JNCC Report
No. 669b

End-user Frequently Asked Questions on DNA-based methods for environmental monitoring

Jones, E.P., Adams, I.P., Walshaw, J., Benucci, M., MacArthur, R., Boonham, N. & Bryce, S.

December 2020

ISSN 0963-8091

For further information please contact:

Joint Nature Conservation Committee
Monkstone House
City Road
Peterborough PE1 1JY

www.jncc.defra.gov.uk

This report should be cited as:

Jones, E.P., Adams, I.P., Walshaw, J., Benucci, M., MacArthur, R., Boonham, N. & Bryce S. (2020) End-user Frequently Asked Questions on DNA-based methods for environmental monitoring. Fera Science report to JNCC, JNCC Report No. 669b. JNCC, Peterborough, ISSN 0963-8091.

JNCC EQA Statement:

This report is compliant with JNCC's Evidence Quality Assurance Policy

https://jncc.gov.uk/about-jncc/corporate-information/evidence-quality-assurance/

Acknowledgements:

This project was led and funded by the Defra DNA Centre of Excellence in 2019/20, which aims to transform Defra group science to support policy and delivery by sharing and demonstrating the wider potential of existing methods; enabling the rapid development and trialling of new DNA based method applications; providing a focal point for leadership in Defra, UK and across other government departments and building shared capability including skills and facilities.

Summary

- There is increasing interest in using DNA-based methods for the environmental monitoring carried out by government organisations. However, end-user understanding of these techniques is often limited, in part due to the rapid evolution and proliferation of approaches.

- This document aims to provide introductory answers to questions that non-technical staff may have on DNA-based methods.

- These Frequently Asked Questions were compiled from end users involved in the Defra DNA Centre of Excellence (CoE) and focus on providing an introduction to when and why the choice of approach can influence the results from monitoring that uses DNA-based methods.

- Readers of these FAQs should be aware that detailed guidance on technical aspects of DNA methods is beyond the scope of the document.

- The answers to questions are necessarily brief and higher level and should not be considered a comprehensive synthesis of all available literature and knowledge.

- The intended audience is end users of the results from DNA-based monitoring methods, who have some knowledge of the techniques but are not specialists.

- Additional information can be found in the complementary JNCC Report No. 669a ‘Guidance for end users on DNA methods development and project assessment’ (Jones *et al.* 2020).

Contents

Background 1
Q1 How do we ensure DNA degradation and cross-contamination is properly controlled during sampling collection and transport? 2
Q2 How should samples be stored to facilitate re-use? 3
Q3 How long can traps be left in the field before DNA degradation becomes a problem? Is this affected by weather conditions (e.g. heavy rain)? 4
Q4 What effect do different approaches to decontamination (e.g. bleach, virkon) of kit from field to laboratory have on results? 5
Q5 What effect do different DNA preservatives have on results? 7
Q6 What effect does filtration vs. ethanol precipitation have on results? 9
Q7 Does temperature during transportation of preserved and unpreserved samples affect results? 10
Q8 What effect does sample volume have on results? 11
Q9 What effect do large differences in organism size have on results? 12
Q10 What are the effects of large amounts of non-target DNA in a sample? 13
Q11 What is the effect of DNA extraction kit choice on results? 14
Q12 What is the effect of the choice of primers on results? 15
Q13 How much effect does inhibition have on results, and how can this be understood and minimised? 17
Q14 What effect does the number of PCR replicates and the way in which they are treated have on results and costs? 18
Q15 What effect does one vs. two-step PCR have on results? 19
Q16 What effect do differences in sequencing depth have on results? 20
Q17 How does the sensitivity of different PCR methods compare with each other? 21
Q18 What are the approaches to removing potentially erroneous sequences from datasets, and do these produce different results? 22
Q19 What are the effects of using different approaches to assigning taxonomy (e.g. algorithms, OTU clustering thresholds)? 24
Q20 What are the effects of using different sequencing platforms? 26
Q21 To what extent does the quality of reference databases affect the reliability of results? 27
Q22 How does metabarcoding compare with targeted assays for single-species detection? 28
Q23 Which aspects of sample processing and data analysis are influenced by the sampling media (e.g. water, soil, bulk specimens, *etc*.) and which can follow similar approaches irrespective? 30
Acknowledgements **31**
References **32**
Appendix 1 ..... **37**

Background

There is increasing interest in using DNA-based methods to support the environmental monitoring carried out by government organisations. However, end-user understanding of these techniques is often limited, in part due to the rapid and ongoing proliferation of approaches. This document aims to provide concise answers to the most frequently encountered questions non-technical staff may have when evaluating or comparing DNA-based studies. It focuses particularly on providing an introduction to when and why the choice of method can influence results. The questions were compiled by end users involved in the Defra DNA Centre of Excellence (CoE). A number of additional questions that were provided by the CoE but could not be included in this FAQ are given at the end of the document (Appendix 1). It may be possible to address these in a future project.

Users should be aware that it is beyond the scope of the current project to provide guidance on all technical aspects relating to DNA method use. Answers to questions are also necessarily brief and higher level; these should not be considered as a comprehensive overview of all the available literature and knowledge. Supporting references are therefore provided as an introduction to the literature rather than the result of a systematic search and synthesis. The intended audience is end users of the results from DNA-based monitoring methods with some knowledge of the techniques but who are not specialists.

Q1 How do we ensure DNA degradation and cross-contamination is properly controlled during sampling collection and transport?

DNA degradation is mostly caused by enzymes (nucleases) or chemical compounds (e.g. phenolics), which can degrade free DNA. It can also be caused by microorganisms in the sample that consume cellular material. Different sample types will differ considerably in propensity to suffer degradation. Degradation will tend to be more prevalent if samples are stored at elevated temperatures for longer periods, which could increase and prolong the duration of the enzyme/chemical action. Enzymes come from living organisms in the sample; increased degradation in some samples could be from rotting or degrading plant or animal material. Sample collection using a preservative (e.g. ethanol) will reduce or retard the likelihood of degradation although care must be exercised to ensure the sample and preservative are well mixed, and the ratio of preservative to sample is high. However, the preservative may not reduce the chemical action.

Particular consideration should be given to the logistical elements of studies where samples are transported without preservatives. A maximum temperature exposure threshold could be determined empirically or with reference to the literature, and the temperature during transit monitored using temperature loggers (electronic, or disposable tags such as Telatemp or WarmMark) and any samples that exceed temperature exposure limits discarded. *(Also see Question 7)*

It is very difficult to monitor DNA degradation within a sample, therefore a spiked control can be used as a proxy. Usually this is a DNA sequence (either of target or a synthetic DNA sequence) added to all samples or contained within the sampling kits at the point of sampling. Note that a control of DNA spiked into the sample will not be captured during filtration as it will pass through the filter but will be captured by ethanol precipitation or similar. Once the sample is received and DNA has been extracted in the lab, the sample can be tested for this spiked control DNA and any samples for which the level of control falls below a quality threshold can be excluded from analysis.

Cross-contamination is controlled by good process (clear protocols to decontaminate equipment, sample handing) and by monitoring (appropriate negative controls including field blanks – although note that these are problematic for metabarcoding samples). Where possible, using single use disposable items for sample collection will minimise cross-contamination. As a minimum, this should include wearing disposable gloves and changing to a fresh pair between samples.

The consequence of failing to control cross-contamination can include: an increase in the false positive rate (i.e. results which find the target is present when it is absent), distortion of the relative abundance ratios in metabarcoding results, and an increase in the failure rate (where the contamination is detected by the field negative controls and so samples have to be discarded). The importance of controlling cross-contamination depend on how serious the impact of a false positive result will be (see sections 3.1.1 to 3.1.3 in the accompanying JNCC Report No. 669a “Guidance for end users on DNA methods development and project assessment”; for further discussion; Jones *et al.* 2020).

Related Questions:

Q4: What effect do different approaches to decontamination (e.g. bleach, virkon) of kit from field to laboratory have on results?

Q7: Does temperature during transportation of preserved and unpreserved samples affect results?

Back to Contents

Q2 How should samples be stored to facilitate re-use?

Note that a separate DNAquaNet Working group is addressing sample and data storage, and the outputs from this may be relevant (https://dnaqua.net/about/).

Documentation
All samples should be appropriately labelled in a standardised format with a unique identifier, and sample records kept in a defined and easily accessible location. Sample records should contain: sample metadata such as the date and location of sampling, the physical location of the sample, a named sample owner, and a note on how long the sample should be stored for.

Storage
The optimum long-term (years) storage of samples depends on the type of sample.

Freezing is often used, but it is important to note that for all sample types, freeze/ thawing will rapidly degrade the integrity of the DNA. If frozen samples are likely to be reused on multiple occasions, it is advisable to store these as separate aliquots that can be used without unfreezing the majority of the sample. Where samples are kept in a freezer or fridge, these should have maintenance protocols and quality processes attached to them to mitigate against catastrophic failures. This is a genuine risk to collections.

It is generally best to store samples as extracted DNA or RNA. There is little literature on how best to store this DNA, but dried or in liquid nitrogen appears to be best for the long-term because it keeps DNA in a vitreous state (Anonymous 2008). For more routine storage, -20°C or -80°C in Tris/EDTA buffer is recommended over storage in water and 4°C (Kim *et al.* 2012), with a general opinion that colder is better. -80°C is often recommended.

For plant and animal tissues, storage at -80°C or in high levels of ethanol is widely accepted as appropriate for long term preservation of DNA (see references in Moreau *et al.* 2013). There can be cost limitations to ultra-low temperature freezers and ethanol preservation, in which case suitable alternatives for plants are freeze drying and storage in a cool, light free location. Note that ethanol preserved tissues require long term maintenance to ensure the ethanol levels don't drop too low.

eDNA samples (e.g. the filters from filtered water samples) should be stored as extracted DNA. Although adding preservative can maintain the stability of the sample over shorter periods (weeks, e.g. Spens *et al.* 2017), we are not aware of studies investigating the preservation of unextracted eDNA for long-term storage. It would therefore appear safer to store these samples as extracted DNA.

The best form of preservation for mixed community samples (soil, invertebrate trap samples, *etc.*) is not clear, and we could not find studies looking at this over the long term. Given that different members of the community may degrade at different rates it may be best to store these as extracted DNA, but preservation of specimens in high concentrations of ethanol (95% or higher), propylene glycol, RNAlater or similar preservatives may also be appropriate (Also see Question 5).

Related Questions
Q5: What effect do different DNA preservatives have on results?

Back to Contents

Q3 How long can traps be left in the field before DNA degradation becomes a problem? Is this affected by weather conditions (e.g. heavy rain)?

The rate of DNA degradation depends on the sample type, the temperature traps are exposed to, and the preservative used within the traps. Some samples will degrade very quickly (soft-tissued or very small organisms, free-DNA) without an efficient preservative. Raised temperatures (above 20^{o}C would be considered high) will cause more rapid degradation and will necessitate an efficient preservative. Anything that causes the preservative to be diluted (rainfall entering the trap, fluids from large-bodied organisms accumulating in the trap) or that causes the preservative to evaporate (temperatures >20^{o}C for ethanol), will have a negative effect on preservation. It is also important to have a high ratio of preservative to sample in the traps.

Because of this range of potential influences, it is not possible to definitively answer the question. For example, with excess volumes of an adequate preservative and in ideal conditions, it may be possible to leave a trap out for a considerable period (months), but under other circumstances more than a few days may be a problem (e.g. where using a preservative is not possible, or the preservative rapidly becomes excessively diluted). For specific projects, it may be necessary to determine the optimal sampling conditions experimentally, e.g. by leaving some traps for a range of time periods under the expected environmental conditions and determining if the degree of sample preservation is sufficient for the desired application.

Possible in-trap preservation strategies include allowing the samples to air dry, or using preservatives such as ethanol or buffers such as Longmire's buffer or polyethylene glycol (PEG) buffer.

Indications that samples have degraded are heavy discolouration of the preservative colour and the samples smelling rotten. If the samples smell strongly of decay, it is unlikely that any but the most sensitive and robust assays that target very short DNA fragments will work.

Related Questions
Q5: What effect do different DNA preservatives have on results?

Back to Contents

Q4 What effect do different approaches to decontamination (e.g. bleach, virkon) of kit from field to laboratory have on results?

The purpose of decontamination is to reduce cross-contamination between samples and from the environment to the samples. Inadequately controlling contamination raises the false positive rate (if cross-contamination goes undetected) or increases the failure rate (if contamination is detected by negative controls, and results are discarded as a consequence). The field of ancient DNA (which also deals with trace amounts of DNA collected from the environment) abounds with early published studies entirely based on contamination rather than real results – e.g. dinosaur DNA that was actually human and chicken DNA; Hedges *et al.* (1995). A secondary purpose of decontamination is to reduce the possibility of spreading pests and diseases between sites, and to reduce the risks to staff from zoonotic diseases.

We were unable to find literature comparing the effectiveness of the different approaches to decontamination for DNA sampling in the field. The control of DNA contamination in the laboratory setting is relatively well studied (see below).

Decontaminating sampling equipment with Virkon (a bactericidal, viricidal and fungicidal agent) is recommended in between different sites to avoid transferring diseases (Tidbury *et al.* 2018), but it does not seem to be an effective decontaminating agent for DNA (Ballantne *et al.* 2015). Alternatives are bleach (sodium hypochlorite) and proprietary products such as DNA-AWAY/Distel. Few studies compare the efficacy of these products directly, although DNA-AWAY appears less effective than bleach (Champlot *et al.* 2010). Appropriately used, bleach is a highly effective low-cost option (Champlot *et al.* 2010), although it can degrade relatively rapidly over time (Camps *et al.* 2009), and diluted bleach solutions should be made up not longer than weekly from tablet or concentrated liquids. Within the laboratory, another commonly used alternative is UV irradiation, which is highly effective where it can be used (Champlot *et al.* 2010).

There are health and safety and environmental risks associated with decontamination products, which are usually stated on the packaging or found on the Material Safety Data Sheet (MSDS) for each chemical. These should be seen as the over-riding source of information for the hazards associated with the chemicals. A short discussion of some relevant points is given below but is not comprehensive.

The use of bleach in the lab is often restricted because it gives off toxic gases if inadvertently mixed with other common laboratory chemicals. Bleach is also a corrosive with risks to users and to equipment, requiring suitable PPE to be worn (e.g. gloves, goggles). Similar care should be taken to minimise the impact on the environment when transporting bleach to and from sites, during its use and while it is stored (bleach resistant plastic containers are necessary), and when it is disposed of. Virkon in powdered form is a high risk from inhalation and poses a risk to the environment.

Should any of the decontamination products enter the samples, the impacts can be very high, ultimately leading to the degradation of all the DNA within a sample. However, we do not know of any studies where this has been mentioned as a problem. Care should always be taken that no sample tubes are open while decontamination is taking place, and that any decontamination products have been fully rinsed of materials that will be in contact with samples (e.g. tubing used to filter eDNA samples).

Note that methodological steps can also be introduced to reduce contamination, reducing the need for decontamination. For example:

- In the field; sampling against the concentration of the target (if known). This involves sampling areas of low/no target presence before sampling sites where the target is known to be present.
- In the laboratory; physically separating areas where samples, DNA and reagents are stored; physically separating areas where DNA extractions, PCR reactions and post-PCR work is done. Removing steps where PCR products are opened (e.g. using real-time PCR, ddPCR over conventional PCR and gel electrophoresis) avoids the risk of PCR amplicon contamination.

Back to Contents

Q5 What effect do different DNA preservatives have on results?

NB: This covers eDNA and specimen DNA

Poor preservation decreases the sensitivity of the test (i.e. failing to detect the target DNA where it was originally present in the sample), leading to higher false negative rates. For targeted assays the samples contributing lower amounts of DNA will be most affected, and for metabarcoding studies the organisms with lower amounts of DNA will be most affected. A degradation control can be included at the sampling step (as in the great crested newt eDNA test approved by Natural England) to monitor for degradation.

Using a preservative is important, but the variation in the end result caused by using different types of preservative may be minor. The best method can only be determined empirically for each application, but an appropriate method can be selected from the literature and tested in some trial samples. For validated tests (in the strict sense of the term 'validation'), the preservation approach is an intrinsic part of the method and a change in preservative should trigger an appropriate re-validation. For example, it would not be appropriate to change the preservative used in the routine eDNA test for great crested newts without demonstrating empirically that the change has no effect.

Preservatives for water filtration samples

For eDNA methods that use in-field filtration as the DNA capture step, filters can be preserved in several ways (e.g. refrigeration, or preservation with Longmire's buffer, RNAlater or ethanol). A number of studies have considered these approaches and reached slightly different conclusions (e.g. Hinlo *et al.* 2017; Spens *et al.* 2017). For example, Spens *et al.* (2017) found that preserving filters in ethanol or Longmire's gave better preservation than RNAlater or freezing, while Hinlo *et al.* (2017) found that ethanol was worse than freezing for some extraction types, but equivalent for other types. Not using any preservative method had a strong detrimental effect on preservation. It is likely that the different buffers have different advantages and disadvantages depending on the sample matrix and target species and the DNA extraction kit they are coupled to, all of which will affect the end result.

Preservatives for aggregated invertebrates

In a review, Liu *et al.* (2019) identified that the most commonly used preservatives for aggregated invertebrates were (in decreasing order): ethanol, Longmire's buffer, DESS and propylene glycol. Vink *et al.* (2005) found that propylene glycol and RNAlater were better preservatives than 95% ethanol, and all were better than 75% ethanol. However, RNAlater may be too expensive for many applications (of the order of £375 per 500ml). Pragmatically, the difference between 95% ethanol and propylene glycol may not have a great effect on the results unless the focus is on sensitivity and limit of detection (e.g. the detection of low abundance DNA is of prime importance). As with eDNA filters, the best method can be determined empirically for each application, but an appropriate method can also be selected from the literature and tested on trial samples. A change in the preservation method should trigger an appropriate re-validation of the modified method.

Preservatives for soil samples

Tzeneva *et al.* (2009) showed that it is possible to obtain ecological information from microbial DNA in soil samples that were thoroughly air-dried at 42°C for 48 hours without adding any preservative to the samples. Andersen *et al.* (2012) extracted vertebrate DNA from soil samples collected in DNA-free 50 mL tubes that were frozen at -20°C; again, not adding any specific preservative but lowering the temperature to reduce degradation (See Question 7 for further information on the effects of temperature on DNA). Tatangelo *et al.* (2014) compared different preserving methods for soil and water samples and demonstrated that freezing or no preservative are both valid for soil whereas water samples need to be

preserved either through refrigeration or the addition of appropriate preservative (e.g. RNAlater, ethanol).

Related Questions:

Question 2: How should samples be stored to facilitate re-use?
Question 3: How long can traps be left in the field before DNA degradation becomes a problem?
Question 7: Does temperature during transportation of preserved and unpreserved samples affect results?

Back to Contents

Q6 What effect does filtration vs. ethanol precipitation have on results?

This question relates to the DNA capture step of eDNA studies on water. In filtration, the water sample is passed through a membrane with a specific pore size. Important variables to consider are: membrane type (e.g. nitrocellulose), pore size (in microns) and volume of water filtered. Ethanol precipitation in this context refers to a series of methodological steps in which the water sample is added to a tube containing a preservative (ethanol) and a chemical that causes the DNA to precipitate out of the solution (sodium acetate), after which the sample is centrifuged at high speed to pellet the DNA at the bottom of the tube(s). This centrifugation step severely limits the amount of water that can be processed.

Several papers compare these two methods (e.g. Spens *et al.* 2017; Hinlo *et al.* 2017; Deiner *et al.* 2015). The two methods principally affect results in terms of sensitivity, contamination (i.e. likelihood of a false positive) and community composition (for metabarcoding studies or shotgun sequencing studies). Differences in operational feasibility are also a highly important consideration.

In relation to method sensitivity, filtration appears to consistently outperform ethanol precipitation for DNA yield and target detection (e.g. Deiner *et al.* 2015; Adrian-Kalchhauser & Burkhardt-Holm 2017; Spens *et al.* 2017). This can partially be attributed to filtration using a higher volume of starting sample than ethanol precipitation (filtration volumes are typically larger than 250ml whereas ethanol precipitation volumes are not greater than 90ml). Variations in filtration methodologies (e.g. pore size, material type) also affect sensitivity (e.g. Hinlo *et al.* 2017: Spens *et al.* 2017). The difference in detectability between filtration and precipitation will affect the reported community composition from metabarcoding studies, both in terms of presence / absence and any estimate of relative abundance.

However, sensitivity is not the only consideration when selecting which DNA capture method to use. Contamination is likely to be controlled better by ethanol precipitation or by filtration using enclosed filters rather than by filtration using open filter receivers, because in the first two approaches the water is not open to the environment during DNA capture. Additional limitations of filtration are that **(i)** filters can become clogged if the sample processed contains high levels of sediment (meaning the agreed volume of water cannot always be filtered) and **(ii)** water either has to be filtered on site (sometimes requiring specialist equipment, and limiting the number of samples that can be processed by hand) or rapidly transported to the laboratory (with associated logistical challenges). Samples for ethanol precipitation are relatively straightforward for non-specialists to collect and to transport to the laboratory (provided correct procedures for posting hazardous goods are followed), and sensitivity can be increased up to a point by increasing the number of replicates used.

If detecting low-level presence of the target is the prime objective, then filtration is likely to be the better method, provided that the volume of water filtered is high (500ml or more, rather than 250ml or less) and operational constraints can be met. However, for practical reasons, ethanol precipitation is sometimes the most appropriate method to use. These operational trade-offs may change as methods become more mature.

An important consequence of the differences in capture methods is that results from studies using different DNA capture methodologies should be compared with a great deal of caution, unless there is suitable intercalibration / validation data, and/ or some statistical modelling.

Related Questions
Q8: What effect does sample volume have on results?

Back to Contents

Q7 Does temperature during transportation of preserved and unpreserved samples affect results?

The extent to which the temperature at which samples are transported is likely to affect results depends on the sample type, the duration and exposure to different temperatures, and whether samples have been preserved in any additional way. In general, high temperatures and high humidity degrade DNA, and the longer a sample is exposed to a higher temperature the greater the degradation. DNA degradation increases the false negative rate, as the amount of target DNA in the sample degrades to below the limit of detection. In surveys targeted to an individual species, this will result in a false negative, in metabarcoding it will tend to result in decreased detection of lower abundance organisms within the reported community.

Ideally, all samples should be kept cool during transport (temperatures will vary by sample, but 5°C or lower is better) and should spend as little time as possible in transit (ideally within 12-24 hours). Adding a preservative will mitigate the effects of higher temperatures, although it is still advisable to keep samples cool as a precaution against degradation.

eDNA in water degrades more rapidly at higher temperatures (Strickler *et al.* 2015) but also degrades at 5°C (Eichmiller *et al.* 2016; reported degradation rate is over days). For ethanol precipitation samples, the ethanol is sufficient to preserve the DNA during normal UK temperature ranges for a short period of time (days). Samples for filtration should be filtered on site or kept cool and transported to the laboratory within 24 hours. It is not clear what preservative could be added to an unfiltered water sample in sufficient volume, without affecting the composition of the DNA particles captured on the filter. If it is not possible to keep water samples cool and transport them to a laboratory quickly, then the better alternative is to process the sample on site (e.g. by filtration) and store the captured DNA (e.g. the filters) in a preservative.

For aggregations of organisms, the organisms will begin to degrade with high temperatures, particularly if they remain wet. This process is significantly delayed by adding the correct preservative, which can allow samples to be stored for years. For bacterial communities, high temperatures might cause not only DNA degradation but also a change in the bacterial community itself, thus skewing the results and their biological interpretation. (Pietikäinen *et al.* 2005). This is also likely to be true of fungal communities. Again, adding preservatives can delay this process and is strongly recommended.

To control for the effect of high temperatures, a DNA degradation control should be introduced into all samples at the point of sampling (or contained within the sampling kit), and any samples for which the degradation control has failed should be excluded from analysis. This is particularly relevant for high importance results – e.g. the standard great crested newt eDNA kit contains a degradation control despite abundant preservative also being part of the kit. Alternatively, a maximum temperature exposure threshold could be determined empirically or with reference to the literature, and the temperature during transit monitored using temperature loggers (electronic, or disposable tags such as telatemp WarmMark) with any samples that exceed this threshold discarded.

Related Questions

Q1: How do we ensure DNA degradation and cross-contamination is properly controlled during sampling collection and transport?

Back to Contents

Q8 What effect does sample volume have on results?

In principle, increasing the sample volume should increase the amount of DNA that enters the capture and/or DNA extraction step, and up to a point will increase the amount and concentration of DNA post extraction. All else being equal, this has two potential benefits:

- Increased sensitivity of the test (increases the probability of detection for rare species)
- Decreases the variability of the test (reduces community variability as less abundant taxa are represented in the sample more reliably)

Empirical evidence for this is somewhat mixed for eDNA studies. Increasing the water volume sampled increased sensitivity for fish detection (e.g. Wilcox *et al.* 2018), although no increase was observed for invertebrate detection from aquatic eDNA samples (Mächler *et al*. 2016).

Several factors influence the upper limit on the sample volume including:

- For eDNA capture from water (and possibly soils, depending on the method): the DNA capture step (filtration, ethanol precipitation and centrifugation) will have an upper limit on the volume of sample that can be processed, dictated (e.g.) by clogging of the filters and centrifuge capacity.
- Overloading the DNA extraction step with DNA. This decreases the amount of DNA recovered (for example, overloading a DNA clean-up spin column such as the Qiagen Blood and Tissue kit column will decrease the DNA recovered compared to loading with an appropriate amount of DNA).
- Increasing the volume of sample that contains inhibitors (e.g. environmental water samples, tissues with inhibitors etc) may also increase the concentration and amount of inhibitors in the DNA post extraction above a threshold that can be dealt with (also see Question 13).
- The cost of extraction reagents may become prohibitive. For example, a practical limit on the volume of soil that can be extracted in a single extraction is imposed by the volume of buffer needed to soak the soil for the first step. This may also apply to aggregations of invertebrates.

An appropriate sample volume should be determined by consulting the wider literature and/or experimental evidence, taking into consideration:

- The purpose of the test (e.g. for rare species detection larger sample volumes may be desirable).
- Operational or cost constraints on the number or volume of sample that can be processed.

Related Questions

Q6: What effect does filtration vs. ethanol precipitation have on results?
Q13: How much effect does inhibition have on results, and how can this be understood and minimised?

Back to Contents

Q9 What effect do large differences in organism size have on results?

The effect for a targeted-detection based project for a small body-sized species in the presence of many large-sized species will depend on the sampling method. Where sampling and extraction are based on the size or weight of the sample, the presence of DNA from a smaller species will be diluted by the excess DNA from the larger species. This would reduce the analytical sensitivity of the assay for the smaller species, although it could be compensated for by using more sequencing (in a metabarcoding method) or by using more but smaller samples (real time PCR or ddPCR). Another attempt to resolve this uses protocols in which samples are sorted and only parts (e.g. a leg) of the larger species are included in the extraction. However, this requires both manual processing and sufficient identification expertise, reducing the likelihood of the method being automated for high-throughput use and increasing time and cost.

If the aim is to investigate relative abundances in populations of species using metabarcoding, the issues can become more involved and difficult to resolve. If the assay target is a single copy nuclear gene, then assessing the number of reads as an estimate of the number of cells of each species present is possible (ploidy issues aside). This could be used to infer something about the relative abundance of each species but is confounded by differences in biomass or growth stage of different individuals and species. The situation becomes even more complex where the target is an organelle such as the mitochondrion (e.g. insect cytochrome oxidase I) or is multicopy (e.g. fungal ITS region). The number of mitochondria per cell will vary between species, individuals, and even specific tissues, whilst the number of copies of multicopy genes will vary between species and individuals. These issues make it difficult to quantify species abundance. It may still be possible to compare differences between sites sampled using the same methods, but in some cases relative abundance from metabarcoding results (e.g. as measured by read count) may be very different from the true relative abundance.

Related Questions:
Q10: What are the effects of large amounts of non-target DNA in a sample?

Back to Contents

Q10 What are the effects of large amounts of non-target DNA in a sample?

The effect of non-target DNA depends on the application and the nature of the non-target DNA. The non-target DNA can either be related organisms (e.g. smooth newts in a test for great crested newt) or unrelated (e.g. plant DNA when screening for plant pathogens). Overall, unrelated DNA is unlikely to cause major issues for targeted assays or for metabarcoding with highly specific primers. Although very high levels of DNA act as an inhibitor within a PCR (DNA accumulation causes DNA production to plateau), this is not a concern for those commissioning a test or study. The effects of high levels of DNA in a PCR can be corrected by diluting the DNA, although this could reduce analytical sensitivity if DNA from the target organisms is at or close to the limit of detection.

Related DNA from non-target species can cause more specific issues. In metabarcoding experiments, related non-target DNA will frequently be amplified and sequenced alongside the target DNA, reducing the analytical sensitivity of the assay. This issue can also cause problems with what appear to be unrelated species. Many bacterial 16S primers will amplify the 16S from plant chloroplasts, and fungal ITS primers can also amplify the ITS region of plants. The theoretical solution to non-target amplification is to make the PCR primers more specific, but there is a difficult balance between amplifying a wide range of target species whilst avoiding undesirable non-target species.

Related non-target DNA can also cause issues in targeted assays. Assays designed to distinguish between closely related species often have some (low) levels of cross reaction. In the presence of large amounts of related non-target DNA, this cross reaction creates competition that can mask detection of the target species if the target is present at low abundance. For example, the real time PCR assay to detect the potato pest *Globodera rostochiensis* cross reacts if *G. pallida* is abundant, meaning that *G. rostochiensis* is effectively undetectable with the assay in this situation. Digital droplet PCR (ddPCR), which relies on thousands of individual reactions, can be a solution to this issue as each reaction should theoretically only contain one or no DNA target molecule and thus there is no competition between target and non-target DNA.

Related Questions
Q9: What effect do large differences in organism size have on results?

Back to Contents

Q11 What is the effect of DNA extraction kit choice on results?

During the DNA extraction step, a sample (or subsample) is treated to recover and concentrate the DNA into a small volume of buffer whilst removing other chemicals and materials. Extraction methods vary hugely in complexity and duration, from methods that can be done in minutes to those that take days or longer. Typically, more complex and difficult sample types require more complex extraction methods. Some methods also require higher quality or longer fragment length DNA, which may involve more specialist DNA extraction methods.

Different DNA extraction methods can have a range of effects on the final results, from negligible to very important. An inappropriate extraction method will reduce analytical sensitivity and increase false negative rates, as DNA is poorly recovered from the original sample or inhibitors remain in the extracted DNA. Note that the choice of extraction method is also likely to influence the community composition in metabarcoding studies, as the different organisms within the sample may respond differently to different DNA extraction methods.

Negligible effects of extraction methods are likely for 'easily extracted' samples with abundant DNA, such as fresh animal tissue, pure bacterial cultures, and 'simple' plant material (e.g. some fresh leaf tissues). The effects (increased false negative rate, changes in community composition) are likely to be higher for:

Physically tough tissue samples
The DNA is protected by some form of outer casing, including: tough plant materials (e.g. woody tissue, tough leaves), organisms with hard outer casings (e.g. some diatoms, fungi, spores, some invertebrates), and organisms with protective mucus coats (e.g. some invertebrates, some bivalves). These casings generally require some disruption in the extraction steps, such as crushing, heating, and/or enzymatic digestion. If the DNA extraction steps omit these, the amount of DNA recovered will typically be far lower than more appropriate DNA extraction methods.

Samples with inhibitors
Inhibitors are substances that inhibit DNA extractions or PCR, and are found in some plant tissues (e.g. tissues with tannins, phenolics, high starch content, polysaccharides), as well as any samples that contain soil or sediment (these contain many inhibitors such as humic acid) – e.g. samples concentrated from water, soil samples, faecal samples. Any DNA extraction on these sample types either should be tailored to 'difficult' extractions or contain a final 'clean-up' step as a separate component. A range of proprietary kits are now available that are tailored to improve extraction efficiency of these difficult samples. Even these extraction methods and kits should be empirically demonstrated as fit for purpose.

Samples with low starting concentrations of DNA
This includes any degraded sample, or samples where the target organism is at very low concentration. Small variations in DNA extraction efficiency or yield will have a greater effect on the end result when the DNA concentration is approaching the threshold of detection.

Related Questions
Q13: How much effect does inhibition have on results, and how can this be understood and minimised?

Back to Contents

Q12 What is the effect of the choice of primers on results?

The primers determine which fragments of DNA will be amplified and therefore detected; thus, primer choice is one of the most important factors for real time PCR, barcoding, or metabarcoding studies. For any given study, the ideal primer sequences would achieve several aims, but in practice the best primer sequences will represent a compromise.

At the broadest level, there is a balance between **(i)** amplifying only fragments of DNA from intended genes/organisms targeted by the study (and from no others), i.e. intended bias, and **(ii)** obtaining amplicons (PCR products) that contain the most informative, taxonomically discriminating sequences (for barcoding/metabarcoding especially, these must distinguish between taxa). There are also necessarily compromises within and between these two aims, as well as practical considerations – e.g. short sequences may amplify more effectively than long sequences but be less effective at allowing species discrimination.

In targeted PCR assays, the primers and probe need to be inclusive enough to detect the whole range of sequence diversity found in a target organism while not giving false positive results with non-target taxa. When selecting primers, off-target binding should be screened for and avoided: even if it does not lead to false positives, any off-target binding of assay primers can lower the efficiency of an assay and potentially lead to reduced analytical sensitivity and increased false positive rate.

Efficiency of targeting intended genes/taxa (intentional bias)
For both barcoding and metabarcoding, the primers must adequately amplify a single gene or DNA sequence without amplifying other DNA regions elsewhere in the genomes. For the purpose of metabarcoding, primers would ideally maximise amplification of the DNA of all types of organism of interest ("target organisms" such as bacteria, fish), and do so with minimal bias between species within that group, and should minimise DNA-amplification of groups not of interest ("off-target"). The drawback of non-target amplification is that it wastes DNA sequencing effort, not that it affects the identification of sequence DNA from the targeted groups. Therefore, a low frequency of off-target amplification is usually acceptable unless sequencing capacity is very limited in relation to the number of samples.

Unwanted bias: different target taxa are amplified disproportionately
Some widely-used primers suffer from unintended bias (Pinto & Raskin 2012; Walters *et al*. 2016) – even though all organisms within the target group are amplified, amplification of some organisms is favoured relative to others. Consequently, resulting metabarcoding data can generally only be treated as indicative of the relative abundances of the organisms detected rather than an accurate measurement of their proportions. Unintended bias can be caused by differences in PCR amplification efficiency between species, e.g. differences in primer-binding affinity to (potentially) slightly different target sequences in each taxon. In addition, some target genes, such as 16S rDNA, are present in different copy numbers per genome in different taxa.

Irrespective of the causes, for a given primer pair, the *relative* biases in amplification are reproducible (Ibarbalz *et al.* 2014). This means that comparisons of ratios of proportional abundances (taxon A: taxon B) between two or more samples, remain valid.

Informativeness of the amplified sequences: identifying to taxon
Regarding the informativeness of the amplified sequences (typically a barcode region), the idealised amplicon sequence would not only be unique in each taxon of interest (e.g. each species), but the differences in the sequence between each species would be substantial. This would mean that sequencing errors introduced by the sequencing platform would rarely

if ever prevent correct identification. However, this is usually an unrealistic expectation, especially for metabarcoding.

Some primers designed to perform well in one aspect (amplifying a large target group, e.g. “all bacteria”) do so at the cost of providing fine resolution between taxa. For example, fungal ITS1 and bacterial 16S amplicons from widely-used primers do not resolve beyond genus level in the general case, although this depends on the taxa: some species can be uniquely identified, while in other cases a small group of several species can be distinguished from other members of the same genus. Conversely, within some families, even reliably resolving particular genera can be problematic.

Using longer amplicon sequences (i.e. longer DNA sequences from the barcode region) or more than one barcode region may provide greater power to resolve down to the desired taxonomic level. In some taxa (e.g. plants, nematodes) multiple different barcoding genes are recommended to resolve the full range of species / genera of interest. Note that in some cases it may not be possible to resolve down to the desired taxonomic level just using a barcoding approach. Prior to metabarcoding it is therefore critically important to specify the taxonomic level to which identification is required in general, and whether there are any particular taxa whose positive identification must be possible in principle. A related basic issue to consider is the availability and richness of reference sequences (i.e. databases) corresponding to the region amplified. The richest databases are dedicated to particular genes and/or intergenic regions and are more detailed than sequences available in general nucleotide sequence databases. Therefore, the choice of primers can also determine the extent to which a DNA metabarcoding dataset can be analysed.

Related Questions:
Q21: To what extent does the quality of reference databases affect the reliability of results?
Back to Contents

Q13 How much effect does inhibition have on results, and how can this be understood and minimised?

Inhibition is likely to be more problematic in sample types that contain inhibitors, such as soil, filtered/precipitated water samples, sediment, faeces, plant samples containing woody or tough plant materials, and animals that produce mucus (bivalves, snails, *etc.*). It is less likely to be a problem in some aggregations of organisms without the matrix they were contained in and in some animal tissues.

The effect of inhibitors depends on whether they are adequately controlled by the method. Factors that can reduce the effect of inhibitors include: the amount of starting material (not too much), correct choice of DNA extraction method, the use of reagents optimised to deal with inhibitors (e.g. inhibitor resistant PCR reagents; see Uchii *et al.* 2019), diluting the DNA prior to spiking into the PCR reaction, the use of DNA and PCR clean-up steps, and the choice of detection technology (both LAMP and ddPCR are more resistant to inhibitors than real-time PCR; Francois *et al.* 2011; Salipante & Jerome 2020).

The effect of inhibitors is better understood for species-specific methods than for metabarcoding. In the former, if not properly controlled for, inhibitors can increase the false negative rate and decrease the analytical sensitivity. In theory the presence of inhibitors would reduce the efficiency of a metabarcoding PCR amplification step, and this could lead to an increase in amplification bias in favour of specific species – this area needs further study.

Inhibition can be detected by including a control (i.e. spiked DNA) into the sample prior to PCR and after extraction. If the spiked DNA can be detected efficiently then the effect of inhibition is likely to be sufficiently small for the purposes of the test. Alternatively, inhibition can be detected in real-time PCR and ddPCR by creating a standard curve based on a dilution series of the sample; if significant inhibitors are present, the detection efficiency will be lower than expected at the highest sample concentrations.

Related Questions:
Q11: What is the effect of DNA extraction kit choice on results?

Back to Contents

Q14 What effect does the number of PCR replicates and the way in which they are treated have on results and costs?

For a targeted detection method that is being used above the analytical limit of detection (LoD), the number of PCR replicates is unlikely to have much effect on the results. The replicates can therefore be minimised (2-3 would be acceptable). However, when a method is expected to be used at, or just above, the LoD, increasing the number of replicates will increase the sensitivity (e.g. GCN detection in eDNA samples).

At the LoD, not all the replicates will be expected to give a positive result even when the target DNA is present. Increasing the number of replicates will increase the probability of at least some being positive. However, increasing the number of replicates also increases the likelihood of false positive detections and may require additional negative control replicates to identify when this is a risk. There will be increased costs with increasing numbers of replicates due to reagent requirements and because fewer different samples can be analysed at the same time (PCR typically uses plates with 96 reaction wells; leaving some wells for controls, a typical assay done in duplicate could process 45 samples but if 12 replicates are used only seven different samples could be run in a single experiment).

Typically, a single positive replicate is sufficient to give a positive diagnostic result. For example, the eDNA test for great crested newt interprets a single positive replicate as a positive diagnostic result. Where there are multiple replicates, a threshold for a positive diagnostic result could be put in place (e.g. $\geq$2 positive PCR replicates necessary for a positive diagnostic result). This can form part of the validation process of the test, allowing the effect on the positive and negative predictive values to be calculated.

Metabarcoding

The effect on metabarcoding is more complex and is influenced by both the random sampling of DNA at the LoD for less abundant organisms and the iterative nature of PCR causing random amplification ("jackpot") events. This can lead to PCR 'drift' favouring particular sequences and disfavouring others (Wagner *et al.* 1994; Kebschull & Zador 2015), potentially producing bias that significantly affects any taxa and not just those at very low abundance.

If truly random, different PCR drifts can be expected to occur in different PCR replicates, and so pooling the amplicons of independent replicates would even out the drift. The nature and size of drift is independent of the polymerase enzyme used and the number of rounds of amplification (Sze & Schloss 2019) and so is in that sense random. Accordingly, comparing pooled triplicate PCR amplicons with the same numbers of reads from single PCR reactions demonstrated a slightly higher repeatability in the former (i.e. when repeating PCR/sequencing runs on the same sample) (Kennedy *et al.* 2014). A counter-argument is that due to improvements in both polymerase fidelity and data analysis, pooled triplicate PCR reactions no longer provide sufficient benefit to warrant the additional costs incurred (Marotz *et al.* 2019). It is not straightforward to say whether replicates do lead to more repeatable results, and this may also depend on the sample type and method.

Back to Contents

Q15 What effect does one vs. two-step PCR have on results?

One and two step PCR protocols both have advantages and disadvantages. A one step protocol is easier to carry out, requires less staff time and has fewer steps so is less likely to lead to contamination. However, one of the disadvantages is that specific indexed barcode primer sets of hundreds of oligonucleotides must be purchased for every gene to be sequenced. When large numbers of samples are planned in single runs this can get very expensive. By contrast, when carrying out two step PCR only a single set of indexed primers is needed, and a new target gene can be trialled with the purchase of two primers. There is also good evidence that the longer more diverse primers used in one-step PCR can cause issues and bias in the PCR process leading to increased variability in replicates caused by potentially large differences in the proportions of the sequenced taxa between replicates (see Zizka *et al.* 2019; O'Donnell *et al.* 2016).

Q16 What effect do differences in sequencing depth have on results?

In this context, sequencing depth is how many sequencing reads are produced per metabarcoding sample. The difference in the results obtained by different depths depends very much on what questions the study is aiming to answer, and also on the complexity of the community being studied.

If the objective is to obtain as detailed a profile of a community as possible, then communities that include many organisms with very low-abundance DNA (e.g. species that are small, present in low numbers, or have DNA that is difficult to extract) will require much deeper sequencing than communities with relatively similar species abundances and only a few different species that each yield similar amounts of DNA. In that regard, many (but not all) microbial, diatom or fungal communities would require deep sequencing. Within the research community, a minimum of 100,000 reads is frequently quoted as a minimum for 16S metabarcoding, but this is based somewhat on arbitrary expectations. It has been shown that even with tens of millions of reads per sample, unique Amplicon Sequence Variant (ASV) counts do not plateau with increasing number of reads, albeit without denoising (Caporaso *et al*. 2011). (*Also see Question 18 on approaches to removing potentially erroneous sequences*)

Planning a metabarcoding project requires considering what is the maximum relative abundance of organisms that would be acceptable to fail to assay accurately (or completely miss). For example, it has been shown that even 1,000 reads per sample successfully enabled determination of taxa at 1% abundance with reasonable accuracy (Hamady & Knight 2009). (Many studies of that era used 454 pyrosequencing, which at that time did not produce read lengths substantially longer than the longest short reads of today and had considerably worse error frequencies). It should also be noted that most measures of within-sample (α-)diversity, such as the Shannon diversity index or the low-magnitude Hill numbers, are relatively robust to changes in numbers and relative abundances of very low-abundance taxa. Measures of between-sample (β-)diversity are even more robust; for example, the same assessment of pyrosequencing study noted that even 100 reads per sample could detect major patterns of variation. A study using > 2 x 10^7 reads per sample (Caporaso *et al*. 2011) demonstrated that β-diversity could be reproduced even when using only 2,000 randomly-selected reads. The Illumina protocol for their iSeq 100 System refers to a range of 15,000 to 100,000 reads per sample depending on the purpose of the study and the target assemblage (Illumina 2018).

Related Questions:

Q18: What are the approaches to removing potentially erroneous sequences from datasets, and do these produce different results?

Back to Contents

Q17 How does the sensitivity of different PCR methods compare with each other?

For species-specific assays, ddPCR is marginally more sensitive than real-time PCR for most applications (e.g. Salipante & Jerome 2020). Although a mass of contradictory studies are available, the consensus is that conventional PCR (in which detection is via gel electrophoresis) is less sensitive than either real-time or ddPCR (see e.g. Lau *et al.* 2003; Hren *et al.* 2007; Paiva-Cavalcanti *et al.* 2010, counter example Bastien *et al.* 2008). As a side note, multiplexed assays, in which more than one test for different targets are run in a single reaction, generally have declining sensitivity as more primer pairs are added.

For most applications, other factors come into play when deciding between PCR methods. Real time PCR is more suited to high thoughput applications than conventional PCR with gel electrophoresis (and possibly ddPCR), and contamination control is also better using real-time PCR than conventional PCR. However, the machines for both real-time PCR and ddPCR are expensive, and so conventional PCR followed by gel electrophoresis may be most appropriate for early phases of method development or for research projects.

Studies comparing metabarcoding with real time PCR for targeted species detection have generally found metabarcoding to be less sensitive (e.g. Bonants *et al.* 2019; Bylemans *et al.* 2019; Wood *et al.* 2019). This fits with theoretical expectations – e.g. competition of targets within the 'untargeted' PCR reaction, ambiguity of interpretation where low read counts are observed. Counter examples can be found but note that for all of these method comparison studies, the sensitivity of the test is likely to be determined by the PCR primers, which vary between the real-time PCR and metabarcoding assays.

Related Questions:
Q23: How does metabarcoding compare with targeted assays for single-species detection?
Back to Contents

Q18 What are the approaches to removing potentially erroneous sequences from datasets, and do these produce different results?

Basic Quality-Control (QC)
The 'normal' quality-control measures applied to Next Generation Sequencing (NGS) reads will generally remove some erroneous sequences. The process includes steps to:

- Filter based on lack of sufficient agreement between paired reads
- Remove expected short sequence motifs such as those matching the primers
- Trim sequences with low quality scores (all platforms give each base a quality score, which is a prediction of how likely the sequence is to be accurate)
- Discard reads that are too short after trimming, or that are too long and thus likely to be non-target amplification.

These procedures usually dramatically improve results, by reducing the numbers of erroneous Operational Taxonomic Units (OTUs) and thus avoiding overinflated measures of diversity (Bokulich *et al*. 2013). As with many processes, differing threshold values and software packages can produce differing results making it difficult to compare results obtained using different methods.

Dealing with the consequences of base-call errors in sequences
Even with a very good average base quality score, in absolute terms many sequence reads will contain one or more errors. Consequences of this include: overestimating the number of different sequences, with implications for estimates of diversity; inability to identify a read to more than a coarse taxonomic level; taxonomic misidentification (in some cases).

Some methods of processing amplicon reads can ameliorate these problems even without explicitly attempting to recognise base-call errors. By clustering very similar sequences, the erroneous versions of some reads will be grouped with their error-free equivalents. The cost of this approach is that it does not distinguish genuine biological sequence variants (e.g. two very similar taxa) from errors, thus placing a limit on the taxonomic resolution. This limit can be adjusted by changing the threshold at which similar sequences are clustered, but it may be difficult to have a 'perfect' threshold that keeps all true taxa separate, whilst ensuring all erroneous sequences are clustered with error-free equivalents.

The problem is addressed through a number of "denoising" approaches that have been developed to recognize sequencing errors. Errors are either corrected on the basis of very similar but much more abundant sequences (the error free sequences are expected to be in higher abundance) (e.g. Callahan *et al*. 2016; Schloss 2020), or the reads containing them are regarded as inadmissible failures that could confound the analysis and so are discarded (e.g. Amir *et al*. 2017). The drawback of discarding errors is that it potentially wastes reads that could be real variation (e.g. true species). On the other hand, attempts to correct perceived errors (typically by comparison with near-identical reads and assessing frequencies of each) can have consequences. For example, denoisers have been shown to artificially inflate the distinct sequence variants, including by a factor of >2 in some studies of mock communities (Nearing *et al.* 2018). Interestingly, the same studies show that these effects can depend on the type of biota studied, with the same denoising software either significantly underestimating or significantly overestimating the number of variants present, depending on the mock community.

In summary, the goal of recent approaches in both 'correction' and 'elimination' categories is to produce unique sequences with associated frequencies, in which each sequence

represents a true amplified sequence (amplicon sequence variant, or ASV). ASVs are in effect the finest-grained possible OTUs.

Dealing with PCR artefacts: chimeras

For any protocol, some chimeric amplicons can be expected, formed by the annealing of partially-synthesised single DNA strands from two or more distinct DNA targets of different biological origin. Chimeras usually account for a high proportion of the unique sequences in a metabarcoding dataset, and therefore can have a dramatic effect on the number of ASVs/OTUs. For example, a nematode benchmark study (Porazinska *et al.* 2012) found that more than 40% of OTUs were the result of chimeras. Depending on how similar the two or more "parent" sequences are, these chimeras may be identifiable as associated with only a high- or very high-level taxon, resulting in a high proportion of essentially unidentifiable reads. However, the chimera frequency is highly protocol-dependent (depending on the polymerase), ranging from 1-18% of all reads generated by 30 PCR rounds in a 16S metabarcoding benchmark study (Sze & Schloss 2019). 18S metabarcoding of nematodes exhibits a similar order of error rate (Porazinska *et al*. 2012).

Many software tools are available for detecting and removing chimeric reads, including both *de novo* methods (comparing reads with each other for consistency) and those using reference databases. It is generally essential to perform such filtering, though none can be expected to eliminate all chimeras. Some tools are specific to particular barcode genes, e.g. for 16S (Ashelford *et al.* 2005) or fungal ITS (Nilsson *et al*. 2010), while others are more general (Edgar *et al.* 2011). Filtering methods in general represent a compromise between sensitivity (lowering the false-negative rate) and specificity (low false-positives). The best sensitivity occurs for chimeras between more divergent sequences, with rates close to 100% (Edgar 2016). However, unsurprisingly, chimera-formation is most common between closely related sequences. While such chimeras are the hardest to detect, the hardest of all may be the least important, because differences between the two "parent" sequences may be too small to make a difference to the identified taxon. Indeed, many chimeras may be literally impossible to detect, because many true biological sequences are identical to hypothetical chimeras of two other, closely related but different biological sequences (Edgar 2016).

Related Questions:

Q19: What are the effects of using different approaches to assigning taxonomy (e.g. algorithms, OTU clustering thresholds)?

Back to Contents

Q19 What are the effects of using different approaches to assigning taxonomy (e.g. algorithms, OTU clustering thresholds)?

The choice of *reference database* has by far the greatest potential influence on the final taxonomic assignment. Any class of taxonomic assignment method (see below) relies on comparing sequence characteristics with those of reference taxa, which must be distinguishable. The presence in the database of a single misidentified reference sequence (also called 'mis-annotated') – e.g. a sequence mis-identified as originating from a different phylum to many other higher similar sequences – can prevent sequence reads from being correctly identified. This can apply even if the great majority of reference sequence identifications to species (or appropriate taxonomic level) are correct.

Approaches to sequence clustering and thresholds

Pairwise Alignment
Sequence identity (usually expressed as a percentage) from a *pairwise alignment* of a metabarcoding read versus another sequence (a taxonomically-identified reference sequence, or other sequences in the experimental data set) can be used as the basis of clustering reads into OTUs. There are several different approaches to this (Kopylova *et al*. 2016):

i. *De novo* clustering compares reads only with each other, and the resulting OTUs are assigned taxa later (e.g. to a single read as the representative of the cluster). Many clustering algorithms exist, and the result of a clustering exercise depends on the method and the population of reads. Note that although details differ, several of the now widely-used denoising approaches (also see Question 18) are in effect *de novo* clustering methods using a very high identity threshold (albeit discarding numerous likely erroneous reads as part of the process).

ii. *Closed-reference* clustering compares each read to a set of reference sequences that have already been clustered into OTUs with taxonomic identifications. Note that this comparison can also be done using pre-aligned reference sequences (see below).

iii. *Open-reference* clustering consists of (ii), followed by (i) applied to reads that did not match a reference sequence. The 'closed' method (ii) inevitably discards many sequences (e.g. that contain base call errors or are chimeric) and will often produce a total OTU count an order of magnitude lower.

Importantly, however, there is generally a many:1 relationship between OTUs and assigned taxa. For microbial communities at least, a similar number (and nature) of taxa are reported in the results irrespective of the clustering method used, as shown by microbial mock-community metabarcoding benchmark studies (e.g. Nearing *et al.* 2018). Remarkably, this analysis shows that even though the number of OTUs from traditional open-clustering methods are between 10 to 100 times those obtained from modern denoising approaches, the number and names of assigned taxa are extremely similar (Nearing *et al.* 2018). At this stage, it is unclear if this is a general trend for all metabarcoding studies, or a feature specific to bacteria and micro-fungi.

Using a pre-aligned set of sequences

When comparing query DNA sequences to reference databases, an alternative to pairwise sequence alignment is *comparison with a pre-aligned set of sequences* representing each taxonomic group. This takes advantage of the differing importance of different positions in the sequence (Schloss 2009). It can provide better comparisons but relies on the availability of a high-quality reference alignment.

Sequence identity thresholds

Whether the approach uses pairwise alignment or sets of reference sequences, sequence identity thresholds need to be defined. Rules of thumb about the threshold for sequence identity are sometimes given for a barcode region for a particular domain or kingdom (e.g. 97% similarity for prokaryote 16S, 95% or 97% for COI for vertebrates and invertebrates), but these rarely seem to hold true across different families. A higher threshold (e.g. 99%) will succeed in separating some species from each other but will also split more species into multiple OTUs.

Thresholds for clustering should be determined empirically from the reference dataset, or with reference to the literature. There may be no optimal threshold. Note also that for some major groups (e.g. bacteria, fungi) single barcode metabarcoding approaches cannot generally resolve reads beyond the genus level (but can resolve some species within some particular genera). However, if a narrower group is the focus, then dedicated primer pairs targeting different gene regions may provide greater resolution within that group.

Using features of sequences

In contrast to assessing overall sequence similarities, some methods characterise taxonomic reference sequences in the terms of the *features* they possess. 'Features' usually means the presence of particular oligonucleotide sequences that are conserved within taxonomic groups (e.g. RDP Classifier uses octamer frequencies; Wang *et al.* 2007). Some of these may identify organisms to a high taxonomic rank only, while others are more specific. Such methods may be relatively robust to sequencing errors elsewhere that could compromise identification based on overall sequence identity. Like the comparison with reference alignments, in effect, the most taxonomically important parts of the sequences are considered. Another advantage is computing speed. However, compared with methods that consider the entire sequence (both experimental reads and in the reference databases), more query reads tend to be unclassified. More sophisticated recent methods use machine-learning approaches to associate a variety of sequence features with taxa, e.g. (Bokulich *et al*. 2018) and show much promise.

Related Questions:

Q18: What are the approaches to removing potentially erroneous sequences from datasets, and do these produce different results?

Back to Contents

Q20 What are the effects of using different sequencing platforms?

There are several different platforms for carrying out High throughput Sequencing (HTS), each with different characteristics. The most common platform used for metabarcoding is the Illumina MiSeq, which can produce up to 30 million paired 300 nucleotide (nt) (i.e. 2x300nt) sequence reads in a single run, allowing barcoding of target genes up to about 500nt long. Larger Illumina systems including the NextSeq, HiSeq and NovaSeq can produce billions of reads but are limited to 2x150nt reads and thus smaller barcode targets. One characteristic of Illumina data is that the error rate increases towards the ends of the reads, particularly 300nt MiSeq reads.

Other platforms include the Thermo Scientific Ion torrent series (PGM and more recent Genstudio and Genexus), which can produce tens of millions of reads up to 600nt long. The Ion Torrent platforms fell out of favour due to higher error rates, particularly homopolymer error (errors in the number of repeated bases e.g.: TTT or TTTT or TT).

The third platform is the Pacific Biosciences series (RII and the more recent Sequel and Sequel II). These platforms produce up to 4 million reads per flowcell and the big advantage is that the reads can be much longer than for the two previous platforms (thousands of kilobases). This platform originally suffered from a high error rate (up to 15%) but by circularising the target DNA and sequencing round the circle multiple times this error rate is now equal to or lower than that obtained from Illumina platforms.

The final platform in mainstream use is the Oxford Nanopore based systems (MinION, GridION and PromethION). These can produce millions to hundreds of millions of reads and the length of the reads is only limited by the length of the DNA provided to it. Like the Pacific Biosciences systems, the Oxford Nanopore systems suffer from higher error rates (currently 5%-10% of bases incorrect but improving). These errors include homopolymer errors and insertions and deletions. Oxford Nanopore is also a very innovative company with new versions of reagents, flowcells, basecalling software and even platforms coming out, sometimes weekly. This is an advantage in that the platform is constantly improving, but also a disadvantage because there is no stable release that can be validated with confidence in a continuing supply. However, this situation is starting to improve. One major advantage of the MinION is that the platform itself is cheap (£1,000 compared to approximately £90,000 for a MiSeq) and with the use of the Flongle adapter it can run cheap flowcells (£75) producing hundreds of thousands of reads.

All platforms have been used for metabarcoding with early comparisons (D'Amore *et al*. 2016; Loman *et al*. 2012) between the MiSeq RSII and Ion Torrent PGM recommending using the MiSeq due to the lower error rates. More recently, the potential advantages of longer read metabarcoding on the Pacific Biosciences platform (Tedersoo *et al*. 2017) or MinION (Krehenwinkel *et al*. 2019) have been explored as it may allow better taxonomic resolution.

Due to their differing library prep methods and sequencing approaches with differing read lengths and particularly error profiles, it is very difficult to compare results obtained across different platforms. A 16S bacterial comparison performed between MiSeq, RSII, IonTorrent PGM and 454 (an old platform that has since been removed) found that "*The observed community composition was always biased, to a degree that depended on the platform, sequenced region and primer choice*" (D'Amore *et al*. 2016).

Back to Contents

Q21 To what extent does the quality of reference databases affect the reliability of results?

The quality of the reference database has an extremely important effect on the results. Where the reference library is complete (all taxa expected to appear are present and from the correct local population) and correct (sequences are good quality, correctly labelled and from correctly identified samples), and the gene region selected can reliably discriminate between the taxa of interest, the final results from this part of the process are highly likely to be correct. Where the reference library is incorrect, the final results are likely to be incorrect as well. The degree to which the results are wrong depends on the degree to which the reference library is wrong.

There are two broad types of database problems. Database gaps are cases in which there is no sequence for a species. A DNA sequence of that species will therefore not be identified at all,or will be assigned to a higher rank (e.g. genus, family) if a sufficiently related taxon is present in the database. Database errors are cases in which key reference sequences are incorrect. This will mean that an organism of that species may be incorrectly identified (because the DNA sequence of the organism 'matches' to a database sequence that has been identified to the wrong species), and/or that a species is incorrectly assessed as absent from the sample (if the database sequence for that species is incorrect, and there are no other matches for the database DNA sequence from the sample).

Sequence quality within a database can also have significant effects on whether sample sequences can be correctly matched to the reference sequence. Factors influencing sequence quality include sequencing errors, incorrectly trimmed primer sequences, and lab-specific artefacts. A sequence with incorrect or unresolved bases is less likely to reach the threshold for a match than it would if it were complete/ correct and will therefore increase the false negative rate for that species. To create a trivial example, a reference sequence 100 base pairs long with five random errors in it is highly unlikely to have a greater than 95% similarity to sequences from the same species. Shorter than expected reference sequences can also have a negative impact. If a percentage pairwise similarity threshold is being used to assign a sequence to a species using a BLAST search, shorter sequences require fewer errors (in absolute terms) to reach that threshold than longer sequences. Bioinformatics pipelines should have steps that ensure this does not occur (i.e. that libraries have sequences of the same length, etc), but this is not always the case. Manual assessment of the sequences and reported statistics like sequence overlap, e-values, and bit-scores can help identify these cases, but need to be interpreted correctly.

Back to Contents

Q22 How does metabarcoding compare with targeted assays for single-species detection?

If the key requirement is to monitor/study a small number of species, then species-specific assays are more likely to be fit for purpose, while if the intention is to study communities, then High Throughput Sequencing (HTS) methods such as metabarcoding are the only currently available methods. Similarly, if the sensitivity of the test is the key requirement, then species specific PCR methods using ddPCR and real-time PCR are the appropriate methods. However, many additional factors may have to be considered. These are outlined below, using real-time PCR and conventional PCR (a PCR reaction with unmodified primers where the result is visualised as bands on a gel) as examples of species-specific PCR methods, and metabarcoding as an example of an 'untargeted' method. ddPCR can be considered analogous to real-time PCR, although it is considered to have a higher sensitivity. Note also that combinations of methods can be used, e.g. using metabarcoding as a screen to identify samples that may contain key taxa (e.g. invasive species), which are tested again using a different test, e.g. a real time PCR assay. The sensitivity of the test will be dictated by the first (screening) test.

Cost
The per sample cost for a species-specific test is low compared to any HTS method, particularly for high numbers of samples. Even with the advent of very cheap HTS flow cells (e.g. the Oxford Nanopore Flongle), samples are unlikely to be cheaper for single species detection than if real-time PCR or conventional PCR are used. Where the survey is for more than one species, it may be more cost effective to run metabarcoding assays than multiple species-specific assays. This would have to be investigated on a case by case basis, taking into consideration that multiplexed (i.e. multiple assays in a single reaction) species-specific assays can be developed and the difficulties of validating metabarcoding.

Complexity of analysis
Interpretation of a real-time PCR assay is relatively straightforward, while HTS data are generally interpreted using bioinformatics pipelines, which can be automated but which can appear as a 'black box' to many users.

Importance of results
If the test result is of high importance (e.g. statutory action will be taken on the detection of a specific species), a consideration is how confidently the uncertainties associated with the method can be determined by validation. There is a well-established framework for validating a real-time PCR, ddPCR and conventional PCR assay, providing a strong degree of confidence in the both the assay (if it passes the method requirements) and in the framework used to conduct the validation.

Currently, there are ways to demonstrate the performance characteristics of the metabarcoding test or to show it returns an equivalent result to an existing test, but the accepted overarching validation framework does not yet exist. Consequently, a real-time or conventional PCR result could be considered more defensible than a metabarcoding result. One approach is to use metabarcoding as a screen and to follow up key findings with confirmatory species-specific test (e.g. Fox *et al.* 2019).

Negative controls
For species-specific assays, negative controls are simple to incorporate and interpret. For metabarcoding, there is not a consensus on what constitutes a 'fair' negative control (typically, a certain number of reads are considered permissible in a negative control), and negative results are often interpreted with reference to thresholds (i.e. a minimum number of

reads are required to provide a positive result). This adds ambiguity to the interpretation of results.

Contamination control

It is easier to control cross-contamination within a real-time PCR system than with metabarcoding. A main source of contamination within PCR based methods is amplicon contamination (where the PCR products at the end of the process are distributed back to the beginning of the process). With a real-time PCR system, the tubes are disposed of unopened at the end of the PCR amplification step, while in metabarcoding (and conventional PCR) the amplified PCR products have to be opened and handled during the subsequent process.

Quantification

If it is necessary to quantify the starting DNA in the reaction, then quantitative PCR (real-time PCR run with a size standard curve or similar) and ddPCR are the appropriate methods. Currently, metabarcoding does not give indications of the absolute starting concentrations of DNA (as samples are normalised for DNA concentrations prior to being sequenced) but can give an indication of relative proportions of DNA starting concentrations within a sample.

Sensitivity

If sensitivity is a key parameter, then species specific assays using real-time PCR or ddPCR should be used. Studies have compared metabarcoding to real-time PCR to detect targeted species and generally found metabarcoding less sensitive (e.g. Bonants *et al.* 2019; Bylemans *et al.* 2019; Wood *et al.* 2019). This fits with theoretical expectations (competition of targets within the 'untargeted' PCR reaction, ambiguity of interpretation where low read counts are observed). Counter examples can be found, but all these method comparison studies suffer in that the sensitivity of the test is likely to be determined by the PCR primers, which are different in the real-time PCR and the metabarcoding assays.

Detecting the unexpected

One important potential advantage of sequencing-based methods is the possibility of detecting the unexpected. This may be variants of existing species or related species not expected at a site (Fox 2019). Targeted assays are usually designed to avoid such detections.

Related Questions

Q17: How does the sensitivity of different PCR methods compare with each other?

Back to Contents

Q23 Which aspects of sample processing and data analysis are influenced by the sampling media (e.g. water, soil, bulk specimens, *etc*.) and which can follow similar approaches irrespective?

(NB: The question scope is mixed samples to be sequenced using HTS)

In general, methods diverge heavily until the point at which samples have been reduced to nucleic acids (i.e. extracted DNA or RNA). The sampling, DNA capture, and DNA extraction steps are therefore highly divergent and specific to the matrix or sample type while the PCR, library preparation, and sequencing steps generally follow the same protocol, provided the same taxonomic group is being targeted. The bioinformatics pipelines depend on the question rather than the base matrix and may or may not diverge.

However, the quality of the DNA recovered may differ between matrices, which potentially affects the choice of PCR, sequencing and analysis methods. For example, the DNA recovered from an eDNA water sample may be degraded and composed of shorter fragments of DNA, while the DNA recovered from an organism within a soil sample may be longer and of higher quality but co-purified with inhibitors. In this example, it is likely that differing PCR and sequencing strategies would be used (e.g. targeting shorter fragments for the poorer DNA), with ultimately different bioinformatics pipelines and end results.

Back to Contents

Acknowledgements

We are very grateful to end users from environmental public bodies for discussions and comments, in particular the following: Colin Bean, Katie Clark, Joan Cottrell, Willie Duncan, Debbie Leatherland, Iveta Matejusova, Mike Nelson, Kerry Walsh and Paul Woodcock.

This project was led and funded by the Defra DNA Centre of Excellence in 2019/20, which aims to transform Defra group science to support policy and delivery by sharing and demonstrating the wider potential of existing methods; enabling the rapid development and trialling of new DNA based method applications; providing a focal point for leadership in Defra, UK and across other government departments, and building shared capability including skills and facilities.

References

Adrian-Kalchhauser, I. & Burkhardt-Holm, P. (2016). An eDNA Assay to Monitor a Globally Invasive Fish Species from Flowing Freshwater. *PloS One*, **11(1)**, e0147558. https://doi.org/10.1371/journal.pone.0147558

Amir, A., McDonald, D., Navas-Molina, J.A., Kopylova, E., Morton, J.T., Zech Xu, Z. & Knight, R. (2017). Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems*, **2(2)**. doi:10.1128/mSystems.00191-16

Andersen, K., Lise Bird, K., Rasmussen, M., Haile, J., Breuning-Madsen, H., Kjær, K.H., Orlando, L., Thomas, M., Gilbert, P. & Willerslev, E., (2012). Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology*, **21(1)**, 1966-1979.

Anonymous. (2008). Strategies for the Storage of DNA. *Biopreservation and Biobanking*, **6**, 251-2.

Ashelford, K.E., Chuzhanova, N.A., Fry, J.C., Jones, A.J. & Weightman, A.J. (2005). At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Applied and Environmental Microbiology*, **71(12)**, 7724-7736. doi:10.1128/AEM.71.12.7724-7736.2005

Ballantyne, K.N., Salemi, R., Guarino, F., Pearson, J.R., Garlepp, D., Fowler, S. & van Oorschot, R.A., (2015). DNA contamination minimisation–finding an effective cleaning method. *Australian Journal of Forensic Sciences*, **47(4)**, 428-439.

Bastien, P., Procop, G.W. & Reischl, U. (2008). Quantitative real-time PCR is not more sensitive than "conventional" PCR. *Journal of Clinical Microbiology*, **46(6)**, 1897-1900.

Bokulich, N.A., Dillon, M.R., Bolyen, E., Kaehler, BD., Huttley, G.A. & Caporaso, J.G. (2018). q2-sample-classifier: machine-learning tools for microbiome classification and regression. *Journal of Open Research Software*, **3(30)**. doi:10.21105/joss.00934

Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R. & Caporaso, J. G. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nature Methods, 10(1), 57-59. doi:10.1038/nmeth.2276

Bonants, P., Griekspoor, Y., Houwers, I., Krijger, M., van der Zouwen, P., van der Lee, T.A. & van der Wolf, J. (2019). Development and Evaluation of a Triplex TaqMan Assay and Next-Generation Sequence Analysis for Improved Detection of Xylella in Plant Material. *Plant Disease*, **103(4)**, 645-655.

Bylemans, J., Gleeson, D.M., Duncan, R.P., Hardy, C.M. & Furlan, E.M. (2019). A performance evaluation of targeted eDNA and eDNA metabarcoding analyses for freshwater fishes. *Environmental DNA*, **1(4)**, 402-414.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J. & Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, **13(7)**, 581-583. doi:10.1038/nmeth.3869

Camps, J., Pommel, L., Aubut, V., Verhille, B., Satoshi, F., Lascola, B. & About, I. (2009). Shelf life, dissolving action, and antibacterial activity of a neutralized 2.5% sodium hypochlorite solution. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, **108(2)**, e66-e73.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J. & Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy Sciences of the USA*, **108 Suppl 1**, 4516-4522. doi:10.1073/pnas.1000080107

Champlot, S., Berthelot, C., Pruvost, M., Bennett, E.A., Grange, T. & Geigl, E.M. (2010). An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. PloS one, 5(9).

D'Amore, R., Ijaz, U.Z., Schirmer, M. *et al*. (2016). A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC Genomics*, **17**, 55. https://doi.org/10.1186/s12864-015-2194-9

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.

Edgar, R.C. (2016). UCHIME2: improved chimera prediction for amplicon sequencing. *bioRxiv*, (074252).

Edgar, R.C. (2018). Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics*, **34(14)**, 2371-2375. doi:10.1093/bioinformatics/bty113

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27(16)**, 2194-2200. doi:10.1093/bioinformatics/btr381

Eichmiller, J.J., Best, S.E. & Sorensen, P.W. (2016). Effects of temperature and trophic state on degradation of environmental DNA in lake water. *Environmental Science & Technology*, **50(4)**, 1859-1867.

Fox, A., Fowkes, A.R., Skelton, A., Harju, V., Buxton-Kirk, A., Kelly, M., Forde, S.M.D., Pufal, H., Conyers, C., Ward, R. & Weekes, R., Adams, I.P. (2019). Using high-throughput sequencing in support of a plant health outbreak reveals novel viruses in *Ullucus tuberosus* (*Basellaceae*). *Plant Pathology*, **68(3)**, 576-587.

Francois, P., Tangomo, M., Hibbs, J., Bonetti, E.J., Boehme, C.C., Notomi, T., Perkins, M.D. & Schrenzel, J. (2011). Robustness of a loop-mediated isothermal amplification reaction for diagnostic applications. FEMS Immunology & Medical Microbiology, 62(1), 41-48

Garnica, S., Schon, M.E., Abarenkov, K., Riess, K., Liimatainen, K., Niskanen, T., Dima, B., Soop, K., Froslev, T.G., Jeppesen, T.S., Peintner, U., Kuhnert-Finkernagel, R., Brandrud, T.E., Saar, G. & Oertel, Ammirati, J. F. (2016). Determining threshold values for barcoding fungi: lessons from Cortinarius (*Basidiomycota*), a highly diverse and widespread ectomycorrhizal genus. *FEMS Microbiology and Ecology*, **92(4)** https://doi.org/10.1093/femsec/fiw045

Gaspar, J.M. & Thomas, W.K. (2013). Assessing the consequences of denoising marker-based metagenomic data. *PLoS One*, **8(3)**, e60458. doi:10.1371/journal.pone.0060458

Hamady, M. & Knight, R. (2009). Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Research*, **19(7)**, 1141-1152

Hedges, S.B., Schweitzer, M.H., Henikoff, S., Allard, M.W., Young, D., Huyen, Y., Zischler, H., Höss, M., Handt, O., von Haeseler, A. & van der Kuyl, A.C. (1995). Detecting dinosaur DNA. *Science*, **268(5214)**, 1191-1194.

Hinlo, R., Gleeson, D., Lintermans, M. & Furlan, E. (2017). Methods to maximise recovery of environmental DNA from water samples. *PloS One*, **12(6)** https://doi.org/10.1371/journal.pone.0179251

Hren, M., Boben, J., Rotter, A., Kralj, P., Gruden, K. & Ravnikar, M. (2007). Real-time PCR detection systems for Flavescence dorée and Bois noir phytoplasmas in grapevine: comparison with conventional PCR detection and application in diagnostics. *Plant Pathology*, **56(5)**, 785-796.

Hunter, M.E., Ferrante, J.A., Meigs-Friend, G. & Ulmer, A. (2019). Improving eDNA yield and inhibitor reduction through increased water volumes and multi-filter isolation techniques. *Scientific Reports*, **9(1)**, 1-9.

Ibarbalz, F.M., Perez, M.V., Figuerola, E.L. & Erijman, L. (2014). The bias associated with amplicon sequencing does not affect the quantitative assessment of bacterial community dynamics. *PLoS One*, **9(6)**, e99722. doi:10.1371/journal.pone.0099722

Illumina Inc. (2018). 16S metagenomics sequencing with the iSeq™ 100 System. Retrieved from https://emea.illumina.com/content/dam/illumina-marketing/documents/products/appnotes/iseq100-16s-app-note-770-2018-009.pdf

Jones, E.P., Adams, I.P., Walshaw, J., Benucci, M., MacArthur, R., Boonham, N. and Bryce, S. (2020) Guidance for end users on DNA methods development and project assessment. JNCC Report No. 669a. JNCC, Peterborough, ISSN 0963-8091. https://hub.jncc.gov.uk/assets/8839bf2f-f934-4a9d-b25d-74a6d20ac875

Kebschull J.M. & Zador A.M. (2015). Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Research*, **43**, e143

Kennedy, K., Hall, M.W., Lynch, M.D.J., Moreno-Hagelsieb, G. & Neufeld J.D. (2014). Evaluating Bias of Illumina-Based Bacterial 16S rRNA Gene Profiles. Appl. Environ. Microbiol. 80, 5717-5722

Kim, Y.-T., Choi, E.-H., Son, B.-K., *et al*. (2012). Effects of storage buffer and temperature on the integrity of human DNA. 대한임상검사과학회지 (*Korean Journal of Clinical Laboratory Science*) **44**, 24-30.

Kopylova, E., Navas-Molina, J.A., Mercier, C., Xu, Z.Z., Mahe, F., He, Y. & Knight, R. (2016). Open-Source Sequence Clustering Methods Improve the State Of the Art. *mSystems*, **1(1)**. doi:10.1128/mSystems.00003-15

Krehenwinkel, H., Pomerantz, A., Henderson, J.B., Kennedy, S.R., Lim, J.Y., Swamy, V. *et al.* (2019). Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale, *GigaScience*, **8 (5)**, giz006, doi.org/10.1093/gigascience/giz006

Lau, L.T., Fung, Y.W.W., Wong, F.P.F., Lin, S.S.W., Wang, C.R., Li, H.L., Dillon, N., Collins, R.A., Tam, J.S.L., Chan, P.K. & Wang, C.G. (2003). A real-time PCR for SARS-coronavirus incorporating target gene pre-amplification. *Biochemical and Biophysical Research Communications*, **312(4)**, 1290-1296.

Liu, M., Clarke, L.J., Baker, S.C., Jordan, G.J. & Burridge, C.P. (2019). A practical guide to DNA metabarcoding for entomological ecologists. *Ecological Entomology*, **45**, 373-385

Mächler, E., Deiner, K., Spahn, F. & Altermatt, F. (2016). Fishing in the water: effect of sampled water volume on environmental DNA-based detection of macroinvertebrates. *Environmental Science & Technology*, **50(1)**, 305-312.

Marotz, C., Sharma, A., Humphrey, G., Gottel, N., Daum, C., Gilbert, J.A., Eloe-Fadrosh, E. & Knight, R. (2019). Triplicate PCR reactions for 16S rRNA gene amplicon sequencing are unnecessary. *Biotechniques*, **67**, 29-32

Moreau, C.S., Wray, B.D., Czekanski-Moir, J.E. & Rubin, B.E., (2013). DNA preservation: a test of commonly used preservatives for insects. *Invertebrate Systematics*, **27(1)**, 81-86.

Muha, T.P., Robinson, C.V., Garcia de Leaniz, C. & Consuegra, S. (2019) An optimised eDNA protocol for detecting fish in lentic and lotic freshwaters using a small water volume. *PLoS ONE*,**14(7)** e0219218.https://doi.org/10.1371/ journal.pone.0219218

Nearing, J.T., Douglas, G.M., Comeau, A.M., & Langille, M.G.I. (2018). Denoising the Denoisers: an independent evaluation of microbiome sequence error-correction approaches. *PeerJ*, **6**, e5364. doi:10.7717/peerj.5364

Nilsson, R.H., Abarenkov, K., Veldre, V., Nylinder, S.P.D.E.W., Brosche, S. & Kristiansson, E. (2010). An open source chimera checker for the fungal ITS region. *Molecular Ecology Resources*, **10(6)**, 1076-1081. doi:10.1111/j.1755-0998.2010.02850.x

O'Donnell, J.L., Kelly, R.P., Lowell, N.C. & Port, J.A. 2016. Indexed PCR primers induce template-specific bias in large-scale DNA sequencing studies. *PloS One*, 11(3). https://doi.org/10.1371/journal.pone.0148698

Paiva-Cavalcanti, M., Regis-da-Silva, C.G. & Gomes, Y.M. (2010). Comparison of real-time PCR and conventional PCR for detection of Leishmania (Leishmania) infantum infection: a mini-review. *Journal of Venomous Animals and Toxins Including Tropical Diseases*, **16(4)**, 537-542.

Pietikäinen, J., Pettersson, M. & Bååth, E. (2005). Comparison of temperature effects on soil respiration and bacterial and fungal growth rates. *FEMS Microbiology Ecology*, **52(1)**, 49–58.

Pinto, A.J. & Raskin, L. (2012). PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS One*, **7(8)**, e43093. doi:10.1371/journal.pone.0043093

Porazinska, D.L., Giblin-Davis, R.M., Sung, W. & Thomas, W.K. (2012). The nature and frequency of chimeras in eukaryotic metagenetic samples. *Journal of Nematology*, **44(1)**, 18-25.

Salipante, S.J. & Jerome, K.R. (2020). Digital PCR—an emerging technology with broad applications in microbiology. *Clinical Chemistry*, **66(1)**, 117-123.

Schloss, P.D. (2009). A high-throughput DNA sequence aligner for microbial ecology studies. *PLoS One, 4(12)*, **e8230**. doi:10.1371/journal.pone.0008230

Schloss, P.D. (2020). Reintroducing mothur: 10 Years Later. *Applied and Environmental Microbiology*, **86(2)**. doi:10.1128/AEM.02343-19

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.

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

Sze, M.A. & Schloss, P.D. (2019). The Impact of DNA Polymerase and Number of Rounds of Amplification in PCR on 16S rRNA Gene Sequence Data. *mSphere*, **4(3)**. doi:10.1128/mSphere.00163-19

Tedersoo, L., Tooming-Klunderud, A. & Anslan, S. (2018). PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. *New Phytologist*, **217(3)**, 1370-1385.

Tzeneva, V.A., Salles, J.F., Numova, N., de Vosa, W.M., Kuikmanc, P.J., Dolfing, J. & Smidta, H. (2009). Effect of soil sample preservation, compared to the effect of other environmental variables, on bacterial and eukaryotic diversity. *Research in Microbiology*, **160**, 89-98.

Uchii, K., Doi, H., Okahashi, T., Katano, I., Yamanaka, H., Sakata, M.K. & Minamoto, T. (2019). Comparison of inhibition resistance among PCR reagents for detection and quantification of environmental DNA. *Environmental DNA*, **1(4)**, 359-367.

Wagner, A., Blackstone, N., Cartwright, P., Dick, M., Misof, B., Snow, P., Wagner, G.P., Bartels, J., Murtha, M. & Pendleton J. (1994). Surveys of gene families using Polymerase Chain Reaction: PCR selection and PCR drift. *Systematic Biology*, **43**, 250-261.

Walters, W., Hyde, E.R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A. & Knight, R. (2016). Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal Internal Transcribed Spacer marker gene primers for microbial community surveys. *mSystems*, **1(1)**. doi:10.1128/mSystems.00009-15

Wang, Q., Garrity, G.M., Tiedje, J.M. & Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, **73(16)**, 5261-5267. doi:10.1128/AEM.00062-07

Wilcox, T.M., Carim, K.J., Young, M.K., McKelvey, K.S., Franklin, T.W. & Schwartz, M.K. (2018). Comment: the importance of sound methodology in environmental DNA sampling. *North American Journal of Fisheries Management*, **38(3)**, 592-596.

Wood, S.A., Pochon, X., Laroche, O., von Ammon, U., Adamson, J. & Zaiko, A. (2019). A comparison of droplet digital polymerase chain reaction (PCR), quantitative PCR and metabarcoding for species-specific detection in environmental DNA. *Molecular Ecology Resources*, **19(6)**, 1407-1419.

Yuan-ling, N., Spratt, D., Sriskantharajah, S. & Gulabivala, K. (2003). Evaluation of protocols for field decontamination before bacterial sampling of root canals for contemporary microbiology techniques. *Journal of Endodontics*, **29(5)**, 317-320.

Zizka, V.M., Elbrecht, V., Macher, J.N. & Leese, F. (2019). Assessing the influence of sample tagging and library preparation on DNA metabarcoding. *Molecular Ecology Resources*, **19(4)**, 893-899.

Appendix 1

Additional questions were also suggested by some public bodies. These are more detailed/technical or more general/open-ended and so were not specifically addressed in this project (though some do overlap with the FAQs). However, they illustrate other important end-user uncertainties and priorities and so are included in case of future relevance and interest.

Experimental Design

- How can we determine the optimum number of samples required from different matrices and what degree of spatial and temporal scale of sampling is appropriate?
- What are the critical factors in determining choice of gene for total community analysis? (Would using multiple genes usefully improve the dataset and increase cost significantly?)
- Is there an advantage to using degenerate primers when a broad taxonomic range is of interest?
- What steps can you take to test primers in silico before choosing a primer pair?
- Should more than one primer pair be used on the same sample?
- What measures can be taken to obtain maximum value from sequencing budget?
- Is it advisable to develop a mock community of the most likely organisms? If yes, should the mock community be morphologically identified and Sanger sequenced individually?
- What would be the minimum quality control steps/items required to produce robust data for qPCR and metabarcoding?

Sample Collection and Transport

- What factors should be considered when sampling different forms of DNA (e.g. particle-associated DNA from cells, organelles, DNA stuck to small particles, free DNA)?
- What are the effects of different filter materials and pore sizes?
- What are the necessary decontamination steps that need to be taken within and between sites?
- Are negative and positive field controls necessary, and if yes how should these be applied?

Sample Processing and Data Analysis

- Is there a benefit to removing some of the specimens of the dominant species to increase the chance of metabarcoding the less common components?
- What are the advantages and disadvantages of prioritising total DNA yield over DNA quality?
- What are the advantages and disadvantages of sequence length when assessing community diversity?
- What are the advantages and disadvantages of clustering-first vs. assignment-first pipelines?
- What would be the minimum quality control steps/items required to produce robust data for qPCR and metabarcoding?

Other

- What is an acceptable level of confidence for DNA-based methods for operational deployment?
- What metadata need to be reported with DNA results?
- What is the minimum methodological information needed to validate and repeat the work?
- How should samples be archived to facilitate future use?
- Are there any steps that can be taken to use the number of reads as a semi-quantitative indicator of abundance?
- Is there a role for MinIon technology in metabarcoding?
- Would it be possible to develop a standard evaluation method to estimate the accuracy of a pipeline for targeted metagenomics?
- What is the relative impact on data of factors such as sequencing errors, choice of gene, length of sequence, chimera identification, and quality of reference database?
- How could a non-expert make a judgement such as ‘these methods are similar enough to allow comparisons between datasets?’ and vice versa ‘these methods are too different to allow comparisons?’