the taxonomic diversity of the group means that it is not possible to design primers that are a perfect match to all taxa. Therefore, primers are designed to be ‘degenerate’ to accommodate variability in the priming sites. This reduces the specificity of the primers and makes it much more difficult to avoid co-amplification of non-target taxa. Where the target taxa form the bulk of the sample mass (e.g. in the case of insect trap samples), this does not pose a particular problem because target DNA (invertebrate metazoans) will dominate the sample, but in an environmental sample the target DNA makes up only a tiny fraction of the total DNA, and the use of degenerate primers typically leads to amplification of microbial or zooplankton DNA that is present at higher concentrations.

A second consequence of using very general primers to amplify a broad taxonomic group is that there will inevitably be variability in how well different taxa match the primers, and this influences amplification efficiency. Taxa that have the best match to the primers will be amplified much more efficiently than those that have some mismatches in the priming sites, and this can have a significant influence on the number of sequences obtained for the different taxa. Essentially this ‘primer bias’ (also known as ‘amplification bias’) disrupts the link between amount of starting DNA and number of sequences obtained, making it difficult to draw quantitative comparisons between taxa.

Invertebrates from traps

Characterising invertebrate assemblages from trap samples is a well-tested application for DNA metabarcoding (e.g. Yu et al. 2012; Ji et al. 2013, Theissinger et al. 2018). Trap samples are relatively simple to analyse because the target DNA is in high concentration. Here we were successfully able to characterise terrestrial invertebrates caught in pitfall and malaise traps set around Lampert Mosses. This landscape is typically species-depauperate, but with an important number of nationally rare insects (e.g. *Spilogona depressiuscula*, *Coenosia paludis*, *Agonum ericeti*, *Carabus nitens*, *Coenonympha tullia*).

For the Lampert Mosses samples, a total of 39 taxa belonging to 19 families and 30 genera were detected across the three samples, with Diptera dominating the Malaise traps.

There are many factors that influence how accurately the metabarcoding output reflects the true species composition on the sample. Fortunately, this is an active field of development and many groups are tackling the effect of methodological factors such as unequal specimen biomass (Elbrecht et al. 2017a), primer choice (Elbrecht et al. 2017b), sequencing and bioinformatic approaches (Wilson et al. 2018), PCR cycle conditions (Clarke et al. 2017), taxonomic placement (Somervuo et al. 2017), etc.

Invertebrates from aquatic eDNA

Sampling for aquatic invertebrate eDNA is difficult primarily because eDNA captured and extracted from filters is shrouded by an overabundance of DNA extracted from organisms (bacteria, small invertebrates) that are also trapped on the filter. We have found that the primers typically used for invertebrate analysis (typically targeting cytochrome oxidase c subunit 1) are inappropriate for eDNA work because they also amplify bacterial DNA, which is massively abundant in aquatic eDNA samples. Moreover, DNA extracted from the filter will

include that extracted from small animals caught in the filter, this DNA will likely overpower true eDNA.

A set of filters from Yardley Chase were analysed with the aim of detecting Odonata, for which this particular location is important. Over 400,000 sequences were generated for two different primer pairs targeting different sections of the COI gene, which resulted in the detection of 51 and 36 taxa. A small proportion of these taxa could be identified as metazoans with a high degree of certainty, including two species of fungi, a mayfly, a copepod, an ostracod and a midge, but no Odonata. Clearly this is a very poor representation of the macroinvertebrate component of freshwater diversity. It is likely that the DNA of many other species is present in the sample but is shrouded by the amplification of non-target (mostly prokaryotic) organisms when these very broad-spectrum primers are used. Indeed, 78% of sequences were attributed to bacteria or fungi and much of the remainder were identified as *Cloeon dipterum*, a nymph that likely to have been caught in the filter. No vertebrate sequences were detected, including fish, which were readily sequenced with more fish specific primers.

Another issue associated with eDNA metabarcoding of aquatic invertebrates is that the amount of DNA shed by certain groups - particularly those with hard exoskeletons - is likely to be far less than is the case for fish, amphibians, and soft-bodied invertebrates such as molluscs or worms. This will cause variance in detectability and also in the number of sequences obtained for a given body mass, adding yet more noise to the relationship between sequence output and number of organisms.

The central challenge in invertebrate eDNA monitoring is finding the balance between primer specificity and universality. Designing species-specific primers for single-species detection is relatively common, but it is not a practical approach for analysing the wider community. An alternative approach would be to design a suite of primers each specific to a different group of invertebrates. It may be the case that COI is not a good gene for any eDNA metabarcoding application, since even when group-specific primers are designed (as has been done for fish), the target taxa still make up only a small minority of sequences obtained (e.g. Bakker et al. 2017). Some researchers have reported better outcomes using other genes (e.g. 16S rDNA - Klymus et al. 2017), but the challenge then becomes the lack of reference sequences available for these genes.

Invertebrates from sediment

Two projects focused on monitoring invertebrates (single species and assemblages) in sediment samples: 1) detecting the rare and endangered violet click beetle (*Limoniscus violaceus*) from wood mould cores and 2) characterising invertebrate assemblages from marine intertidal sediment samples.

The first project was challenging because it was not possible to ascertain whether the tree mould samples were expected to contain violet click beetle DNA, meaning that we do not know whether our results represent true negatives or false negatives. Across all of the samples obtained for the wood mould samples a total of 1.45 million unique sequences were obtained, which were clustered into a total of 344 taxa (including 100 Metazoa). A significant proportion of the sequences could not be identified to species-level, which is likely because a lot of the taxa present in the wood mould has not been sequenced for this particular gene.

No violet click beetle DNA was found in the environmental samples. It is possible that violet click beetle DNA is not detected due to insensitivity, inhibition, or being overwhelmed by more prevalent non-target DNA. However, it is also possible that our results represent true negatives since these are difficult organisms to find and the chances of traces of DNA being left behind in a detectable state in the small core used is low. Greater sampling effort may be needed. Since carrying out this work, we have used DNA from the tissue sample obtained from Germany to design species-specific primers for the violet click beetle and are in the process of testing and optimising them.

eDNA detection of invertebrate in sediment cores is hampered by the same issues as eDNA detection of invertebrates in water; namely, overabundance of non-target DNA present in the sample, non-target primer binding and amplification, and low levels of DNA sloughing. Moreover, sediment samples are also hampered by the dominance of inhibitors (e.g. humic acids) and contain a multitude of unsequenced and unknown species for which no reference sequences are available.

Characterising the community in marine sediment is a very different challenge to finding an elusive single species in tree mould. We used two different sets of primers targeting two different genes (COI and 18S) for which there is a history of use in marine benthic metabarcoding (e.g. Leray and Knowlton 2015; Lejzerowicz et al. 2015). A single sample of marine sediment was collected and split into 5 subsamples, which were subsequently sequenced using COI (Leray et al. 2013) and 18S hypervariable region 4 (Bradley et al. 2016) primers. After bioinformatic processing, only 7 and 33 metazoan taxa were identified respectively by the different primer sets. A substantial amount of sequence data (68% and 60% for COI and 18S, respectively) was discarded because it had no close match in any available reference database or because it was identified as non-metazoan. The most prevalent taxa were the molluscs Laver spire shell (*Hydrobia ulvae*), Baltic clam (*Macoma balthica*), and common cockle (*Cerastoderma edule*), and a copepod (*Centropages* species). These species were present in both the COI and 18S datasets.

There are notable discrepancies between the two datasets, which is due to a combination of the facts that 18S evolves at a slower rate than the COI marker and so species will typically be less divergent than their COI counterparts (Tang et al. 2012), the 18S marine database is more populated than the COI marine database, and that COI and 18S primers detect different subsets of diversity.

18S detected a wider range of taxa, including some phyla that were not detected at all by COI. This is particularly striking in the case of platyhelminths (flatworms), which are a key group of marine benthic organisms. On the other hand, 18S is less able to separate closely-related species than is COI, so if species identity is of importance then 18S may not be adequate on its own. It is therefore possible that a multi-marker approach may be required, and there are other genes / primers that have not been tested here but may return useful data at the species level.

Plants

Identifying plants from their roots is challenging using molecular taxonomy, but a key benefit is the potential for year-round surveys of vegetation. A DNA based approach may help to avoid the subjectivity and biases of morphological taxonomy, which are a particular issue

when non-specialists (e.g. volunteers) are responsible for identification. Adopting a high-throughput DNA based methodology for the identification of plants from their roots is possible but is challenging in terms of both laboratory workflows and data analysis.

Unique to this application is the need for a high-throughput means of separating root material from large soil samples in such a way that does not risk cross-contamination occurring between samples. The process developed here was a slow and labour intensive series of manual hand washes, which inevitably resulted in finer root hairs being washed away and a lot of soil was ultimately still retained. A cleaner means of extracting root material would result in purer plant DNA free from inhibitors, which might alleviate some of the post DNA extraction limitations (i.e. species dropout due to low concentrations or DNA or inhibition).

As ever, there is a trade-off among *rbcL* and ITS2 in species resolution, primer universality, and also the completeness of the reference database. We found that, *rbcL* is able to detect more taxa and has a more complete reference database. This is not surprising given that *rbcL* is one of two 'plant DNA barcoding' genes (CBOL Plant Working Group 2009) chosen because of its high quality sequence data over its whole length and its universality across a wide range of taxa. As such this has resulted in an ever expanding reference database (e.g. de Vere et al. 2012). For ITS2, the taxonomic resolution of those taxa detected was higher than for *rbcL*, which is consistent with published research (e.g. Chen et al. 2010) and has already led others to suggest the official adoption of ITS as an alternative or complementary DNA barcode gene (Yao et al. 2010; Hollingsworth 2011). An additional challenge of applying ITS metabarcoding to root samples is that ITS is a very effective gene for metabarcoding of fungi, and a significant portion of the sequences returned were identified as fungal taxa. This means that the analysis is less efficient for plant biomonitoring, and means that greater sequencing depth per sample may be needed to compensate for the data lost to fungi.

Despite these difficulties, sequencing the 20 root samples using two different primer sets / genes (*rbcL* and ITS2) yielded a total of 31 and 20 plant taxa respectively, belonging to 19 different families. An initial comparison with the conventional survey data indicates that the main taxa are similar, although the DNA data are less well resolved (many closely-related species pairs cannot be distinguished from one another) and there are some spurious determinations. It should be noted that the field surveys incorporate the whole 40 cm x 40 cm above ground cell while the soil sample used for the DNA covered a much smaller area in the centre of the cell. It should also be noted that DNA metabarcoding of clean leaf tissue is much more straightforward and easy than it is for buried root tissue.

The benefit of a high throughput plant community identification system is huge, and these data go some way in showing that even with dirty soiled roots it is possible to obtain sensible botanical data. Of course there is still a great deal of research and development needed for these methods both before and after DNA extraction. Moving forward, a detailed evaluation of how well DNA metabarcoding compares to traditional taxonomy with either above-ground material or with root material (in soil) is clearly warranted.

Conclusions

DNA monitoring is a new and fast-developing field, and it is inevitable that some teething

problems will be encountered as we attempt to combine these new techniques and technologies with the more familiar ones. However, it is also likely that issues that seem to present significant barriers today will soon have solutions as research advances.

Despite the myriad challenges, the use of DNA for detecting and identifying species has the potential to dramatically increase the scale and resolution at which we can obtain data on species in the environment and help to shine a light on taxa that have been difficult to observe.

The work undertaken in this collaboration is a useful starting point for exploring the promises and pitfalls of applying DNA metabarcoding to real-world monitoring challenges. Due to limitations in time and resources together with the breadth of applications considered, it was not possible to design and test new primers for individual projects. Therefore, the projects consisted very much of a first-pass analysis using published primer sets. In many cases, development of more targeted primers is required to overcome the challenges associated with environmental samples in which the concentration of target DNA is extremely low, both in absolute terms and in relation to DNA from non-target taxa.

The most successful pilots were those that focused on eDNA monitoring of fish. Fish consistently shed DNA into the water, which increases their detectability relative to many other organisms. Moreover, the primers used (MiFish 12S primers) were designed specifically for this group. This is a useful demonstration of the importance of targeted primer design, since the MiFish primers were able to overcome the low ratio of target DNA to non-target DNA. Indeed, DNA extracts that yielded many fish species using the MiFish primers yielded none using more general metazoan primers, suggesting that greater success is likely to be achieved for other groups with appropriate effort applied to primer design and optimisation. While limitations do exist for surveying fish using eDNA, these are fairly well understood and are being actively addressed by the wider research community (e.g. via DNAqua-Net). Furthermore, it should be remembered that *all* survey methodologies have associated limitations and biases, and in the case of fish monitoring these do not seem to be vastly greater for eDNA approaches than for conventional ones. While eDNA cannot yield size or age-class data and is only semi-quantitative, it consistently detects more species than conventional methods such as electrofishing and gillnetting (e.g. Valentini *et al*., 2016; Environment Agency, 2017). A key area for further research is the optimal sampling strategy in different types of water bodies, since there is as yet little consensus on the volume of water that should be filtered, or the number or spatial arrangement of sampling or sub-sampling points.

The next most developed application is DNA metabarcoding of invertebrate trap samples, which is facilitated by the fact that the DNA of the invertebrates is readily extractable from the specimens themselves. Having high concentration target DNA increases the efficiency of all of the downstream processes (e.g. DNA extractions are most concentrated, PCR reagents are not wasted, sequencing depth is more targeted, etc.). This is to a large extent also true of the marine benthic invertebrates, where large organisms such as molluscs were easily detectable since they contributed large amounts of DNA to the sample. Nonetheless, the sediment sample yielded more non-metazoan diversity that could obscure smaller metazoan taxa. More work is needed to fully evaluate the species lists that come from traditional and molecular methods in both cases and it is possible that multiple primer sets would ultimately be required to capture the full breadth of diversity often contained in these

samples. Moreover, it remains a challenge to make this application truly scalable in the laboratory. Homogenisation equipment that allows the parallel grinding of large numbers of samples are typically designed for samples of just a few millilitres in volume, and this is far too small for the contents of most insect traps.

Botanical surveys from roots present in soil cores were successful in that many expected identifications were retrieved. Here the physical presence of the plant roots resulted in a moderate ratio of plant to non-target DNA, although a substantial amount of sequencing depth was lost to fungal taxa, particularly in the ITS2 dataset. It is already widely accepted that molecular identification of plants usually requires more than a single gene region due to lack of genetic variability within some families (here we found that the grasses are particularly unresolved) compounded by factors such as hybridisation and lack of consensus in morphological taxonomy. Here too there is a challenge in scaling this approach from a few sample to a national-scale monitoring scheme that would require hundreds or thousands of samples to be processed. Since the majority of DNA work in soil has focused on microbial diversity, DNA extraction kits and associated equipment are designed for much smaller samples - typically just a few millilitres, while the soil cores processed here were each several hundred millilitres in volume. Handling large numbers of such samples - especially in a way that avoids cross-contamination - is currently a major challenge. Freeze grinding followed by subsampling would be an interesting approach to explore going forward.

The least success was achieved in the projects that sought to detect invertebrate taxa in environmental samples (water or wood mould). While we can't be certain that violet click beetle DNA was present in the wood mould samples, we can be sure that the metazoan diversity recovered from the water samples in both freshwater ponds and tidal lagoons was a vast underestimate of the true diversity. This was due to the tiny ratio of target DNA to non-target DNA, and the use of very non-specific primers that amplified the more abundant non-metazoan DNA in the samples. Slightly greater success has been reported in recovering invertebrate community data from flowing water than from standing water, which may be because standing water supports a much higher density of microbial and planktonic organisms compared with streams and rivers. Further work is required to design more species- or group-specific primers to target violet click beetle, Odonata, lagoon sand shrimp and starlet sea anemone, rather than pursuing a community-based approach. For instance, Bista et al. (2017) were successful in characterising chironomid diversity with a chironomid specific set of COI primers.

The findings of these eight pilots indicate that DNA applications can make a significant contribution to a multitude of challenging tasks. These pilots have identified the likely potential and risks of specific DNA applications for various difficult-to-survey taxa. We have shown that it is possible to obtain fish diversity data for elusive communities of fish, characterise whole invertebrate communities that would otherwise have had to be identified by eye, and even obtained floristic information from roots buried in soil. While we acknowledge there are several important knowledge gaps in all of these pilots, we highlight avenues of research that can help to overcome them.

References

Andruszkiewicz, E. A., Starks, H. A., Chavez, F. P., Sassoubre, L. M., Block, B. A., & Boehm, A. B. (2017). Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding. PloS one, 12(4), e0176343.

Arreguín-Sánchez, F. (1996). Catchability: a key parameter for fish stock assessment. Reviews in fish biology and fisheries, 6(2), 221-242.

Arribas, P., Andujar, C., Hopkins, K., Shepherd, M., & Vogler, A. P. (2016). Metabarcoding and mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil. Methods in Ecology and Evolution, 7(9), 1071-1081.

Baird, D. J., & Hajibabaei, M. (2012). Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. Molecular ecology, 21(8), 2039-2044.

Bakker, J., Wangensteen, O. S., Chapman, D. D., Boussarie, G., Buddo, D., Guttridge, T. L., Hertler, H., Mouillot, D., Vigliola, L. & Mariani, S. (2017). Environmental DNA reveals tropical shark diversity in contrasting levels of anthropogenic impact. Scientific reports, 7(1), 16886.

Barnes, M. A., Turner, C. R., Jerde, C. L., Renshaw, M. A., Chadderton, W. L., & Lodge, D. M. (2014). Environmental conditions influence eDNA persistence in aquatic systems. Environmental science & technology, 48(3), 1819-1827.

Bhadury, P., Austen, M. C., Bilton, D. T., Lambshead, P. J. D., Rogers, A. D., & Smerdon, G. R. (2006). Development and evaluation of a DNA-barcoding approach for the rapid identification of nematodes. Marine Ecology Progress Series, 320, 1-9.

Biggs, J., Ewald, N., Valentini, A., Gaboriaud, C., Griffiths, R. A., Foster, J., Wilkinson, J., Arnett, A., Williams, P. & Dunn, F. (2014). Analytical and methodological development for improved surveillance of the Great Crested Newt. Defra Project WC1067. Freshwater Habitats Trust: Oxford.

Bista, I., Carvalho, G. R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., Christmas, M. & Creer, S. (2017). Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. Nature communications, 8, 14087.

Bohmann, K., Monadjem, A., Noer, C. L., Rasmussen, M., Zeale, M. R., Clare, E., Jones, G., Willerslev, E. & Gilbert, M. T. P. (2011). Molecular diet analysis of two African free-tailed bats (Molossidae) using high throughput sequencing. PloS one, 6(6), e21441.

Bradley, I. M., Pinto, A. J., & Guest, J. S. (2016). Design and evaluation of Illumina MiSeq-compatible, 18S rRNA gene-specific primers for improved characterization of mixed phototrophic communities. Applied and environmental microbiology,

82(19), 5878-5891.

Cai, W., Ma, Z., Yang, C., Wang, L., Wang, W., Zhao, G., Geng. Y. & Douglas, W. Y. (2017). Using eDNA to detect the distribution and density of invasive crayfish in the Honghe-Hani rice terrace World Heritage site. PloS one, 12(5), e0177724.

CBOL Plant Working Group (2009). A DNA barcode for land plants. Proceedings of the National Academy of Sciences, 106(31), 12794-12797.

Chambert, T., Pilliod, D. S., Goldberg, C. S., Doi. H. & Takahara, T. (2018). An analytical framework for estimating aquatic species density from environmental DNA. Ecology and evolution.

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Yingjie Zhu, Ma, X., Gao, T., Pang, X., Luo, K., Li, Y., Li, X., Jia, X., Lin, Y. & Luo, K. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PloS one, 5(1), e8613.

Chen, J., Li, Q., Kong, L., & Yu, H. (2011). How DNA barcodes complement taxonomy and explore species diversity: the case study of a poorly understood marine fauna. PloS one, 6(6), e21326.

Civade, R., Dejean, T., Valentini, A., Roset, N., Raymond, J. C., Bonin, A., Taberlet, P. & Pont, D. (2016). Spatial representativeness of environmental DNA metabarcoding signal for fish biodiversity assessment in a natural freshwater system. PloS one, 11(6), e0157366.

Clarke, L. J., Beard, J. M., Swadling, K. M., & Deagle, B. E. (2017). Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. Ecology and evolution, 7(3), 873-883.

Council of Europe (1979). Convention on the Conservation of European Wildlife and Natural Habitats, Bern.

De Vere, N., Rich, T. C., Ford, C. R., Trinder, S. A., Long, C., Moore, C. W., Satterthwaite, D., Davies, H., Allainguillaume, J., Ronca, S., Tatarinova, T., Garbett, H., Walker, K. & Wilkinson, M.J. (2012). DNA barcoding the native flowering plants and conifers of Wales. PLoS One, 7(6), e37945.

De Vere, N., Jones, L. E., Gilmore, T., Moscrop, J., Lowe, A., Smith, D., Hegarty, M.J., Creer, S. & Ford, C. R. (2017). Using DNA metabarcoding to investigate honey bee foraging reveals limited flower use despite high floral availability. Scientific reports, 7, 42838.

Deiner, K., Fronhofer, E. A., Mächler, E., Walser, J. C., & Altermatt, F. (2016). Environmental DNA reveals that rivers are conveyer belts of biodiversity information. Nature communications, 7, 12544.

Deiner, K., & Altermatt, F. (2014). Transport distance of invertebrate environmental DNA in a natural river. PloS one, 9(2), e88786.

Deiner, K., Walser, J. C., Mächler, E., & Altermatt, F. (2015). Choice of capture and

extraction methods affect detection of freshwater biodiversity from environmental DNA. Biological Conservation, 183, 53-63.

Díaz, J., Villanova, G. V., Brancolini, F., del Pazo, F., Posner, V. M., Grimberg, A., & Arranz, S. E. (2016). First DNA barcode reference library for the identification of South American freshwater fish from the lower Paraná river. PloS one, 11(7), e0157419.

Dincă, V., Zakharov, E. V., Hebert, P. D., & Vila, R. (2011). Complete DNA barcode reference library for a country's butterfly fauna reveals high performance for temperate Europe. Proceedings of the Royal Society of London B: Biological Sciences, 278(1704), 347-355.

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(7), e0130324.

Elbrecht, V., Peinert, B., & Leese, F. (2017a). Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. Ecology and evolution, 7(17), 6918-6926.

Elbrecht, V., Vamos, E. E., Meissner, K., Aroviita, J., & Leese, F. (2017b). Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. Methods in Ecology and Evolution, 8(10), 1265-1275.

Elbrecht, V., & Leese, F. (2017c). Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. Frontiers in Environmental Science, 5, 11.

English Nature (1994). Lampert Mosses citation. Sites of Special Scientific Interest.

English Nature (1990). Yardley Chase citation. Sites of Special Scientific Interest.

Environment Agency. (2017). eDNA-based metabarcoding as a monitoring tool for fish in large lakes: Report SC140018.

European Commission (1992). Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. Council of the European Communities (CEC). Official Journal of the European Communities, 206, 7–50.

Fahner, N. A., Shokralla, S., Baird, D. J., & Hajibabaei, M. (2016). Large-scale monitoring of plants through environmental DNA metabarcoding of soil: recovery, resolution, and annotation of four DNA markers. PLoS One, 11(6), e0157505.

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., Gielly, L., Lopes, C.M., Boyer, F., Pompanon, F. & Rayé, G. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Molecular Ecology Resources, 15(3), 543-556.

Folmer, O., Hoeh, W. R., Black, M. B., & Vrijenhoek, R. C. (1994). Conserved primers

for PCR amplification of mitochondrial DNA from different invertebrate phyla. Molecular Marine Biology and Biotechnology, 3, 294-299.

Fontaneto, D., Flot, J. F., & Tang, C. Q. (2015). Guidelines for DNA taxonomy, with a focus on the meiofauna. Marine Biodiversity, 45(3), 433-451.

Garrick-Maidment, N. (2011). British Seahorse Survey 2011, The Seahorse Trust, Escot, Devon.

Gentry-Shields, J., Wang, A., Cory, R. M., & Stewart, J. R. (2013). Determination of specific types and relative levels of QPCR inhibitors in environmental water samples using excitation–emission matrix spectroscopy and PARAFAC. Water research, 47(10), 3467-3476.

Gouix, N., Sebek, P., Valladares, L., Brustel, H., & Brin, A. (2015). Habitat requirements of the violet click beetle (Limoniscus violaceus), an endangered umbrella species of basal hollow trees. Insect Conservation and Diversity, 8(5), 418-427.

Gouix, N., & Brustel, H. (2012). Emergence trap, a new method to survey Limoniscus violaceus (Coleoptera: Elateridae) from hollow trees. Biodiversity and conservation, 21(2), 421-436.

Hajibabaei, M., Baird, D. J., Fahner, N. A., Beiko, R., & Golding, G. B. (2016). A new way to contemplate Darwin's tangled bank: how DNA barcodes are reconnecting biodiversity science and biomonitoring. Phil. Trans. R. Soc. B, 371(1702), 20150330.

Hänfling, B., Lawson Handley, L., Read, D. S., Hahn, C., Li, J., Nichols, P., Blackman, R., Oliver, A. & Winfield, I. J. (2016). Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. Molecular ecology, 25(13), 3101-3119.

Hayhow, D. B., Burns, F., Eaton, M. A., Al Fulaij, N., August, T. A., Babey, L., Bacon, L., Bingham, C., Boswell, J., Boughey, K.L., Brereton, T., Brookman, E., Brooks, D.R., Bullock, D.J., Burke, O., Collis, M., Corbet, L., Cornish, N., De Massimi, S., Densham, J., Dunn, E., Elliott, S., Gent, T., Godber, J., Hamilton, S., Havery, S., Hawkins, S., Henney, J., Holmes, K., Hutchinson, N., Isaac, N.J.B., Johns, D., Macadam, C.R., Mathews, F., Nicolet, P., Noble, D.G., Outhwaite, C.L., Powney, G.D., Richardson, P., Roy, D.B., Sims, D., Smart, S., Stevenson, K., Stroud, R.A., Walker, K.J., Webb, J.R., Webb, T.J., Wynde, R. & Gregory, R.D. (2016). State of nature 2016.

Hebert, P. D., Zakharov, E. V., Prosser, S. W., Sones, J. E., McKeown, J. T., Mantle, B., & La Salle, J. (2013). A DNA 'Barcode Blitz': Rapid digitization and sequencing of a natural history collection. PLoS One, 8(7), e68535.

Hollingsworth, P. M. (2011). Refining the DNA barcode for land plants. Proceedings of the National Academy of Sciences, 108(49), 19451-19452.

Hope, P. R., Bohmann, K., Gilbert, M. T. P., Zepeda-Mendoza, M. L., Razgour, O., & Jones, G. (2014). Second generation sequencing and morphological faecal

analysis reveal unexpected foraging behaviour by *Myotis nattereri* (Chiroptera, Vespertilionidae) in winter. Frontiers in zoology, 11(1), 39.

Hu, Q., Liu, Y., Yi, S., & Huang, D. (2015). A comparison of four methods for PCR inhibitor removal. Forensic Science International: Genetics, 16, 94-97.

Hunter, M. E., Oyler-McCance, S. J., Dorazio, R. M., Fike, J. A., Smith, B. J., Hunter, C. T., Reed, R.N. & Hart, K. M. (2015). Environmental DNA (eDNA) sampling improves occurrence and detection estimates of invasive Burmese pythons. PloS one, 10(4), e0121655.

Hyman, P. S. (1992). A review of the scarce and threatened Coleoptera of Great Britain: Part 1. Joint Nature Conservation Committee.

Illumina Inc. (2013). 16S Metagenomic Sequencing Library. Illumina.com, (B), 1–28. Available at: http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf.

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

Jane, S. F., Wilcox, T. M., McKelvey, K. S., Young, M. K., Schwartz, M. K., Lowe, W. H., Letcher, B.H. & Whiteley, A. R. (2015). Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. Molecular ecology resources, 15(1), 216-227.

Jerde, C. L., Chadderton, W. L., Mahon, A. R., Renshaw, M. A., Corush, J., Budny, M. L., Mysorekar, S. & Lodge, D. M. (2013). Detection of Asian carp DNA as part of a Great Lakes basin-wide surveillance program. Canadian Journal of Fisheries and Aquatic Sciences, 70(4), 522-526.

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P.M., Woodcock, P. Edwards, F.A., Larsen, T.H., Hsu, W.W., Benedick, S., Hamer, K.C., Wilcove, D.S., Bruce, C., Wang, X., Levi, T., Lott, M., Emerson, B.C. & Yu, D.W. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. Ecology letters, 16(10), 1245-1257.

Joint Nature Conservation Committee (2010). Council Directive 92/43/EEC on the Conservation of natural habitats and of wild fauna and flora. [online]. Available at: http://jncc.defra.gov.uk/page-1374.

Joint Nature Conservation Committee (2007). Second Report by the United Kingdom under Article 17 on the implementation of the Directive from January 2001 to December 2006. Peterborough, 1-9.

Joint Nature Conservation Committee (1981). Wildlife and Countryside Act 1981, London: The Stationary Office.

Kartzinel, T. R., Chen, P. A., Coverdale, T. C., Erickson, D. L., Kress, W. J., Kuzmina,

M. L., Rubenstein, D.I., Wang, W. & Pringle, R.M. (2015). DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. Proceedings of the National Academy of Sciences, 112(26), 8019-8024.

Kelly, R. P., Port, J. A., Yamahara, K. M., Martone, R. G., Lowell, N., Thomsen, P. F., Mach, M.E., Bennett, M., Prahler, E., Caldwell, M.R. & Crowder, L. B. (2014). Harnessing DNA to improve environmental management. Science, 344(6191), 1455-1456.

Klymus, K. E., Marshall, N. T., & Stepien, C. A. (2017). Environmental DNA (eDNA) metabarcoding assays to detect invasive invertebrate species in the Great Lakes. PloS one, 12(5), e0177643.

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

Leese F, Altermatt F, Bouchez A, Ekrem T, Hering D, Meissner K, Mergen P, Pawlowski J, Piggott J, Rimet F, Steinke D, Taberlet P, Weigand A, Abarenkov K, Beja P, Bervoets L, Björnsdóttir S, Boets P, Boggero A, Bones A, Borja Á, Bruce K, Bursić V, Carlsson J, Čiampor F, Čiamporová-Zatovičová Z, Coissac E, Costa F, Costache M, Creer S, Csabai Z, Deiner K, DelValls Á, Drakare S, Duarte S, Eleršek T, Fazi S, Fišer C, Flot J, Fonseca V, Fontaneto D, Grabowski M, Graf W, Guðbrandsson J, Hellström M, Hershkovitz Y, Hollingsworth P, Japoshvili B, Jones J, Kahlert M, Kalamujic Stroil B, Kasapidis P, Kelly M, Kelly-Quinn M, Keskin E, Kõljalg U, Ljubešić Z, Maček I, Mächler E, Mahon A, Marečková M, Mejdandzic M, Mircheva G, Montagna M, Moritz C, Mulk V, Naumoski A, Navodaru I, Padisák J, Pálsson S, Panksep K, Penev L, Petrusek A, Pfannkuchen M, Primmer C, Rinkevich B, Rotter A, Schmidt-Kloiber A, Segurado P, Speksnijder A, Stoev P, Strand M, Šulčius S, Sundberg P, Traugott M, Tsigenopoulos C, Turon X, Valentini A, van der Hoorn B, Várbíró G, Vasquez Hadjilyra M, Viguri J, Vitonytė I, Vogler A, Vrålstad T, Wägele W, Wenne R, Winding A, Woodward G, Zegura B, Zimmermann J (2016) DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. Research Ideas and Outcomes 2: e11321.

Lejzerowicz, F., Esling, P., Pillet, L., Wilding, T. A., Black, K. D., & Pawlowski, J. (2015). High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. Scientific reports, 5, 13932.

Leonard, J. A., Shanks, O., Hofreiter, M., Kreuz, E., Hodges, L., Ream, W., Wayne, R.K. & Fleischer, R. C. (2007). Animal DNA in PCR reagents plagues ancient DNA research. Journal of Archaeological Science, 34(9), 1361-1366.

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Joehm, J.T. & Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Frontiers in zoology, 10(1), 34.

Leray, M., & Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized

samples reveal patterns of marine benthic diversity. Proceedings of the National Academy of Sciences, 112(7), 2076-2081.

Mallet, J., & Willmott, K. (2003). Taxonomy: renaissance or Tower of Babel? Trends in Ecology & Evolution, 18(2), 57-59.

McKee, A. M., Spear, S. F., & Pierson, T. W. (2015). The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. Biological Conservation, 183, 70-76.

Merkes, C. M., McCalla, S. G., Jensen, N. R., Gaikowski, M. P., & Amberg, J. J. (2014). Persistence of DNA in carcasses, slime and avian faeces may affect interpretation of environmental DNA data. PLoS One, 9(11), e113346.

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., 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(7), 150088.

Nieto, A. & Alexander, K.N. (2010). European red list of saproxylic beetles. IUCN (International Union for Conservation of Nature).

O'Donnell, J. L., Kelly, R. P., Shelton, A. O., Samhouri, J. F., Lowell, N. C., & Williams, G. D. (2017). Spatial distribution of environmental DNA in a nearshore marine habitat. PeerJ, 5, e3044.

Oliveira, L. M., Knebelsberger, T., Landi, M., Soares, P., Raupach, M. J., & Costa, F. O. (2016). Assembling and auditing a comprehensive DNA barcode reference library for European marine fishes. Journal of fish biology, 89(6), 2741-2754.

Perrow, M. R., & Jowitt, A. J. (1999). The habitat and management requirements of spined loach *Cobitis taenia*. Nature Conservancy Council for England.

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

Port, J. A., O'Donnell, J. L., Romero-Maraccini, O. C., Leary, P. R., Litvin, S. Y., Nickols, K. J., ... & Kelly, R. P. (2016). Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. Molecular ecology, 25(2), 527-541.

Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Resources, 7(3), 355-364.

Roux, K. H. (1995). Optimization and troubleshooting in PCR. Genome Research, 4(5), S185-S194.

Schnell, I. B., Sollmann, R., Calvignac-Spencer, S., Siddall, M. E., Douglas, W. Y., Wilting, A., & Gilbert, M. T. P. (2015). iDNA from terrestrial haematophagous leeches as a wildlife surveying and monitoring tool–prospects, pitfalls and avenues to be developed. Frontiers in zoology, 12(1), 24.

Schnell, I. B., Thomsen, P. F., Wilkinson, N., Rasmussen, M., Jensen, L. R., Willerslev, E., Bertelsen, M.F. & Gilbert, M. T. P. (2012). Screening mammal biodiversity using DNA from leeches. Current biology, 22(8), R262-R263.

Schrader, C., Schielke, A., Ellerbroek, L., & Johne, R. (2012). PCR inhibitors–occurrence, properties and removal. Journal of applied microbiology, 113(5), 1014-1026.

Shen, Y. Y., Chen, X., & Murphy, R. W. (2013). Assessing DNA barcoding as a tool for species identification and data quality control. PLoS One, 8(2), e57125.

Shirt, D. ed. (1987). British Red Data Books: 2. Insects, Peterborough: Joint Nature Conservancy Council.

Sickel, W., Ankenbrand, M. J., Grimmer, G., Holzschuh, A., Härtel, S., Lanzen, J., Steffan-Dewenter, I. & Keller, A. (2015). Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. BMC ecology, 15(1), 20.

Smart, A. S., Tingley, R., Weeks, A. R., van Rooyen, A. R., & McCarthy, M. A. (2015). Environmental DNA sampling is more sensitive than a traditional survey technique for detecting an aquatic invader. Ecological applications, 25(7), 1944-1952.

Somervuo, P., Koskela, S., Pennanen, J., Henrik Nilsson, R., & Ovaskainen, O. (2016). Unbiased probabilistic taxonomic classification for DNA barcoding. Bioinformatics, 32(19), 2920-2927.

Somervuo, P., Yu, D. W., Xu, C. C., Ji, Y., Hultman, J., Wirta, H., & Ovaskainen, O. (2017). Quantifying uncertainty of taxonomic placement in DNA barcoding and metabarcoding. Methods in Ecology and Evolution, 8(4), 398-407.

Song, H., Buhay, J. E., Whiting, M. F., & Crandall, K. A. (2008). Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proceedings of the national academy of sciences, 105(36), 13486-13491.

Spens, J., Evans, A. R., Halfmaerten, D., Knudsen, S. W., Sengupta, M. E., Mak, S. S., Sigsgaard, E.E. & Hellström, M. (2017). Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. Methods in Ecology and Evolution, 8(5), 635-645.

Steinke, D., R deWaard, J., Gomon, M. F., Johnson, J. W., Larson, H. K., Lucanus, O., Moore, G.I., Reader, S. & Ward, R. D. (2017). DNA barcoding the fishes of Lizard Island (Great Barrier Reef). Biodiversity data journal, (5).

Stoeckle, M. Y., Soboleva, L., & Charlop-Powers, Z. (2017). Aquatic environmental DNA detects seasonal fish abundance and habitat preference in an urban estuary. PloS one, 12(4), e0175186.

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. Molecular ecology, 21(8), 1789-1793.

Takahara, T., Minamoto, T., Yamanaka, H., Doi, H., & Kawabata, Z. I. (2012). Estimation of fish biomass using environmental DNA. PloS one, 7(4), e35868.

Tang, C. Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T. G., & Fontaneto, D. (2012). The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. Proceedings of the National Academy of Sciences, 109(40), 16208-16212.

Telfer, A. C., Young, M. R., Quinn, J., Perez, K., Sobel, C. N., Sones, J. E., Levesque-Beaudin, V., Derbyshire, R., Fernandez-Triana, J., Rougerie, R., Thevanayagam, A., Boskovic, A, Borisenko, A.V., Cadel, A., Brown, A., Pages, A., Castillo, A.H., Nicolai, A., Glenn Mockford, B.M., Bukowski, B., Wilson, B., Trojahn, B., Lacroix, C.A., Brimblecombe, C., Hay, C., Ho, C., Steinke, C., Warne, C.P., Garrido Cortes, C., Engelking, D., Wright, D., Lijtmaer, D.A., Gascoigne, D., Hernandez Martich, D., Morningstar, D., Neumann, D., Steinke, D., Marco DeBruin, D.D., Dobias, D., Sears, E., Richard, E., Damstra, E., Zakharov, E.V., Laberge, F., Collins, G.E., Blagoev, G.A., Grainge, G., Ansell, G., Meredith, G., Hogg, I., McKeown, J., Topan, J., Bracey, J., Guenther, J., Sills-Gilligan, J., Addesi, J., Persi, J., Layton, K.K., D'Souza, K., Dorji, K., Grundy, K., Nghidinwa, K., Ronnenberg, K., Lee, K.M., Xie, L., Lu, L., Penev, L., Gonzalez, M., Rosati, M.E., Kekkonen, M., Kuzmina, M., Iskandar, M., Mutanen, M., Fatahi, M., Pentinsaari, M., Bauman, M., Nikolova, N., Ivanova, N.V., Jones, N., Weerasuriya, N., Monkhouse, N., Lavinia, P.D., Jannetta, P., Hanisch, P.E., McMullin, R.T., Ojeda Flores, R., Mouttet, R., Vender, R., Labbee, R.N., Forsyth, R., Lauder, R., Dickson, R., Kroft, R., Miller, S.E., MacDonald, S., Panthi, S., Pedersen, S., Sobek-Swant, S., Naik, S., Lipinskaya, T., Eagalle, T., Decaëns, T., Kosuth, T., Braukmann, T., Woodcock, T., Roslin, T., Zammit, T., Campbell, V., Dinca, V., Peneva, V., Hebert, P.D., deWaard, J.R. (2015). Biodiversity inventories in high gear: DNA barcoding facilitates a rapid biotic survey of a temperate nature reserve. Biodiversity data journal, (3).

Theissinger, K., Kästel, A., Elbrecht, V., Leese, F., & Brühl, C. (2018). Using DNA metabarcoding for assessing chironomid diversity and community change in mosquito controlled temporary wetlands. Metabarcoding and Metagenomics, 2, e21060.

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA–An emerging tool in conservation for monitoring past and present biodiversity. Biological Conservation, 183, 4-18.

Thomsen, P. F., Møller, P. R., Sigsgaard, E. E., Knudsen, S. W., Jørgensen, O. A., & Willerslev, E. (2016). Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater fishes. PloS one, 11(11), e0165252.

Tillotson, M. D., Kelly, R. P., Duda, J. J., Hoy, M., Kralj, J., & Quinn, T. P. (2018). Concentrations of environmental DNA (eDNA) reflect spawning salmon abundance at fine spatial and temporal scales. Biological Conservation, 220, 1-11.

Ushio, M., Murakami, H., Masuda, R., Sado, T., Miya, M., Sakurai, S., Yamanaka, H., Minamoto, T. & Kondoh, M. (2018). Quantitative monitoring of multispecies fish environmental DNA using high-throughput sequencing. Metabarcoding and

Metagenomics, 2, e23297.

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

Wägele, H., Klussmann-Kolb, A., Kuhlmann, M., Haszprunar, G., Lindberg, D., Koch, A., & Wägele, J. W. (2011). The taxonomist-an endangered race. A practical proposal for its survival. Frontiers in zoology, 8(1), 25.

Webb, J., Heaver, D., Lott, D., Dean, H.J., van Breda, J., Curson, J., Harvey, M.C., Gurney, M., Roy, D.B., van Breda, A., Drake, M., Alexander, K.N.A. and Foster, G. (2017). Pantheon - database version 3.7.4. [online] Available at: http://www./brc.ac.uk/pantheon/ [Accessed 28/08/2017].

Whitehead, P. F. (2003). Current knowledge of the violet click beetle *Limoniscus violaceus* (PWJ Müller, 1821) (Col., Elateridae) in Britain. In Proceedings of the second pan–European conference on saproxylic beetles. People's Trust for Endangered Species, Royal Holloway, University of London (pp. 57-65).

Whitworth, T. L., Dawson, R. D., Magalon, H., & Baudry, E. (2007). DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). Proceedings of the Royal Society of London B: Biological Sciences, 274(1619), 1731-1739.

Willette, D. A., Simmonds, S. E., Cheng, S. H., Esteves, S., Kane, T. L., Nuetzel, H., Pilaud, N., Rachmawati, R. & Barber, P. H. (2017). Using DNA barcoding to track seafood mislabeling in Los Angeles restaurants. Conservation Biology, 31(5), 1076-1085.

Williams, K. E., Huyvaert, K. P., & Piaggio, A. J. (2017). Clearing muddied waters: Capture of environmental DNA from turbid waters. PloS one, 12(7), e0179282.

Wilson, J. J., Brandon-Mong, G. J., Gan, H. M., & Sing, K. W. (2018). High-throughput terrestrial biodiversity assessments: mitochondrial metabarcoding, metagenomics or metatranscriptomics? Mitochondrial DNA Part A, 1-8.

Woodward, G., Gray, C., & Baird, D. J. (2013). Biomonitoring for the 21st century: New perspectives in an age of globalisation and emerging environmental threats. Limnetica, 32(2), 159-174.

Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y. & Chen, S. (2010). Use of ITS2 region as the universal DNA barcode for plants and animals. PloS one, 5(10), e13102.

Yoccoz, N. G., Bråthen, K. A., Gielly, L., Haile, J., Edwards, M. E., Goslar, T., Von Stedingk, H., Brysting, A.K., Coissac, E., Pompanon, F., Sønstebø, J.H., Miquel, C., Valentini, A., De Bello, F., Chave, J., Thuiller, W., Wincker, P., Cruaud, C.,

Gavory, F., Rasmussen, M., Gilbert, M.T., Orlando, L., Brochmann, C., Willerslev, E. & Taberlet, P. (2012). DNA from soil mirrors plant taxonomic and growth form diversity. Molecular Ecology, 21(15), 3647-3655.

Yu, D. W., Ji, Y., Emerson, B. C., 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(4), 613-623.

Zettler, M. L., Proffitt, C. E., Darr, A., Degraer, S., Devriese, L., Greathead, C., Kotta, J., Magni, P., Martin, G., Reiss, H., Speybroeck, J., Tagliapietra, D., Van Hoey, G. & Ysebaert, T. (2013). On the myths of indicator species: issues and further consideration in the use of static concepts for ecological applications. PLoS One, 8(10), e78219.

Annex 1. Detailed methods and materials

Sampling

eDNA sampling with filters

eDNA in aquatic environments can be intracellular (skin, excreta) or extracellular (free-floating), and needs to be captured for DNA extraction. Unlike the standardised GCN protocol, which uses a precipitation method and is limited to a total of 90 ml per sample, filtration-based eDNA capture methods consistently yield higher DNA concentrations primarily because they concentrate eDNA sources from much larger volumes of water. Higher eDNA yields are particularly important for the detection of low density targets, and so filtration was used for all aquatic eDNA projects.

Many different types of filters exist for the purification of water, and these have been repurposed for the capture of eDNA. To facilitate on-site filtration and to reduce contamination associated with open filtration systems, closed capsule and portable Sterivex filters were used to filter water and capture eDNA as recommended by Spens et al. (2016). For all of the eDNA sampling, the same protocol was used and the resulting sample was three individual 0.45 µm Sterivex filters per sample site.

eDNA sampling from water: Protocol

1. Select 20 sampling points equally spaced around the water body.
2. Put on a pair of the gloves provided and try not to contact the water with any bare skin during the sampling process.
3. To make an extended sampling tube, attach the uncapped centrifuge tube to a stick with the cable tie.
4. Go to the first sampling point. From the bank, use the extended sampling tube to collect 50 ml of water and carefully pour this into the Whirlpak bag.
5. Repeat this for the remaining 19 sampling points. The Whirlpak bag should be full by the end of the process.
6. Mix the water and fill the syringe with 50 ml of water.
7. Attach the syringe to the filter inlet and press the plunger to push the water through the filter.
8. Repeat steps 6 – 7 until no more water can be passed through the filter. By the end it may take some force to push the water through the filter, particularly if there are high levels of sediment. Make a note on the sample sheet of how much water has passed through that filter.
9. Detach the syringe from the filter and pull back the plunger to fill the syringe with

air. Reattach the syringe to the filter and push the air through to expel the water from the filter. Repeat until the filter is dry. Cap the filter inlet and outlet.

10. Attach the blunt needle to the syringe and draw up with the preservative solution. Uncap the inlet for each of the three filters and fill with the preservative solution (The solution can be harmful to aquatic life, so perform this step away from the water's edge and try to avoid spills). Recap the filter inlets and return any unused preservative solution to the bottle.
11. Repeat steps 6 – 10 for the remaining 2 filters.
12. Total up the amount of water passed through the 3 filters.
13. Place the blunt needle into the preservative solution bottle.
14. Place filters and any remaining preservative solution into the resealable bag and into jiffy envelope.

Tissue homogenisation

Trap samples

These traps provide an excellent source of tissue for DNA metabarcoding (Yu et al. 2012; Ji et al. 2013), but first the samples need to be isolated from their preservative and homogenised so that DNA can be extracted from the community. Malaise and pitfall trap samples were used to collect invertebrate samples.

Homogenisation of invertebrate tissue from trap samples: Protocol

Tissue extraction

1. Pour off the preservative from the sample and transfer the sample onto a disposable drying tray.
2. Use a pipette to remove any excess preservative.
3. Cover with a paper towel and allow to air dry overnight.
4. Choose a homogeniser: for small traps (<25 mL) use the OmniPrep, otherwise use the NutriBullet.
    a. Optional: Large specimens will tend to have more DNA and that this overabundance of DNA can shroud that of smaller specimens. You can choose to size sort and subsample large individuals at this stage (e.g. the larger individuals, DNA can be extracted from legs and not the whole specimens).

Tissue homogenisation

5. Transfer the dried material into the homogeniser vessel.
6. Add enough PBS to the sample to either submerge the OmniPrep probe tip or to cover the sample (NutriBullet).
7. Secure the homogeniser to the vessel and blend the sample until it is smooth, some samples will take longer to homogenise.
8. Transfer the homogenate to a 50 ml centrifuge tube and flush any remnant homogenate from the vessel with a small amount of PBS. This is not necessary for the OmniPrep as it is already homogenised in a 50 ml tube.
9. Centrifuge at 5000 x *g* for 2 minutes to pellet homogenate and pour off any PBS.
10. If using a NutriBullet, clean the NutriBullet with detergent and 500 ml of freshly diluted 10% bleach, and soak the vessel and blades into a bucket of 10% bleach for 30 minutes. If using the OmniPrep then dispose of the probe tip.

Roots from soil

Soil cores were used to sample the plant roots. We developed a method to homogenise the plant roots without excessive soil carryover. To remove the roots from the soil required several sieving and transferring steps interspersed with UV sterilisation and bleach decontamination. Isolated roots were then homogenised.

Homogenisation of root material from soil cores: Protocol

Root separation

Steps 1-8 should be performed in a DNA free room (i.e. room separate to the DNA extraction lab).

1. Place the soil core/sample in a 2 L container and add 1.5 L of water.
2. Crush the sample by hand to loosen the soil.
3. Shake the container to help break up the sample.
4. Leave the sample to settle.
5. Label a 180 ml sample pot with the sample number on both the lid and the side.
6. Scoop any floating material off the water's surface and add it to the sample pot.
7. Sieve the sample with a 0.5 mm sieve and flush with water to wash off as much soil as possible from the roots through the sieve.
8. Transfer the contents on the sieve into the sample pot.
9. Transfer the sample pot to lab and decontaminate the sieve and pots with UV irradiation and an overnight 10% bleach soak.

Root Homogenisation

10. Pour the sample into a NutriBullet cup.
11. Some sample will remain in sample pot. Add ~30 ml PBS into the pot, swirl, and then add this to NutriBullet cup.
12. Secure the rotary blade to NutriBullet vessel and blend the sample until it is smooth, this will take approximately 10 seconds (additional blending may be required for some samples).
13. Transfer the homogenate to a new sample pot.
14. Add a small amount of PBS to the NutriBullet vessel (use remaining space in sample vessel to judge max amount of PBS), replace the blade, and shake to recover as much homogenate as possible. Transfer the recovered homogenate to

the corresponding sample pot.

15. Put the sample pot to one side and leave it to settle.
16. Clean the NutriBullet with detergent and 500 ml of freshly diluted 10% bleach, and soak the vessel and blades into a bucket of 10% bleach for 30 minutes.
17. Pour off the PBS from the settled homogenate.
18. Add 25 ml of the homogenate to a 50 ml centrifuge tube and centrifuge at 5000 x *g* for 2 minutes to pellet homogenate and pour off any PBS.

DNA extraction

DNA extraction: Filters

Samples were provided to the lab in the form of 0.45 µm Sterivex filters. These samples were preserved with Longmire's solution as provided in the NatureMetrics eDNA Sampling Kits. Each sample consisted of 3 separate filters. DNA was separately extracted from each of these broadly following protocols from Spens et al. (2016).

DNA extraction from filters: Protocol

Briefly, the preservative solution was removed from the filter units, which were air dried. The unit was then filled with lysis buffer (ATL buffer and Proteinase K from Qiagen) and incubated overnight at 56 °C. Following incubation, the lysate was removed from the filter unit and processed with a DNEasy Blood and Tissue Kit, following manufacturer's instructions.

DNA extraction as performed in a pre-PCR DNA extraction laboratory and all surfaces and equipment were sterilised and disinfected with 10% bleach and 70% ethanol and UV radiation.

1. Displace the Longmire's solution out of the Sterivex filters with air being careful not to cross contaminate the samples with the same syringe.
2. Air dry the filters for 10 minutes.
3. Lyse the filters with 720 µl of ATL and 80 µl of Proteinase K from the DNeasy Blood and Tissue extraction kit.
4. Incubate the capped Sterivex overnight at 56 °C.
5. Remove lysis buffer by pushing the lysate into a 1.5 ml low bind microcentrifuge tube.
6. Pipette equal amounts of AL buffer and lysate into a tube and incubate at 56 °C for 30 minutes.
7. Add equal amount of ice cold absolute EtOH as lysate and incubate for 10 minutes at room temperature.
8. Transfer contents to a spin column and centrifuge for 1 minute at 6,000 x *g*. Discard flow through and repeat until all lysate is used.
9. Transfer spin column to new collection tube and add 500 µl of AW1 to spin column and centrifuge for 1 minute at 6,000 x *g*. Discard flow-through and collection tube.
10. Transfer spin column to new collection tube and add 500 µl of AW2 to spin column and centrifuge for 3 minutes at 20,000 x *g*. Discard flow-through and collection tube.

11. Transfer spin column to a 1.5 ml low bind microcentrifuge tube, add 100 µl of AE buffer and incubate for 30 minutes at room temperature. Then centrifuge for 1 minute at 6,000 x *g*.

12. Repeat step 11.

DNA extraction: Homogenate

Homogenate samples can be from animal tissue or plant tissue. At NatureMetrics we use appropriate kits for animal and plant tissues because they have different biological compositions (including inhibitory substances) that confound DNA extraction. For homogenate from animal tissue we extract DNA using the DNeasy Blood and Tissue kit, while the homogenate from plants we use the DNeasy PowerSoil Kit (which we found to be better for removing inhibitors carried over from the soil as well as from the plant).

DNA extraction from animal tissue homogenate: Protocol

Three separate extractions are performed for each homogenate, and the remaining homogenate is stored at -20 °C in 100% ethanol.

1. Subsample 25-50 mg of the homogenate with a spatula into a 2 ml low bind tube.
2. Lyse the homogenate with 180 µl of ATL and 20 µl of Proteinase K from the DNeasy Blood and Tissue extraction kit.
3. Incubate the overnight at 56 °C.
4. Pipette equal amounts of AL buffer and lysate into a tube and incubate at 56 °C for 30 minutes.
5. Add equal amount of ice cold absolute EtOH as lysate and incubate for 10 minutes at room temperature.
6. Transfer contents to a spin column and centrifuge for 1 minute at 6,000 x *g*. Discard flow through and repeat until all lysate is used.
7. Transfer spin column to new collection tube and add 500 µl of AW1 to spin column and centrifuge for 1 minute at 6,000 x *g*. Discard flow-through and collection tube.
8. Transfer spin column to new collection tube and add 500 µl of AW2 to spin column and centrifuge for 3 minutes at 20,000 x *g*. Discard flow-through and collection tube.
9. Transfer spin column to a 1.5 ml low bind microcentrifuge tube, add 100 µl of AE buffer and incubate for 30 minutes at room temperature. Then centrifuge for 1 minute at 6,000 x *g*.
10. Repeat step 9.

DNA extraction from root homogenate: Protocol

Three separate extractions are performed for each homogenate, and the remaining homogenate is stored at -20 °C in 100% ethanol.

1. Subsample 25-50 mg of the homogenate with a spatula into a 1.5 ml low bind tube.
2. Add the PowerBeads and vortex to mix for 5 seconds.
3. Add 60 µl of Solution C1 and vortex briefly for 5 seconds.
4. Secure the tubes horizontally to the vortex and vortex at maximum speed for 10 minutes.
5. Centrifuge the tubes at 10,000 x *g* for 30 seconds.
6. Transfer the supernatent to a clean 2 ml tube.
7. Add 250 µl of Solution C2 and vortex for 5 seconds. Incubate at 4°C for 5 minutes.
8. Centrifuge the tubes at room temperature for 1 minute at 10,000 x *g*.
9. Transfer 600 µl of supernatant to a clean 2 ml tube.
10. Add 200 µl of Solution C3 and vortex for 5 seconds.
11. Incubate at 4°C for 5 minutes.
12. Centrifuge the tubes at room temperature for 1 minute at 10,000 x *g*.
13. Transfer 750 µl of supernatant into a clean 2 ml tube.
14. Add 1200 µl of Solution C4 to the supernatant and vortex for 5 seconds.
15. Load approximately 675 µl onto a Spin Filter and centrifuge at 10,000 x *g* for 1 minute at room temperature. Discard the flow through and add an additional 675 µl of supernatant to the Spin Filter and centrifuge at 10,000 x *g* for 1 minute at room temperature. Load the remaining supernatant onto the Spin Filter and centrifuge at 10,000 x *g* for 1 minute at room temperature.
16. Add 500 µl of Solution C5 and centrifuge at room temperature for 30 seconds at 10,000 x *g*. Discard the flow through.
17. Centrifuge again at room temperature for 1 minute at 10,000 x *g*.
18. Carefully place spin filter in a clean 2 ml tube.
19. Add 100 µl of Solution C6 to the centre of the white filter membrane.
20. Centrifuge at room temperature for 30 seconds at 10,000 x *g*.
21. Repeat steps 19 and 20.

DNA purification

DNA extraction of environmental samples often results in the co-extraction of inhibitory matrices. Environmental water samples are often inhibited by humic and fulmic acids associated with dissolved organic matter (Gentry-Shields et al. 2013), debris, metal ions, and polyphenol (Schrader et al. 2012); soil may contain humic and fulmic acids, which inhibit PCR even at low concentrations (Schrader et al. 2012); plants may contain pectin, polyphenols, xylan, and polysaccharides (Schrader et al. 2012); and stool may contain complex polysaccharides, bile salts, lipids, and urate (Schrader et al. 2012).

All of these co-extracted matrices are known as PCR inhibitors whose mechanistic action include co-precipitation, degradation, incomplete melting, sequestration, cross-linking, binding, or adsorption of DNA, reduction in the specificity of primers, degradation or inhibition of polymerases, chelation of metal ions or co-factors, or competition with the template DNA (see Schrader et al. 2012 for a review).

Because of the complexity of the co-extracted inhibitors associated with different types of samples, many different DNA extractions kits exist to cope with these co-extracts. For example, there are different Qiagen stool, soil, plant, and tissue DNA extractions kits. These kits all include the bind wash elute steps, but include different steps to cope with known inhibitors, the stool kit for example used InhibiteX tablets. These kits do not completely remove inhibitors, and with a lot of the NatureMetrics samples the final DNA extract is not colourless (indicative of some inhibitors).

Many methods exist to remove this inhibition. Among the most commonly used methods is simply to dilute away the inhibitors (e.g. as recommended in the great crested newt standard protocol - Biggs et al. 2014). The thinking here is that all components of the DNA extract will reduce in concentration, but that PCR of the template DNA will increase in efficiency without the inhibitors.

The addition of DNA enhancers such as BSA, DMSO, formamide, glycerol, detergents, polyethylene glycol, powdered milk, proteinase inhibitors, or gp32 has been shown to help with certain inhibitors, but these are not effective against all inhibitors and can affect the chemistry of the PCR reaction.

Lastly, DNA can be purified, which can be done with either kits or with a phenol-chloroform extraction. The trade-off between these methods is cost, time, and health and safety. Phenol-chloroform extraction is cheap, but it takes a long time and cannot be done without a fume hood. Kits (e.g. MO BIO's PowerClean Pro, or Zymo's One Step) are quick, can be done without a fume hood, but are more expensive. This additional cost is worth it though, McKee et al. (2015) show that the Zymo kit performs better than dilution, while Hu et al. (2015) show that PowerClean kits perform better than Chelex or phenol-chloroform re-extraction.

DNA purification is much more desirable than the other methods because complete removal of all inhibitors is much more desirable than dealing with inhibition, and this only has to be done once. Here we use MO BIO's PowerClean Pro DNA Clean-Up kit based on the recommendation of Hu et al. (2015).

DNA purification with PowerClean Pro kits: Protocol

1. Add up to 100 µl of DNA sample to a 2 ml collection tube.
2. Add 50 µl of Solution DC1 to the DNA. Vortex for 5 seconds and spin for 10 seconds.
3. Add 50 µl of Solution DC2 to the DNA mix. Vortex for 5 seconds and put it in the centrifuge.
4. Place the tubes in the centrifuge and orientate them the same way so that the pellet will always be on the same side of each tube. Centrifuge the tube at 13,000 x *g* for 2 minutes at room temperature.
5. Avoiding the pellet, transfer the entire supernatant to the second 2 ml collection tube. Close the cap of the first collection tube and throw it away.
6. Vortex Solution DC3 for 5 seconds. Add 400 µl to collection tube 2. Vortex for 5 seconds and spin for 10 seconds.
7. Load all of the DNA mix from collection tube 2 (~600 µl) to the labelled spin filter. Spin the filter at 10,000 x *g* for 30 seconds at room temperature. Carefully take the spin filter out of its collection tube and, without touching the tip of the tube to any surface, pour the flow through on the blue roll in the bin. Place the spin filter back into the same, newly-empty, collection tube.
8. Add 500 µl of Solution DC4 to the spin filter. Spin the filter at 10,000 x *g* for 30 seconds at room temperature. Carefully take the spin filter out of its collection tube and, without touching the tip of the tube to any surface, discard the flow. Place the spin filter back into the same, newly-empty, collection tube.
9. Repeat step 8.
10. Spin the spin filter at maximum speed for 2 minutes.
11. Place the spin filter into a new collection tube.
12. Add 50 µl of Solution DC5 to the centre of the spin filter membrane. Close the cap and tap the bottom of the collection tube onto the table to disperse DC5 throughout the membrane. Incubate the collection tube at room temperature for 1 minute and then spin the filter at 10,000 x *g* for 30 seconds at room temperature.
13. Repeat step 12.
14. Discard the Transfer the entirety of the eluate to the labelled low bind 1.5 ml microcentrifuge tube.

Polymerase Chain Reaction (PCR)

A specific and standardised region of the DNA is amplified from trace amounts found in the sample (eDNA or tissue) by a technique called polymerase chain reaction (PCR). PCR involves precise heating and cooling of DNA in the presence of buffered enzymes and genetic building blocks (dNTPs and primers), which results in the enzymatic replication of the region of DNA targeted by the chosen primers. The result of PCR is millions of copies of a specific DNA sequence (or amplicon), which can be sequenced and compared to reference database to identify the original source.

Different markers can be used to target different taxa, and these can be changed using different reagents in the laboratory (Table 11.1). For example, we target 12S rDNA with primers designed with fish mitochondrial genomes (Miya et al. 2015) to amplify fish DNA, while we use *rbcL* chloroplast primers to target plant DNA (de Vere et al. 2017). All PCRs were prepared in a dedicated UV-sterilised and bleached cabinet with dedicated pipettes and filter tips in a PCR-free building. Reaction composition for each project and marker, and the cycle conditions of the PCR (conducted in a different building) are shown in Table 11.1. All PCRs were carried out in triplicate in the presence of both a template negative control and a positive control (mock community with a known composition). PCR amplification success was determined by gel electrophoresis.

Table 11.1 PCR reaction compositions and cycle conditions for each marker accompanied with references. Primers were modified with overhanging Illumina adapter sequences in preparation for the subsequent nested PCR.

Project	Marker	Reaction composition	Cycle conditions	Reference
Violet click beetle Bredon Hill/Windsor Forest	COI - short 2	1X AmpliTaq Gold buffer, 0.5 µM of each primer, 1 mM of MgCl2, 0.8 nM of dNTPs, 0.0025 units/µl of AmpliTaq Gold, 2 µL of DNA, and up to 25 µL with H2O	94°C for 3 m, 35 x [94°C for 30 s, 46°C for 30 s, 72°C for 1 m], 72°C for 5 m	Arribas et al. 2016
Violet click beetle Bredon Hill/Windsor Forest	COI - Leray	1X AmpliTaq Gold buffer, 0.3 µM of each primer, 0.85 mM of MgCl2, 0.8 nM of dNTPs, 0.0025 units/µl of AmpliTaq Gold, 2 µL of DNA, and up to 25 µL with H2O	94°C for 3 m, 35 x [94°C for 30 s, 46°C for 30 s, 72°C for 1 m], 72°C for 5 m	Leray et al. 2013
Terrestrial invertebrates Lampert Mosses	COI - short 2	1X AmpliTaq Gold buffer, 0.5 µM of each primer, 1 mM of MgCl2, 0.8 nM of dNTPs, 0.0025 units/µl of AmpliTaq Gold, 2 µL of DNA, and up to 25 µL with H2O	94°C for 3 m, 35 x [94°C for 30 s, 46°C for 30 s, 72°C for 1 m], 72°C for 5 m	Arribas et al. 2016
Terrestrial invertebrates Lampert Mosses	COI - Leray	1X AmpliTaq Gold buffer, 0.3 µM of each primer, 0.85 mM of MgCl2, 0.8 nM of dNTPs, 0.0025 units/µl of AmpliTaq Gold, 2 µL of DNA, and up to 25 µL with H2O	94°C for 3 m, 35 x [94°C for 30 s, 46°C for 30 s, 72°C for 1 m], 72°C for 5 m	Leray et al. 2013
Freshwater fish Nene Washes	MiFish	1X Phusion Green Mastermix, 0.3 µM of each primer, 1 mM of MgCl2, 2 µL of DNA, and up to 25 µL with H2O	95°C for 3 m, 45 x [98°C for 20 s, 65°C for 15 s, 72°C for 15 s], 72°C for 5 m	Miya et al. 2015
Freshwater invertebrates Yardley Chase	COI - short 2	1X AmpliTaq Gold buffer, 0.5 µM of each primer, 1 mM of MgCl2, 0.8 nM of dNTPs, 0.0025 units/µl of AmpliTaq Gold, 2 µL of DNA, and up to 25 µL with H2O	94°C for 3 m, 35 x [94°C for 30 s, 46°C for 30 s, 72°C for 1 m], 72°C for 5 m	Arribas et al. 2016
Freshwater invertebrates Yardley Chase	COI - Leray	1X AmpliTaq Gold buffer, 0.3 µM of each primer, 0.85 mM of MgCl2, 0.8 nM of dNTPs, 0.0025 units/µl of AmpliTaq Gold, 2 µL of DNA, and up to 25 µL with H2O	94°C for 3 m, 35 x [94°C for 30 s, 46°C for 30 s, 72°C for 1 m], 72°C for 5 m	Leray et al. 2013
Freshwater	COI -	1X Phusion Green Mastermix, 0.5 µM of each primer, 1 mM of MgCl2 2 µL of DNA, and up to 25 µL with H2O	94°C for 3 m, 35 x [94°C for 30 s, 46°C for 30 s, 72°C	Elbrecht et

Project	Marker	Reaction composition	Cycle conditions	Reference
invertebrates Yardley Chase	BF2BR 1		for 1 m], 72°C for 5 m	al. 2017b
Lagoon invertebrates Bembridge/Saltons	COI - short 2	1X AmpliTaq Gold buffer, 0.5 µM of each primer, 1 mM of MgCl2, 0.8 nM of dNTPs, 0.0025 units/µl of AmpliTaq Gold, 2 µL of DNA, and up to 25 µL with H2O	94°C for 3 m, 35 x [94°C for 30 s, 46°C for 30 s, 72°C for 1 m], 72°C for 5 m	Arribas et al. 2016
Lagoon invertebrates Bembridge/Saltons	COI - Leray	1X AmpliTaq Gold buffer, 0.3 µM of each primer, 0.85 mM of MgCl2, 0.8 nM of dNTPs, 0.0025 units/µl of AmpliTaq Gold, 2 µL of DNA, and up to 25 µL with H2O	94°C for 3 m, 35 x [94°C for 30 s, 46°C for 30 s, 72°C for 1 m], 72°C for 5 m	Leray et al. 2013
Lagoon invertebrates Bembridge/Saltons	COI - BF2BR 1	1X Phusion Green Mastermix, 0.5 µM of each primer, 1 mM of MgCl2 2 µL of DNA, and up to 25 µL with H2O	94°C for 3 m, 35 x [94°C for 30 s, 46°C for 30 s, 72°C for 1 m], 72°C for 5 m	Elbrecht et al. 2017b
Lagoon invertebrates Bembridge/Saltons	COI - Full	1X Phusion Green Mastermix, 0.5 µM of each primer, 1 mM of MgCl2 2 µL of DNA, and up to 25 µL with H2O	94°C for 3 m, 35 x [94°C for 30 s, 46°C for 30 s, 72°C for 1 m], 72°C for 5 m	Folmer et al. 1994
Lagoon invertebrates Bembridge/Saltons	18S V4	1X Phusion Green Mastermix, 0.3 µM of each primer, 3% DMSO, 2 µL of DNA, and up to 25 µL with H2O	95°C for 5 m, 10 x [94°C for 30 s, 57°C for 45 s, 72°C for 1 m], 20 x [94°C for 30 s, 47°C for 45 s, 72°C for 1 m], 72°C for 10 m	Bradley et al. 2016
Lagoon invertebrates Bembridge/Saltons	18S V8	1X Phusion Green Mastermix, 0.3 µM of each primer, 2 µL of DNA, and up to 25 µL with H2O	95°C for 3 m, 30 x [98°C for 20 s, 65°C for 15 s, 72°C for 15 s], 72°C for 10 m	Bradley et al. 2016
Seahorse Poole Harbour	MiFish	1X Phusion Green Mastermix, 0.3 µM of each primer, 1 mM of MgCl2, 2 µL of DNA, and up to 25 µL with H2O	95°C for 3 m, 45 x [98°C for 20 s, 65°C for 15 s, 72°C for 15 s], 72°C for 5 m	Miya et al. 2015
Plants in soil Derbyshire Dales	rbcL	1X Phusion Green Mastermix, 0.2 µM of each primer, 2 µL of DNA, and up to 25 µL with H2O	95°C for 2 m, 35 x [95°C for 30 s, 50°C for 90 s, 72°C for 40 s], 72°C for 5 m	de Vere et al. 2017
Plants in soil Derbyshire Dales	ITS2	1X Phusion Green Mastermix, 0.33 µM of each primer, 2 µL of DNA, and up to 25 µL with H2O	94°C for 4 m, 37 x [94°C for 40 s, 49°C for 40 s, 72°C for 40 s], 72°C for 5 m	Sickel et al. 2015
Marine invertebrates	COI - Leray	1X AmpliTaq Gold buffer, 0.3 µM of each primer, 0.85 mM of MgCl2, 0.8 nM of dNTPs, 0.0025 units/µl of AmpliTaq Gold, 2 µL of DNA, and up to 25 µL with H2O	94°C for 3 m, 35 x [94°C for 30 s, 46°C for 30 s, 72°C for 1 m], 72°C for 5 m	Leray et al. 2013

Sequence library preparation

High throughput sequencing, unlike Sanger sequencing, requires that additional adaptor sequences are appended to the target amplicons so that they can physically attach to the high throughput sequencer, which has complementary adaptors to those appended to the amplicon - this process of appending appropriate adaptors is called library preparation.

Sequencing libraries were prepared according to the methodology developed by Illumina for amplicon sequencing on the Illumina MiSeq System (Illumina Inc. 2013), which is available in detail on their website (Illumina's "16S rRNA Sequencing Protocol"). Each sequencing library comprise of a single sample for a single marker. Briefly, this requires:

1. Purification of the amplicon to remove primer dimers and unincorporated reagents with AMPure XP beads according to the manufacturer's protocol. Quality of the purification was determined by gel electrophoresis.
2. Extending the purified amplicon with sequencing adaptors and indexes were added with a Nextera XT kit (Illumina) according to the manufacturer's protocol.
3. Purification of the extended amplicon to remove unincorporated reagents with AMPure XP beads according to the manufacturer's protocol. Quality of the purification was determined by gel electrophoresis.
4. Dilution to the correct concentration (4 nM) for the sequencer with H_2O according to concentrations obtained using a Qubit high sensitivity kit according to the manufacturer's protocol.
5. Pooling those sequencing libraries into a single final sequencing library.
6. Denaturation of the final sequencing library.
7. Sequencing on the MiSeq using a MiSeq 2 × 300 kit at a final concentration of 10-15 pM with a 5-10% PhiX spike.

Bioinformatics

The raw output of the high throughput sequencer is millions of sequences that need to be quality controlled and processed efficiently in more manageable sizes. NatureMetrics has developed a custom bioinformatic pipeline that effectively achieves this. Briefly, our pipeline merges raw R1 and R2 reads with usearch, trims the primer sequences with cutadapt, quality filtered and dereplicated with usearch, and clustered at 97% similarity with usearch to form OTUs (operational taxonomic units akin to species determinations). Each of these steps filter the number of superfluous sequences (replicated sequences) while maintaining the original links. A worked example with read numbers is shown in Figure A.1.

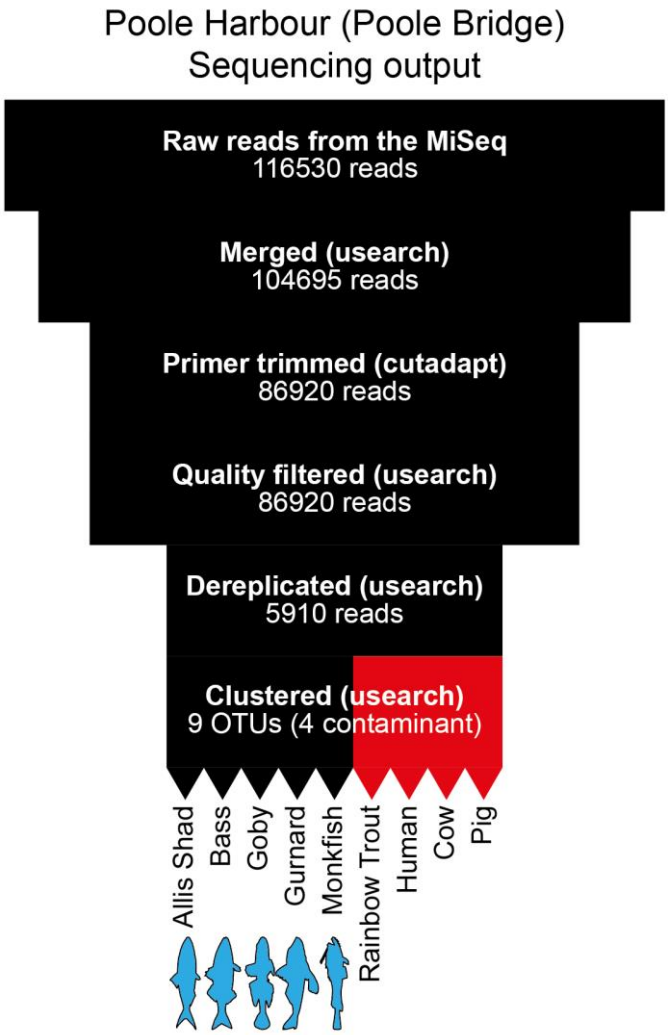


Figure A.1 Bioinformatic pipeline. The NatureMetrics metabarcoding pipeline takes the raw reads from a MiSeq and uses various programs to merge, trim, quality filter, dereplicate, cluster, and identify them. In this example a 12S rDNA single library from Poole Harbour targeting fish is processed using the pipeline. The number of reads goes from 116530 raw reads to 5910 high quality unique reads. These were clustered into 9 OTUs, of which 4 are known to be common contaminants associated with reagents used in the lab work and the human contamination along the way. The eDNA sample is found to contain five fish taxa.

After filtering and clustering, OTUs were identified using BOLD and/or GenBank. BOLD (Ratnasingham and Hebert 2007) is a curated database specifically for animal (COI), plant (*rbcL* and *matK*), and fungal (ITS) identification, but is limited to certain genes (COI, *rbcL*, *matK*, and ITS). GenBank is another DNA reference database that is open access, publicly available and is the most comprehensive database of its kind. GenBank contains much more information than BOLD, including other genes (for our purposes 12S rDNA, 18S rDNA resources are particularly useful), but also contains more uncurated and unidentified sequences.

By default the BOLD identification algorithm only provides species-level identifications if the OTU sequence matches the reference by at least 98%. The species-level identification is the top hit on the BOLD or GenBank database. If multiple reference sequences match equally to the query sequence then all of those references are given in the table. Any OTU sequences that have no species-level identification are given a genus-level identification where possible. Note that unidentified or misidentified OTUs can result from incomplete or incorrect reference databases, and missing OTUs can result from low quality DNA, environmental contaminants, or overrepresentation of certain species owing to imperfect size sorting.

The bioinformatics steps take the millions of raw sequences from the MiSeq and summarise these into a ‘species’ by sample table.

Marine invertebrate metabarcoding - Bangor University

Sample collection

Five samples were collected from the Wash SSSI in Norfolk, during December 2016, using standard protocols for marine invertebrate sampling. The samples were preserved in Longmire's buffer immediately after collection and posted to Bangor where they were stored at 4°C until DNA extraction (approximately 4 weeks).

Primer selection

For this work we used a 313 bp amplicon of the Cytochrome Subunit Oxidase I (COI) barcoding region targeting marine invertebrate diversity, using a primer pair published by Leray et al. (2013). These primers were designed based on DNA barcodes from the Moorea project using wide diversity of marine invertebrate species, while they are degenerate and suitable for targeting marine invertebrate sediment species.

DNA extraction and clean-up

Extraction of DNA directly from the collected sample, without pre-sorting, could be advantageous for quick processing of samples, while also enabling collection of information from the entire community without subsampling. Here, DNA extraction was performed directly from the collected sample, without pre-sorting. In order to facilitate the process, large empty shells (e.g. from mussels) were removed manually prior to homogenisation. For DNA extraction the MoBio PowerMax Soil DNA isolation kit was used. At the first instance the samples were homogenised using a commercial blender. The homogenate was added to the beads tubes and mixed briefly by vortexing. Lysis was performed by incubation at 56°C in a shaker at medium speed for 1 hour, with the addition of 0.4 mg/ml of Proteinase K. All reusable equipment used for extraction was decontaminated by immersion in 10% TriGene solution for 20 minutes, followed by thorough rinsing with distilled water. Subsequently, DNA clean-up was performed using MoBio Power Clean Pro DNA clean up kit. A volume of 100µl source DNA was used per reaction.

Library preparation - PCR protocols for MiSeq library preparation.

Library preparation was performed using a two-step PCR protocol according to Bista et al. (2017). Round 1 PCRs were performed using Illumina-tailed primers and Round 2 using Illumina indexes. PCRs were performed in 25 ml reaction volumes containing:

- Round 1: 12.5 µl Hot Start Taq 2X Master Mix (New England Biolabs), 8.9 µl PCR water, 0.8 ml (10 nmol/µl) of each forward and reverse primer and 2 µl DNA (10 ng/µl).
- Round 2: 12.5 µl Hot Start Taq 2X Master Mix (New England Biolabs), 5.9 µl PCR water, 0.8 µl of each forward and reverse primer and 5 µl Purified PCR product from Round 1.

The following thermocycling parameters were used:

- Round 1: Denaturation at 95 °C for 30 s, 23 cycles of: 95°C for 30 s, 49 °C for 40 s, 68 °C for 60 s, followed by a 5 min extension at 68 °C, hold at 4 °C.
- Round 2: Denaturation at 95 °C for 30 s, 15 cycles of: 95 °C for 30 s, 55 °C for 40 s, 68 °C for 60 s, followed by a 5 min extension at 68 °C, cool at 4 °C for 10 min.

Cleanup of PCR product between Rounds 1 and 2 was performed using an EXO-TSAP protocol (0.1 µl Exonuclease, 0.2µl TSAP, 0.7µl PCR water), incubation 37 °C for 30 min, 80 °C for 20 min and cooling at 20 °C for 30 sec. Multiplexing of the samples was performed using a dual index strategy using Illumina i5 and i7 indexes. See Table A.2 for the sequences of indexes used for each sample.

Table A.1 Primers used for library preparation. Round 1: universal tails and template specific sequence. Round 2: Illumina adapter, index and universal tail sequence.

Primer	**Round 1**	**Direction**
mlCOIintF	Forward Universal tail Template specific primer ACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNN GGWACWGGWTGAACWGTWTAYCCYCC	Forward
jgHCO2198	Forward Universal tail Template specific primer ACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNN GGWACWGGWTGAACWGTWTAYCCYCC	Reverse
Primer	**Round 2**	**Direction**
Forward	P5 Illumina adapter - Index 2 (i5) - Forward Universal tail AATGATACGGCGACCACCGAGATCTACAC - i5 Index - ACACTCTTTCCCTACACGACGCTC	Forward
Reverse	P7 Illumina adapter - Index 1 (i7) - Reverse Universal tail CAAGCAGAAGACGGCATACGAGAT - i7 Index - GTGACTGGAGTTCAGACGTGTGCTC	Reverse

Table A.2 Round 2 Illumina indexes.

Sample	**i7**	**Index sequence**	**i5**	**Index sequence**
1	**N701**	TAAGGCGA	**S505**	GTAAGGAG
2	**N702**	CGTACTAG	**S505**	GTAAGGAG
3	**N703**	AGGCAGAA	**S506**	ACTGCATA
4	**N704**	TCCTGAGC	**S506**	ACTGCATA
5	**N705**	GGACTCCT	**S506**	ACTGCATA

The samples were extracted and cleaned before library preparation. See Table A.3 for DNA concentration. Due to large volume of sample NE2 two columns were used for extraction (NE2a, NE2b) and the DNA was pooled for PCR. After optimization of the library preparation

protocol, samples 1, 3 and 5 performed well (strong bands, Figure A.2). Samples 2 and 4 underperformed and were repeated to increase the volume of available due to the low concentration of PCR product. The PCR products were sent to Nature Metrics for pooling as part of a more extended library, for run on Illumina MiSeq.

Table A.3 DNA concentrations per sample.

Sample	Extraction	Post-clean up
NE1	9.8	7.25
NE2a	12.33	3.4
NE2b	8.01	6.17
NE3	8.48	8
NE4	8.25	5.15
NE5	10.87	10.41

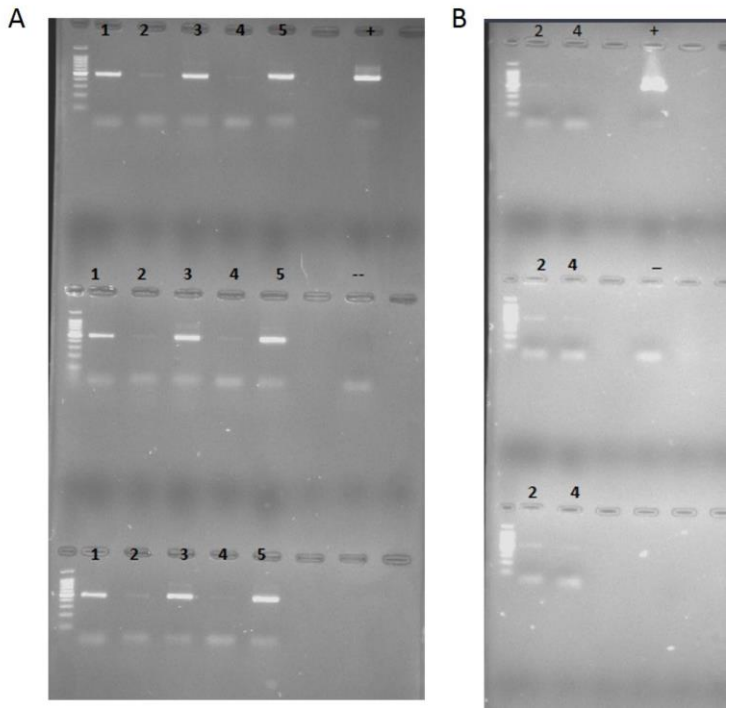


Figure A.2 Electrophoresis gel of prepared library samples after round 2, in triplicates. Each column represents one sample (1-5) and each line one replicate. (+) positive control, (-) negative control. A) all samples, in triplicates, B) second run of samples 2 and 4, in triplicates.

Annex 2. Glossary

Bioinformatics

Refers to a data processing pipeline that takes the raw sequence data from High Throughput Sequencing (often 20 million sequences or more) and transforms it into usable ecological data. Key steps for Metabarcoding pipelines include quality filtering, trimming, merging paired ends, removal of sequencing errors such as Chimeras, clustering of similar sequences into molecular taxonomic units (each of which approximately represents a species), and matching one sequence from each cluster against a reference database. The output is a species-by-sample table showing how many sequences from each sample were identified as each species.

Community DNA

Refers to DNA extracted from a mixture of different organisms. Could be eDNA (environmental samples almost always contain DNA from a mixture of species) or Organismal DNA (e.g. homogenised insect trap samples).

DNA barcodes (genes/markers)

Refers to genes that can be used for species identifications. Different regions of DNA mutate at different speeds. Fast-changing regions are useful for population studies and paternity testing, while the most stable regions can be used for assessing deep evolutionary relationships between groups of organisms. Certain regions change at just the right rate to be stable within a species but different between species. These are known as barcode genes. The official barcode gene for animals is Cytochrome Oxidase 1 (COI or *cox-1*). Other genes used as animal barcodes include 12S, 16S, 18S and Cytochrome-b (*cytb*). For plants, the most commonly used genes are *matK*, *rbcL*, *trnL* and ITS.

DNA barcoding

Refers to the process of linking a voucher specimen to a voucher sequence (DNA barcode) to either populate a reference database or to use the reference database to identify the specimen. Organismal DNA is extracted, amplified by PCR, sequenced using Sanger sequencing, and linked by to reference databases to identify the original specimen.

DNA metabarcoding

Refers to identification of species assemblages from Community DNA using Barcode Genes. DNA metabarcoding is a high throughput version of DNA barcoding. PCR is carried out with non-specific Primers, followed by High Throughput Sequencing and Bioinformatics processing. Can identify hundreds of species in each sample, and 100+ different samples can be processed in parallel to reduce sequencing cost. DNA metabarcoding has been used to rapidly and accurately identify whole communities, which is directly amenable for trap samples (Ji et al. 2013) or kick samples (Elbrecht et al. 2017b). The common theme in all comparative DNA metabarcoding vs. traditional sampling studies is that the DNA methods are quicker, higher resolution, more speciose, while also providing equivalent information

with equivalent environmental management decisions.

Environmental DNA (eDNA)

Short for 'environmental DNA'. Refers to DNA deposited in the environment through excretion, shedding, mucous secretions, saliva, etc. eDNA can be found in soil, water, air, etc. and is characterised by a complex mixture of DNA originating from multiple biological sources and from intact cells and free DNA (Taberlet et al. 2012). This can be collected in environmental samples and used to identify the organisms that it originated from in a non-invasive way rather than rely on catching or even observing the target. eDNA in water is broken down by environmental processes over a period of days to weeks. It can travel some distance from the point at which it was released from the organism, particularly in running water. eDNA in soil can bind to organic particles and persist for a very long time (sometimes hundreds or thousands of years). eDNA is sampled in low concentrations and can be degraded (i.e. broken into short fragments), which limits the analysis options.

Environmental DNA (eDNA) metabarcoding

eDNA metabarcoding is the combination of eDNA as a source and DNA metabarcoding as a technique. Instead of tissue samples that could feasibly be sorted and identified using the traditional taxonomic methods, eDNA sources are difficult or impossible to analyse with traditional means. This is a burgeoning and young field, and has been applied to assess aquatic samples (Valentini et al. 2016; Hänfling et al. 2016), gut contents (Kartzinel et al. 2015), faeces (Hope et al. 2014), honey (de Vere et al. 2017), leeches (Schnell et al. 2012; 2015), etc.

High throughput sequencing technology

Technology developed in the 2000s that produces millions of sequences in parallel. Enables thousands of different organisms from a mixture of species to be sequenced at once, so Community DNA can be sequenced. Various different technologies exist to do this, but the most commonly used platform is Illumina's MiSeq. Also known as Next-Generation Sequencing (NGS) or parallel sequencing.

Organismal DNA

Refers to DNA sampled directly from the organism through whole organism collection (e.g. invertebrates), swabbing, blood sampling, clipping, etc. Usually high concentration and non-degraded. The location of the organism at the time of sampling is definitively known. Overall there are fewer uncertainties than for eDNA.

OTU (Operational Taxonomic Unit)

Refers to a unit of diversity defined bioinformatically based on a similarity threshold. These units are clustered at a percentage similarity that is akin to species units (e.g. COI sequences are clustered into a single OTU when they are similar for 97% of the sequence).

PCR (Polymerase Chain Reaction)

A process by which millions of copies of a particular DNA segment are produced through a series of heating and cooling steps. Known as an 'amplification' process. One of the most

common processes in molecular biology and a precursor to most sequencing-based analyses.

Primers

Short sections of synthesised DNA that bind to either end of the DNA segment to be amplified by PCR. Can be designed to be totally specific to a particular species (so that only that species' DNA will be amplified from a community DNA sample), or to be very general so that a wide range of species' DNA will be amplified. Good design of primers is one of the critical factors in DNA based monitoring.

Reference databases

Refers to libraries of DNA sequences (usually from DNA barcode genes) that have been generated from species of known identity. Sequences from unidentified organisms – obtained either by Sanger Sequencing or High Throughput Sequencing – are compared against a Reference Database to make species identifications. Databases can be curated (e.g. the Barcode of Life Database – BOLD – www.boldsystems.org) or uncurated (e.g. Genbank – www.ncbi.nlm.nih.gov). In curated databases, identifications are scrutinised and verified; in uncurated databases they are not. GenBank is therefore far more extensive than BOLD, but contains many errors.

Sanger sequencing

Traditional DNA sequencing. Each reaction produces a single sequence so it only works on amplified DNA of a single species. A sequence is a series of nucleotide bases represented by the letters A, T, C & G. Here is the sequence of part of the 12S gene for a minnow (*Phoxinus phoxinus*):
CACCGCGGTTAAACGAGAGGCCCTAGTTAATAATTGACGGCGTAAAGGGTGGTTAGGG
GGTGTAATGTAATAAAGCCGAATGGCCCTTTGGCTGTCATACGCTTCTAGGTGTCCGAA
GCCCAACATACGAAAGTAGCTTTAAGAAAGTCCACCTGACGCCACGAAAACTGAGAAA

Sequencing library

Refers to the state of the sequencing-ready DNA. DNA needs to have specific adaptors (short sections of synthesised DNA) attached so that it can physically bind to the high throughput sequencer. The process of making DNA sequencer ready is called library preparation.