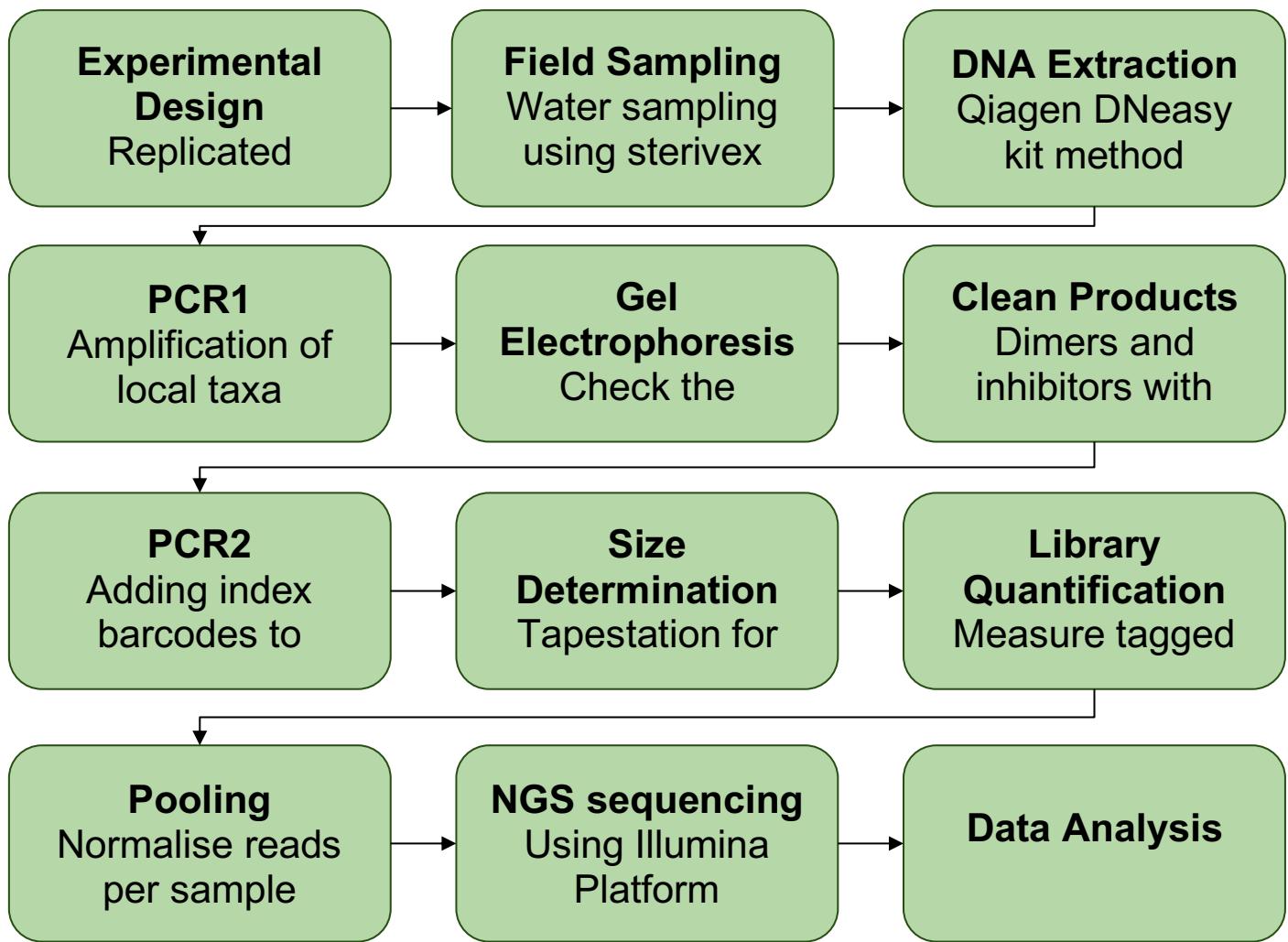


Metabarcoding and Environmental DNA Workshop:

Assessing biodiversity of a local pond using DNA metabarcoding

Date	Time	Activity	Location
08/09	From 13:30	Arrival - Welcome and Introductions	Alfred Denny Conference Room A225
08/09	14:00 - 14:30	1. Experimental Design	Alfred Denny Conference Room A225
08/09	14:30 - 15:00	2. Field Sampling	Weston Park
08/09	15:00 - 15:30	Refreshments	Alfred Denny Conference Room A225
08/09	15:30 - 17:00	3. DNA Extraction	Perak 1 Laboratory
09/09	9:30 - 11:00	3. DNA Extraction (continued)	Perak 1 Laboratory
09/09	11:00 - 11:30	Refreshments	Hicks Building, Lecture Theatre 4
09/09	11:30 -12:30	4. PCR1 Amplification	Perak 1 Laboratory
09/09	12:30 - 13:00	5. Quantification	Perak 1 Laboratory
09/09	13:00 - 14:00	Lunch	Hicks Building, Lecture Theatre 4
09/09	14:00 - 15:00	6. Gel Electrophoresis	Perak 1 Laboratory
09/09	15:00 - 15:30	Refreshments	Hicks Building, Lecture Theatre 4
09/09	15:30 - 17:00	7. Visualise Gel and Lab Tour	Perak 1 Laboratory
10/09	9:30 - 10:15	8. Bead Cleaning	Perak 1 Laboratory
10/09	10:15 - 10:30	9. PCR2 Indexing	Perak 1 Laboratory
10/09	10:30 - 11:00	10. Library Quantification and Pooling	Perak 1 Laboratory
10/09	11:00 - 11:30	Refreshments	Firth Court, Council Room
10/09	11:30 - 12:15	11. NGS Sequencing and Q&A	Firth Court, Council Room
10/09	12:15	Walk to Pam Liversidge Building for Lunch	Pam Liversidge Building - Atrium

Workflow



Introduction to the Metabarcoding Process

Metabarcoding allows us to:

- Identify a range of species in an environmental sample simultaneously, through genetic sequencing.
- This allows scientists to carry out a range of studies including diet analysis and population surveys.

How does it work:

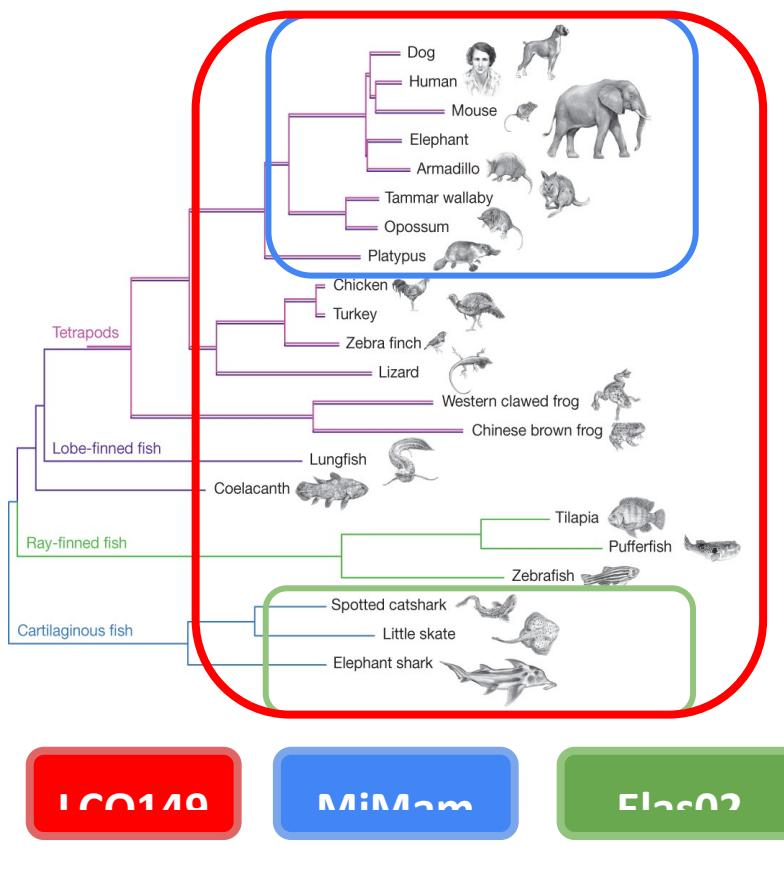
- Generic primers are used to target whole taxonomic groups, whilst bioinformatic analysis allows species identification.
- To be useful as a barcoding marker, the region amplified must be highly conserved within a species, but variable between species.
- The barcoding gene should be flanked by sequences that are conserved within the target phylogeny.
- Primer sets are designed to target the conserved region, which is then sequenced.
- Resulting sequence data can be compared to existing reference sequences for species identification.

Primer selection and the Diversity vs. Precision trade off

Evolutionary trees can show us how closely related different species are to each other. There are highly conserved regions of the genome that will be shared with all these species, and genetic regions that vary between species.

When designing your project, you will want to determine which taxonomic groups you wish to detect using genetic analysis. But bear in mind, the broader the taxonomic range you target, the more species you are likely to detect. And herein lies a tradeoff, between diversity and sequencing depth.

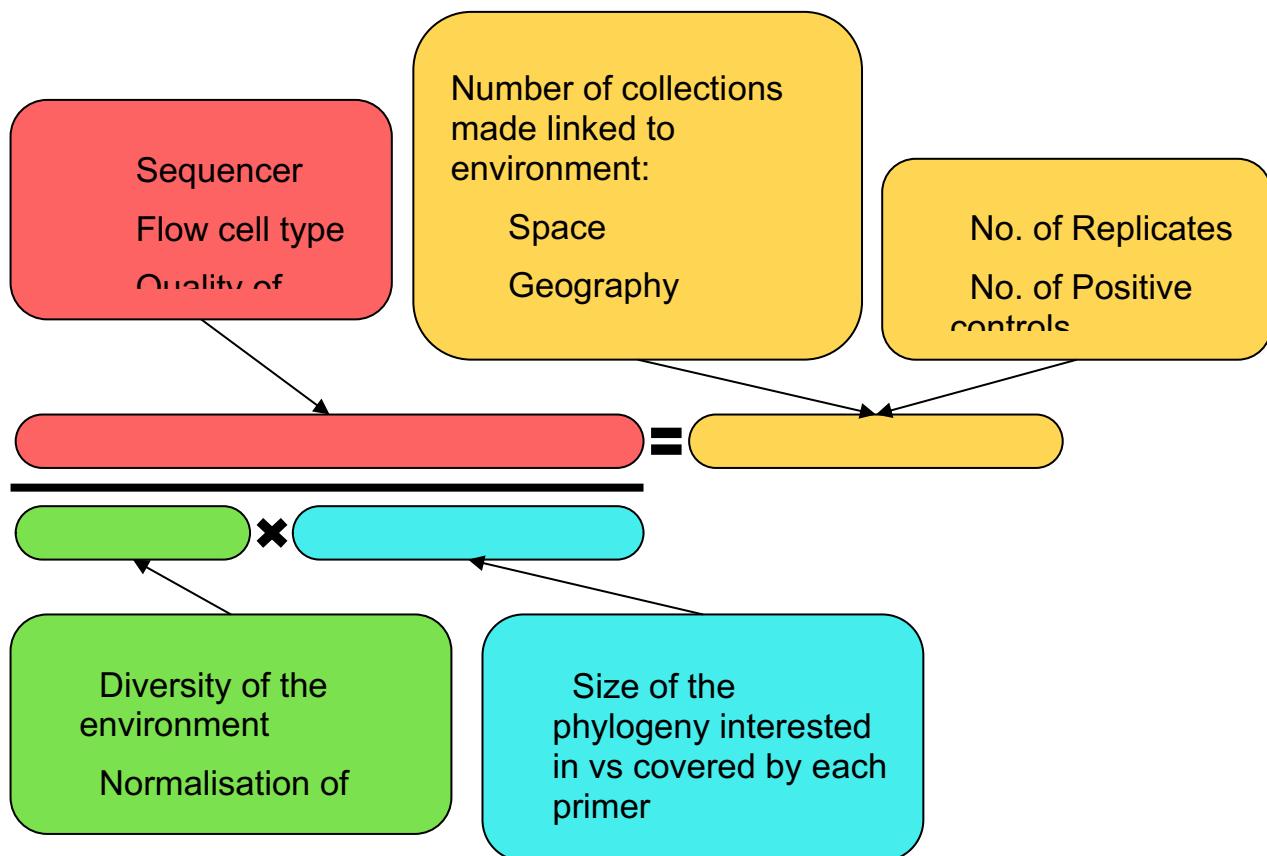
The flow cell on which your sequencing is run offers a limited number of reads, which can be allocated either to sequencing a few species multiple times - the sequencing depth, which increases the accuracy of the result - or allocated to sequencing many species fewer times, reducing the sequencing depth. If you want to sequence all species, at great



sequencing depth, you will require multiple flow cells, which will increase your costs. We must therefore calculate the number of species we can sequence at a reasonable depth to generate reliable results.

1. Experimental Design

When planning your experiment it is important to think about your project holistically, considering your research aims, the power you will need to answer your research questions, and the cost both in terms of finance and time.



Space The number of collections you make should be appropriate for the space you are monitoring. i.e. Larger area = Larger number of samples.

Geography The number of collections you make should be appropriate for the niche diversity of an area.

Seasonality Collections should be made in a time series if diversity is likely to be impacted by Season/migration patterns.

Grab or Passive sampling Grab sampling is limited to that point in time. Passive sampling

Samples, replicates, positives, negatives.

Example study

Miseq v2 2x150bp = 12 million reads

2 primer sets of interest both of which require 50,000 reads

12 million x 0.8 (presume adding 20% PhiX) = 96 samples

PhiX - PhiX is spiked in at higher concentration in metabarcoding runs in order to increase diversity.

Limited normalisation of DNA/PCR product concentration - Normalisation helps reduce the chance that highly concentrated samples will swamp the run.

Sequencer and flow cell size - Different flow cell sizes yield different numbers of overall

Number of Primer sets - Having a greater number of primers increases the power to detect a greater number of taxa.

Having more than 1 primer set with overlapping species coverage - If the species captured by two primer sets overlap this leads to a) more certainty of detection, b) increases the chance of capturing rare species and c) a greater power to calculate abundance.

Diversity of the environment/size of the phylogeny covered by the primer - 50,000 reads per sample per primer is a good rule of thumb, however, this may need to be increased for primers that target a greater numbers of species, such as Riaz12s which captures almost all vertebrate groups. Conversely, a marker like EwA-F 16s might be limited to family or order and so need fewer reads to be confident.

Unwanted species/blocking primers - If species of no interest are abundant in the sample that can be amplified by the chosen primer set (e.g. host insects in a parasitoid wasp study). Reads will be wasted unless blocking primers are incorporated into the experimental design.

Sample replicate - Different sample collections from the same location.

Extraction replicate - Independent extractions from the same sample.

PCR replicate - Separate PCR reactions from the same extraction.

Positive (Xeno) controls:

If a positive fails to produce a signal then something has likely gone wrong at that stage in the process. These should be DNA of high enough quality that should produce clear and quantifiable amounts of PCR product every time. However, ideally they should be a) alien either synthetic DNA or species not present in your targeted environment. This is to minimise any risk of cross contamination from the positive leading to false assignments of species in the unknown samples b) ideally be a mixture (mesocosm) of a number of species c) ideally each species should be quantified and input at known concentrations. d) the amount of Positive in the final pool should be normalised so it does not “use up” an excess of available NGS reads (swamping).

Negatives:

You generally want to incorporate a new negative at every stage of the process. This enables one to establish if contamination occurs and at what stage it happened.

Field Negative - A negative that has been through field processes (e.g. Sterile water passed through a filter in the field at the same time as sample collection). Tests for clean technique within the field.

DNA Extraction Negative - A blank (again normally sterile water) that is extracted at the same time as the samples. Tests for clean extraction technique

PCR negative - A blank (again normally sterile water) not extracted but simply amplified as a sample at the PCR stage (PCR1 and PCR2).

2. Field Sampling

Field Equipment:

Ladle	Caulking
Sterivex 0.45µm filter unit gun	
Syringe	Box of ice
250 mL bottles	10% Bleach
Field notebook	Tissue



1. To be performed in groups of three. Complete and sign a risk assessment. Label bottles and filter units.
2. Gloves should be worn at all times during the sampling process and samples should be collected without entering the water, preventing disturbance of the substrate and limiting contamination from other sources.
3. Within the sampling site, identify locations where the water will be collected. The location of subsamples should be spaced as evenly as possible.
4. Use the ladle to collect water and pour into a labelled bottle.
5. Clean the outside of the bottle with bleach and wipe dry with clean tissue paper.
6. Before filtration, shake the bottle in order to mix the DNA across the whole volume of the water sample.
7. Attach a 0.45µm sterile filter capsule to a sterile syringe. Pull the top part off the syringe, pour 50 mL of water (measure with the marks on the syringe) and push the water through the filter capsule. A caulking gun may be used to help filter samples with a lot of sediment.
8. Using the same filter and syringe, repeat this until the 200 mL has been filtered.
9. Store the filter inside a labelled 50 mL tube, and keep on ice.
10. For each water sample, repeat this 3x using new bottles, syringes and filter units each time.
11. Store samples in a -20°C freezer until extraction (or store in the fridge if DNA extraction is performed on the same day).

Recommended Sample Storage



HMW DNA

Long-read sequencing
for PacBio & ONT

RNA

cDNA construction,
RNA-Seq & Iso-Seq

DNA

SNP genotyping, NGS
Sequencing, Real-Time
PCR

Snap-freeze in Liquid Nitrogen

Store long-term at -80°C

Pros:

Produces the best quality DNA & RNA.
Eliminates extra step of removing
storage buffer.

Cons:

Impractical in the field.
Expensive to transport internationally.

RNAlater

Submerge in ~8x volumes. Stable for 1 day
at 37°C, 1 week at 25°C, 1 month at 4°C, and
long-term at -20°C or -80°C

Pros:

Maintains RNA & DNA integrity. Easy
to transport. Long term storage.

Cons:

Expensive.
Limited time at room temperature .

DNA/RNA Shield

Submerge in ~8x volumes. Store for 1 week
at 35-40°C, 1 month at 5-30°C and
long-term at -80°C

Pros:

Produces good quality DNA & RNA.
Inactivates infectious agents. Easy to
transport.

Cons:

Expensive.
Incompatible with long-read
sequencing.

Freezing

Keep samples at 4°C and freeze at -20°C or
-80°C as soon as possible (<24hrs)

Pros:

Cheap. -20°C freezers are easy to
access. Eliminates extra step of
removing storage buffer.

Cons:

Incompatible with long-read & RNA
sequencing.
Can be hard to transport.

Ethanol

20% sample to 80% ethanol ratio
Long term storage at room temperature or
-20°C

Pros:

Cheap.
Ideal for long term storage for NGS,
SNP genotyping & R-T PCR.

Cons:

Incompatible with long-read & RNA
sequencing.
Can be hard to transport.

Dry in Silica

Plant material, hair & feathers

Pros:

Cheap. Easy to transport. Eliminates
extra step of removing storage buffer.

Cons:

Not for all sample types.
Incompatible with long-read & RNA
sequencing.

Upward arrow indicating increasing DNA/RNA Quality from bottom to top.

3. DNA Extractions

DNA extraction is the process of separating DNA from cell membranes, proteins, and other cellular components by using physical and/or chemical techniques. There are many different methods of DNA extraction which usually depend on the *sample type*, *project aim* and *budget*.

Soil/stool extractions often require an extra step to remove any inhibitors that may be present in the sample, this may be as a chemical or an inhibitor removal spin column.

Plant samples may need stronger chemicals and/or physical grinding to help break down cell walls.

Hair/feather/swab extractions work with very low concentrations of DNA so methods usually focus on recovering as much DNA as possible.

eDNA samples, like water filters, are best extracted in a clean room to reduce the chance of external contamination.



If the aim of the project is to PCR target amplicons which usually range 50bp-1kb (metabarcoding, genotyping, qPCR), the chosen extraction method can be relatively harsh, chemically and/or physically, as shearing the DNA slightly will not cause a problem. This often allows for a higher recovery of DNA.

For long read full genome sequencing that requires good-quality high molecular weight (HMW) DNA (PacBio, Oxford Nanopore), methods that are more gentle are recommended. This can involve wide-bore tips, slower centrifugation, shorter lysis and gentle agitation.

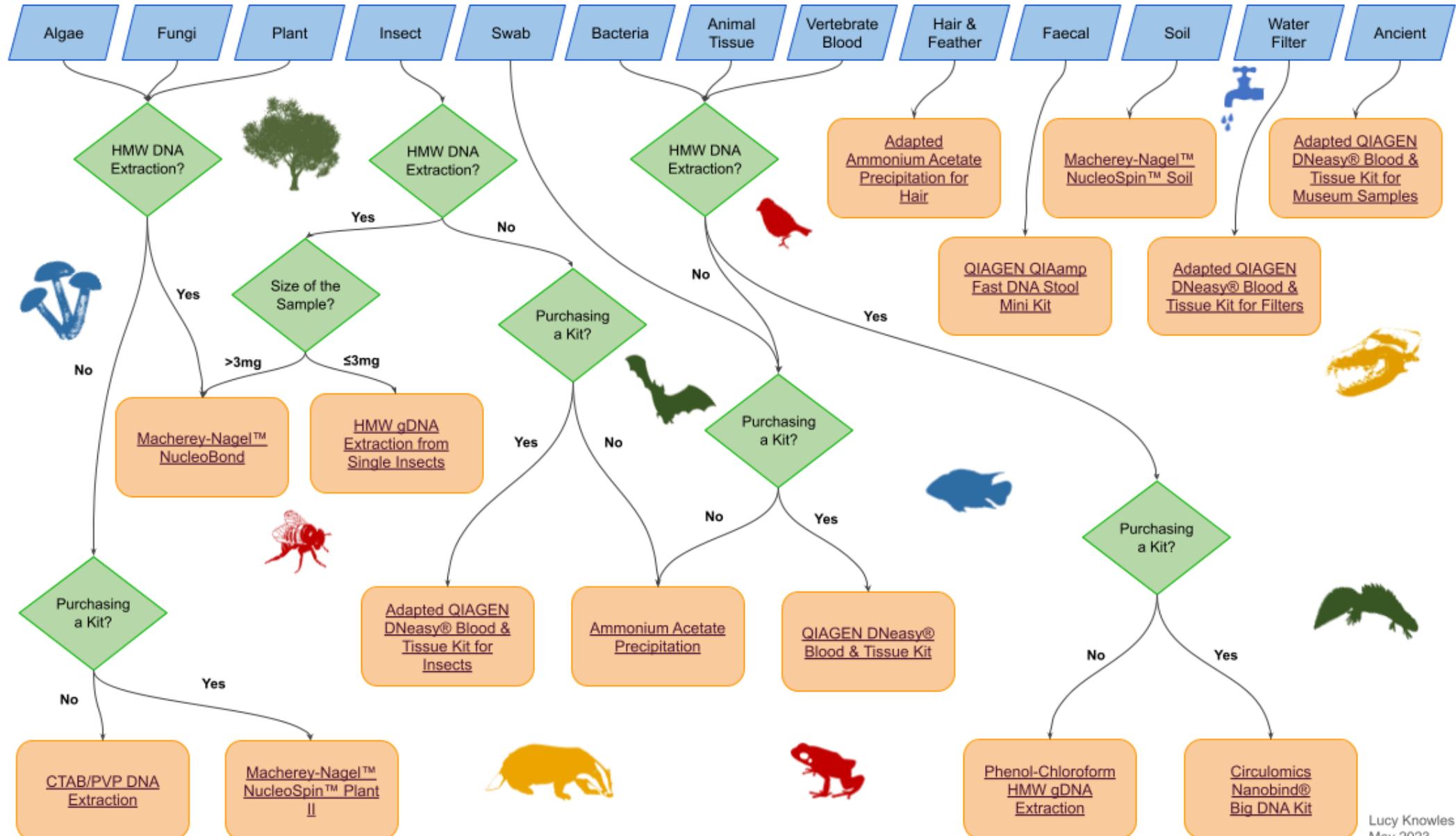
A large budget can allow for automated DNA extraction, using magnetic bead based extractions on robots.

Alternatively, there are many commercially available kits from companies such as QIAGEN, Macherey-Nagel, Zymo, Circulomics & New England Biolabs. These are generally easy to use and contain all of the necessary reagents, however, can also be expensive.

Home-brewed methods can also produce good quality, cost-effective results. They can be found in journals, [Protocols.io](#) or by inquiring with lab users, each lab often has their own favourite protocols. One of the oldest and most popular methods used for animal tissue/bird blood, known as the phenol-chloroform DNA extraction, is no longer used in the NEOF lab due to the hazardous chemicals involved; we have found the ammonium acetate precipitation method produces similar great results whilst being safer and easier.

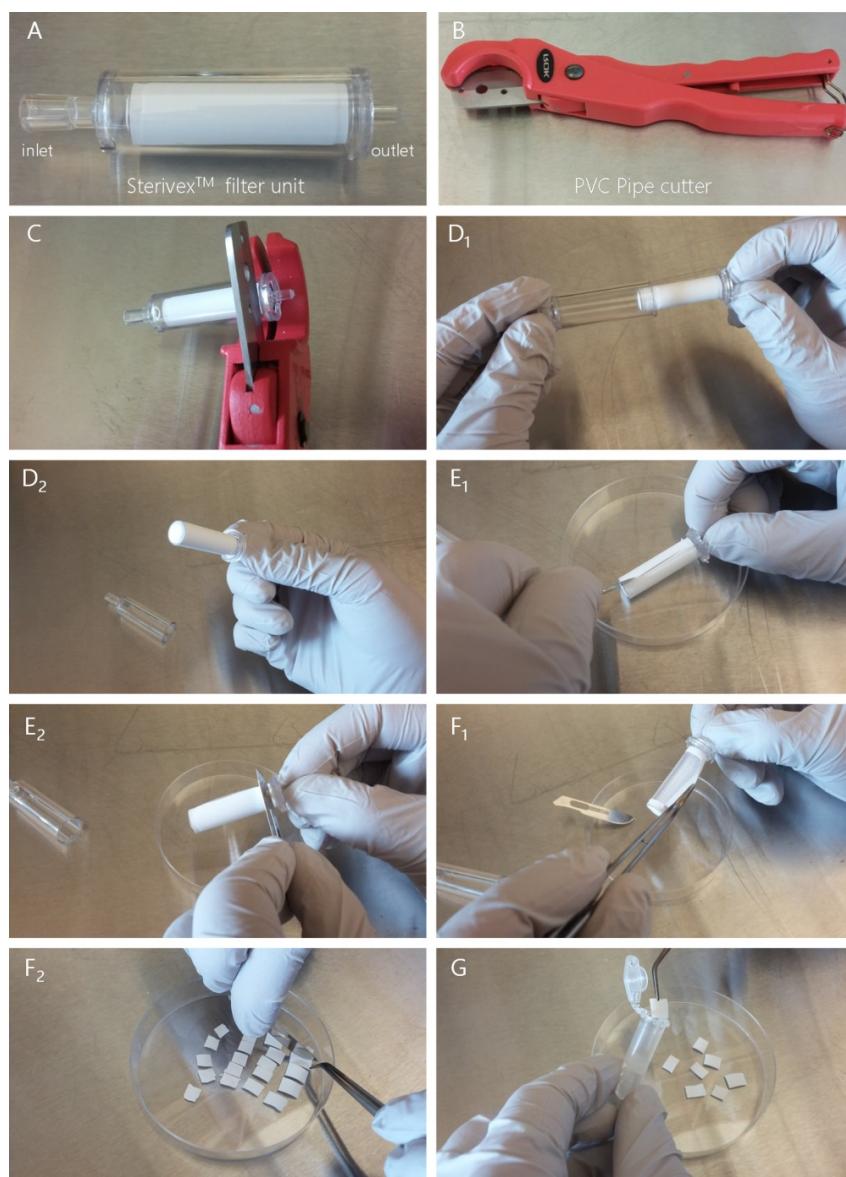
The flow chart on the following page includes some of our preferred protocols and recommended kits, links to the protocols are available at <https://ei16.short.gy/neofDNA>. Today we will be using a modified method for a commercial kit (QIAGEN Blood & Tissue).

Recommended DNA Extraction Methods



3.1 Removing the filter paper from the plastic unit

1. Clean the bench and equipment with 10% bleach.
2. Prepare a labelled **2 mL** tube with **400 µl Buffer ATL** and **10 µl proteinase K**. The lysis buffer ATL breaks open cell and nuclear membranes which exposes DNA to proteins which the enzyme proteinase K denatures. **This is where we create an extraction negative using the same reagents.**



(Craaud et al. 2017)

⚠ Wear safety glasses. ⚠ Hold filter at inlet end, away from the blade.

(A) Sterivex filter unit, showing inlet and outlet. **(B)** PVC pipe cutter.

(C) Cutting of the bottom of the Sterivex unit (outlet end) with the PVC pipe cutter.

(D1, D2) Removal of the filter from the casing.

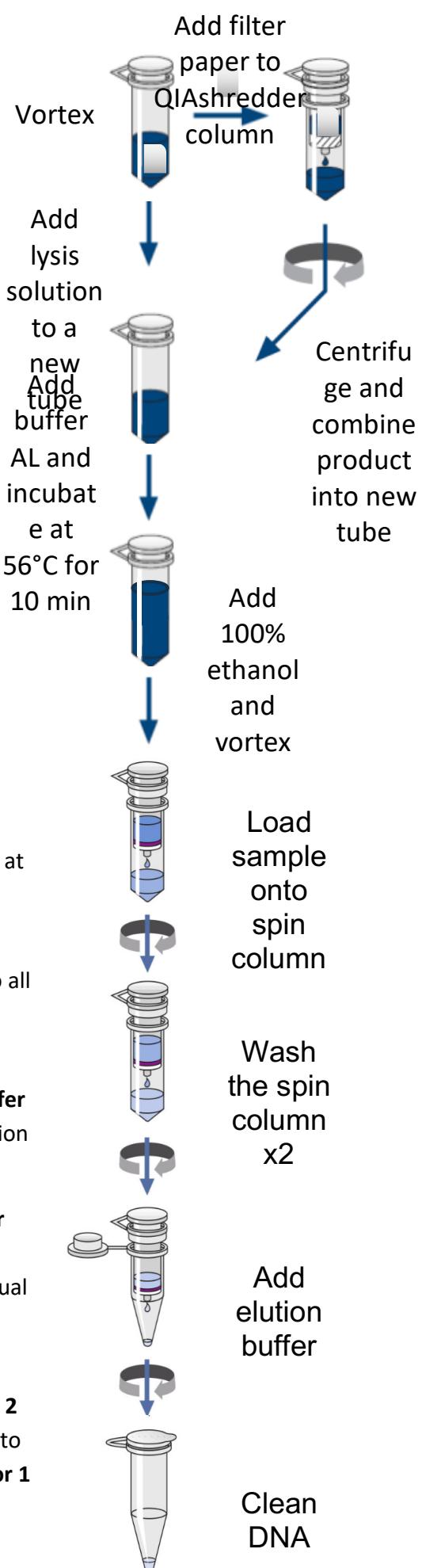
(E1, E2) Cutting of the filter attached to the inner part of the casing: first, longitudinal cut (E1) and second, transverse cut (E2).

(F1, F2) Detachment of the filter from the casing with sterilised forceps into a Petri dish and the dissection of the filter into small pieces with a sterile scalpel.

(G) Insert into the labelled microcentrifuge tube containing the lysis buffer. Incubate **overnight at 56°C**.

3.2 DNA extraction using a QIAGEN DNeasy Kit

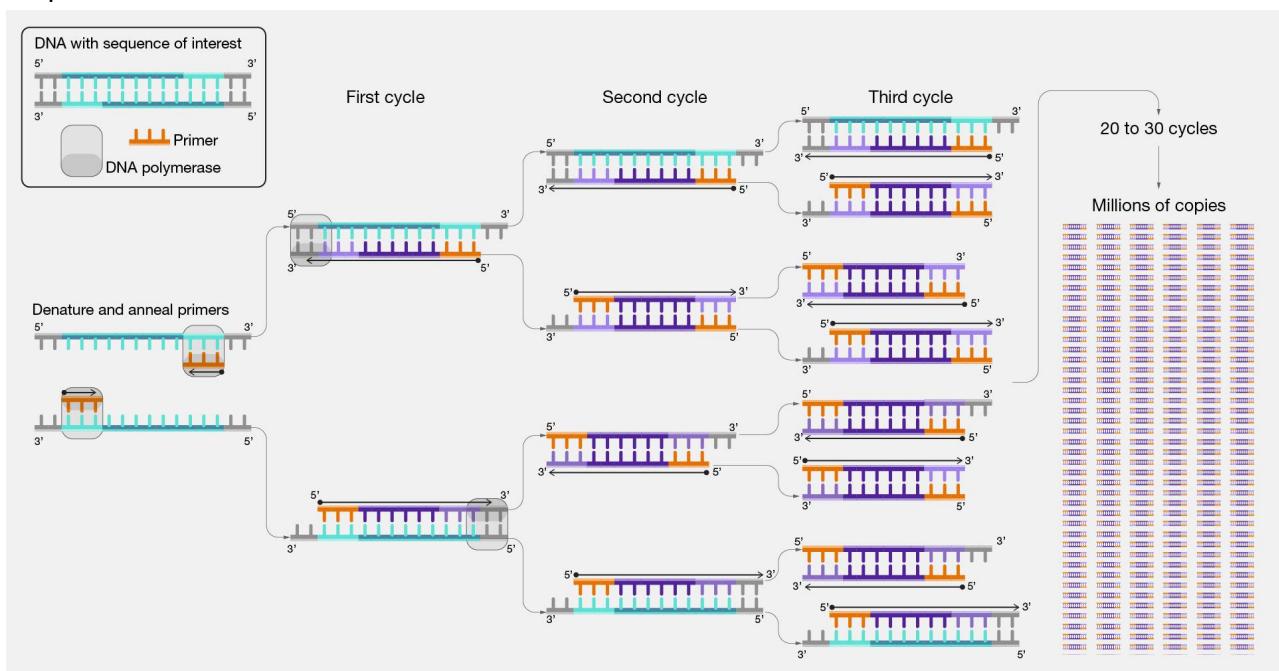
1. Clean the bench and equipment with 10% bleach.
2. Label a new **2 mL** tube.
3. Vortex the microfuge tube containing the filter sample for **15 seconds**. Pipette as much solution as possible into the new tube.
4. Place remaining filter paper in a QIAshredder spin column (lilac) using clean tweezers and centrifuge at **10,000rpm for 3 minutes**.
5. Combine the product (liquid that passes through the column to the collection tube) into the new tube with the rest of the solution. Discard the QIAshredder column and collection tube.
6. Quickly spin the original lysis tube down and pipette any remaining solution into the new tube with the rest of the solution.
7. Add **400 µl Buffer AL** to the 2 mL microfuge tubes with the lysis solution and vortex. Incubate at **56°C for at least 10 minutes**. This inactivates nucleases and promotes nucleic acid binding to the silica membrane.
8. Add **400 µl 100% ethanol** & vortex to precipitate the DNA.
9. Label a DNeasy spin column.
10. Add **700 µl of the sample solution** to a DNA spin column. Centrifuge at **10,000 rpm for 1 minute**. Material will pass through the silica membrane, which attracts DNA and allows debris to pass through.
11. Place the spin column in a new collection tube and repeat step 10 so all the solution has passed through the spin column. Discard the old collection tube and waste solution.
12. Place the spin column into a new collection tube and add **500 µl Buffer AW1**. Centrifuge at **10,000 rpm for 1 minute**. Discard the old collection tube and waste solution.
13. Place the spin column in a new collection tube and add **500 µl Buffer AW2**. Centrifuge at **13,000rpm for 3 minutes**. Discard the old collection tube and waste solution. These washes help remove residual impurities.
14. Place the spin column into a new labelled **1.5 mL** tube and add **50 µl Buffer AE** to the spin column and incubate at **room temperature for 2 minutes**. This rehydrates the membrane and DNA allowing the DNA to become unbound and elute into solution. Centrifuge at **8,000 rpm for 1 minute**. Discard the spin column.



4. PCR1 Amplification

1. **Denaturation** - The double-stranded DNA is separated into single strands when heated.
2. **Annealing** - The single-stranded primers bind to their complement in the target DNA when the temperature is lowered. The exact temperature depends on the annealing temperature of your primers.
3. **Extension** - The primers then direct the *Taq* polymerase to the sequence in your DNA sample. When the temperature is raised a copy of the DNA strand is made by the *Taq* polymerase enzyme. Extension time is influenced by the length of your amplicon.

The PCR process then replicates this sequence repeatedly over 20 to 30 cycles. The number of replicates grows exponentially, resulting in millions of exact copies of your genetic sequence, known as amplicons.



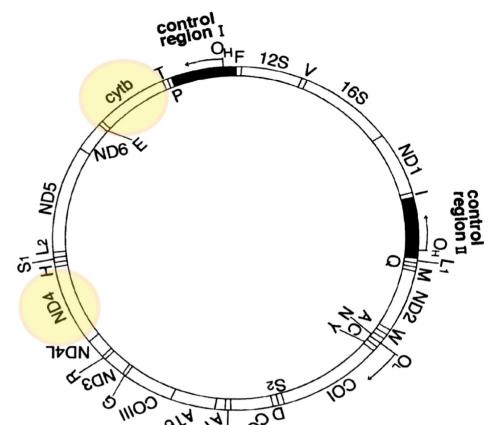
The Polymerase Chain Reaction is achieved in a thermocycler. We programme the thermocycler to control the timings and temperature conditions of the sample. These settings are tailored to the sequence you are replicating.

4.1 Primer design

Vertebrates (mtDNA):
COI
16S
12S
ND4
Cytb

Plants (chloroplast):
ITS
RBCL

Fungi - ITS
Bacteria - 16s or 18s



- **Gene choice**

It is vital to have reference sequences for every species potentially present to compare against.

- **Primer length**

We design primer sets to be as specific as possible to the target region, however there is always a chance that the primer will bind to similar sequences, replicating non-desired regions of the DNA. The specificity is generally controlled by the length of the primer and the annealing temperature of the PCR reaction. Shorter primers will be less specific and therefore more likely to target a wider range of genetic regions, or species. Longer primers will be more specific, but avoid making them so long that they fail to capture the range of individuals being sampled, which vary by their unique DNA mutations.

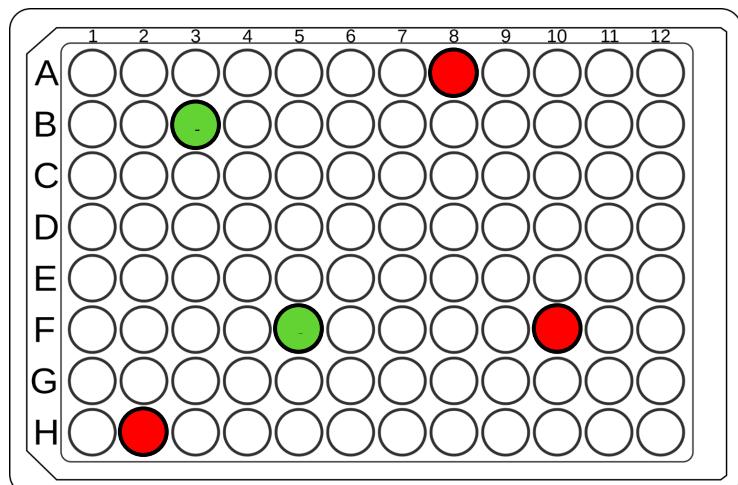
- **Amplicon length**

Our primers come in pairs, forward and reverse. They bind to either end of the DNA sequence of interest. This is the length of the DNA sequence being targeted for amplification, known as the amplicon. Shorter amplicons are better for the amplification of degraded DNA. Clearly though the shorter the amplicon the less sequence information present. Conversely, the longer the amplicon the more power you have to discriminate between species. This assumes an equal abundance of variants/mutations in a gene.

In silico software can help us in all stages of primer design, validating our primers against libraries of genetic sequence information.

4.2 Quality Control

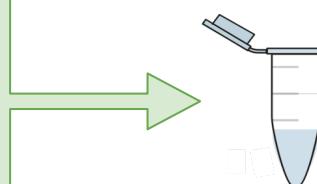
- Including a known **positive** will help determine if the programme ran as expected
 - Positives should be synthetic DNA or species not present in your targeted environment but that you have confirmed will be amplified by your PCR1 primer set.
- Including a **negative control** will highlight any contamination, which might influence all the samples in your run. Types of Negative
 - Field negative - A negative that has been through field processes.
 - DNA Extraction negative - A blank that is extracted at the same time as the samples.
 - PCR negative - A blank not extracted but simply loaded as a sample at the PCR stage.
- Positives and negatives should ideally be distributed at random across the plate and should be of reasonable number. This helps sample tracking and identifying contamination during analysis.



4.3 Setting up a PCR

1. Dilute your DNA sample by adding **1 µl** to a new labelled 1.5 mL tube and adding **49 µl ddH₂O**.
2. Label two 0.2 mL PCR tubes with your initials on the top and side.
3. Pipette up and down at least 10 times to mix.
4. Add the following reagents to a labelled 0.2 mL PCR tube:

- 2 µl diluted DNA sample
- 1 µl 16s Forward Primer
- 1 µl 16s Reverse Primer
- 2 µl ddH₂O
- 4 µl MyTaq HS Mix



5. Pipette up and down at least 10 times to mix.
6. Repeat steps 4 and 5 with 2 µl of ddH₂O instead of DNA to create a negative control and add 'neg' to the label.
7. Place into the thermocycling machine where it will follow the program:

95°C for 3 mins

x30 cycles of

94°C for 30 secs

50°C for 45 secs

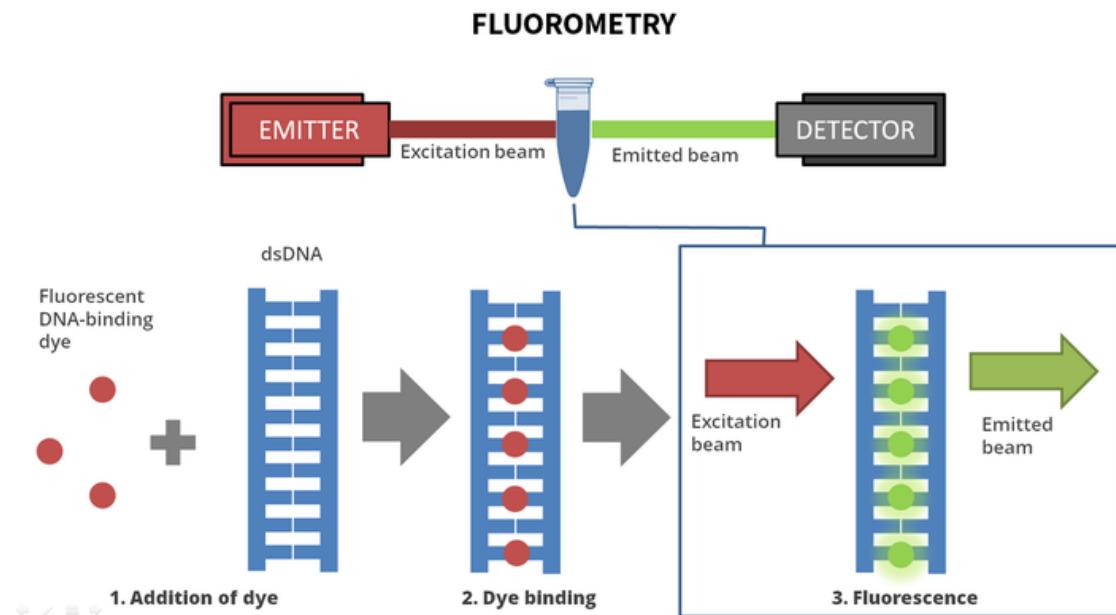
72°C for 45 secs

72°C for 10 mins

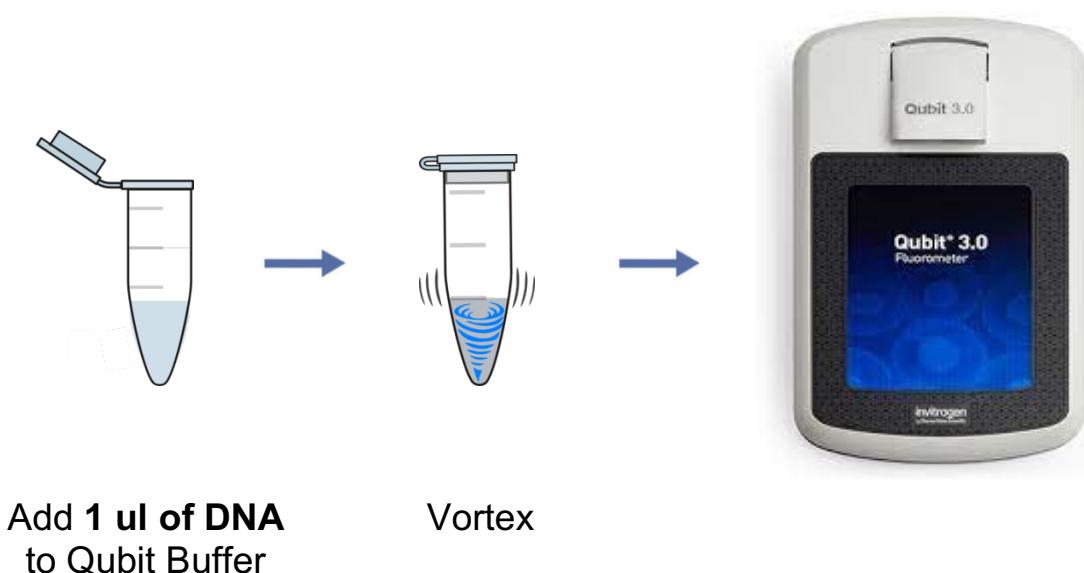


5. Quantification

Fluorometers utilise fluorescent dyes that bind to the DNA. Fluorescent dyes absorb a particular wavelength of light then emit light of another wavelength. The amount of fluorescence is directly proportional to the amount of DNA in the sample.



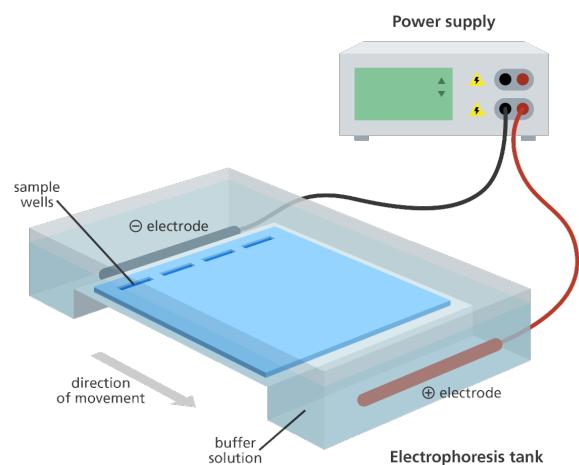
1. Add **1 μ L DNA sample** to the Qubit Buffer in 0.2 mL tube.
2. Close and vortex for **15 seconds**.
3. Place the tube on the Qubit Fluorometer and press “Read tube”.



6. Gel Electrophoresis

Running your PCR products on an agarose gel is a useful way to visualise your products post-amplification.

The PCR products are added to the gel in lanes, and drawn through the gel by an electrical current. As the phosphate backbone of the DNA molecule is negatively charged, the fragments will migrate to the positively charged anode (red). The shorter, and therefore the lightest, fragments will travel the quickest and the longer, heavier fragments will remain closest to the sample wells.

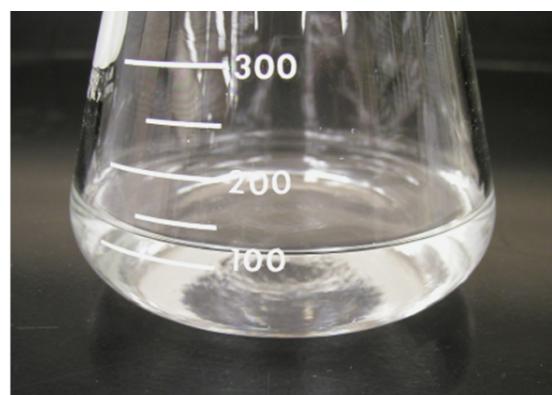
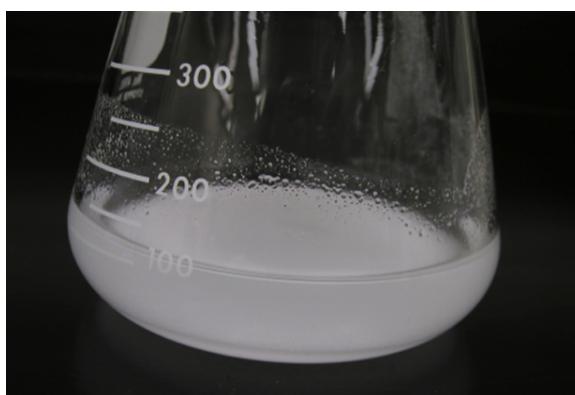


6.1 Making the Gel

1. Please work in pairs.
2. Use masking tape to seal off the ends of a plastic gel tray and place a comb at one end.
3. Weigh out **1 g of Agarose** and add to a 500 ml flask.
4. Use a measuring cylinder to pour out **100 ml of Water** and add to the 500 ml flask.
5. Heat the mixture in a microwave. Swirl the flask every 30 seconds at the beginning and every 10 seconds when it begins to bubble. The mixture will be completely clear when ready.

⚠ Wear heat proof gloves and take care as the mixture becomes superheated.

⚠ Always watch the flask so it does not bubble over.



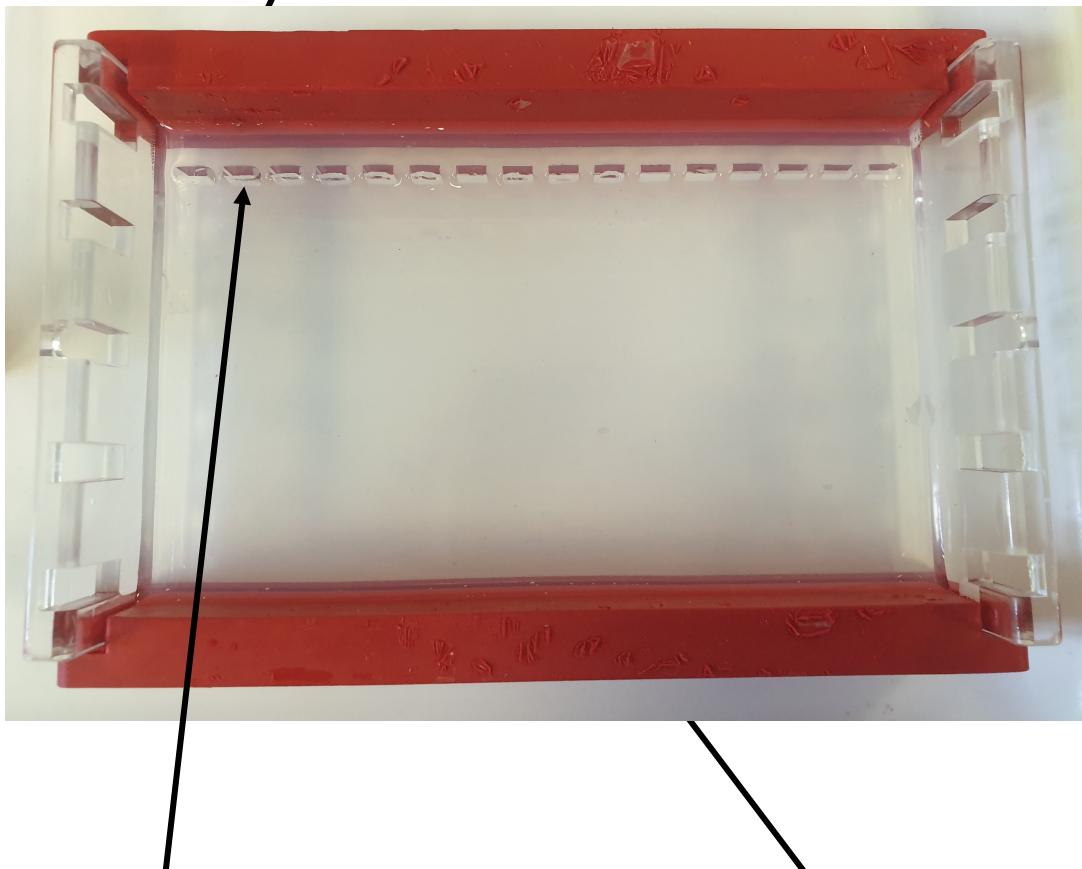
6. Cool under a tap whilst swirling the flask until warm to the touch (approximately 1-2 mins).
7. Add **4 µl of Gel Stain** and mix by swirling.
8. Pour gently into the prepared plastic gel tray.
9. If any bubbles are present, use a tip to either pop or drag to the side of the gel.

10. Leave to set which will take approximately 30 minutes.

6.2 Loading the Gel

Split into 4 groups. You are provided with a pre-made 1% agarose gel, 100bp ladder and a tube of orange G loading dye.

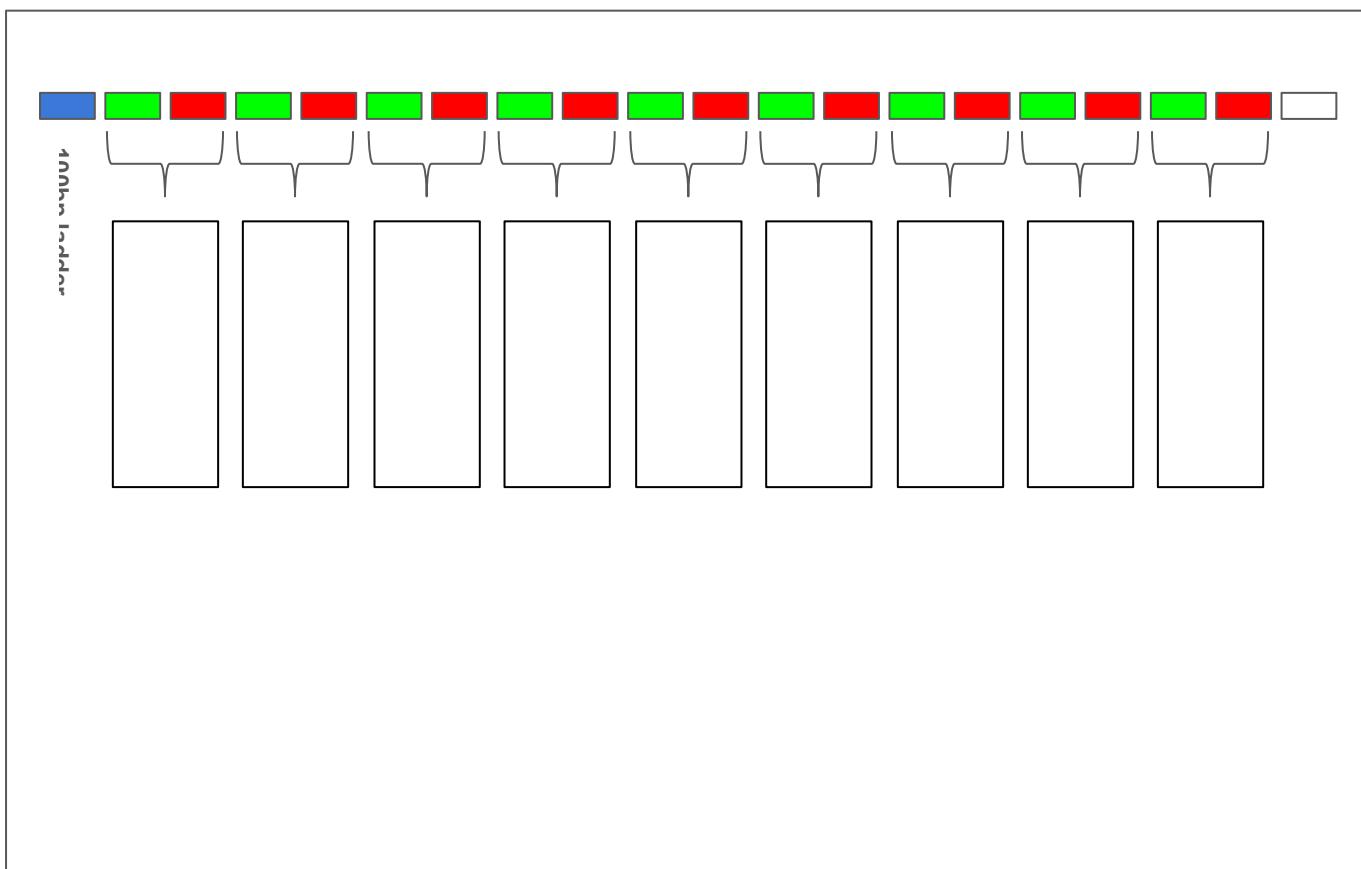
1. Pipette **4µl of your PCR products** into a new labelled 1.5ml tubes (one for the DNA sample and one for the PCR negative).
2. Add **6µl of orange G loading dye** to each and pipette up and down at least 10 times to mix.
3. One group member: remove red bumpers/tape and carefully lower the tray and gel into the tank.
4. One group member: Add **5µl of the 100bp ladder** to the first well, being careful not to pierce the gel.



5. Add **10µl of your PCR product mix** to an empty well. Add **10µl of your PCR negative mix** to the next well. Record which gel lane you have added your samples to on the next page.
6. Run at 120 volts for 40 minutes.

Gel Lane Map

Once you've loaded the gel with your samples, write your name in the box below to signify which gel lanes your PCR product has been loaded into.

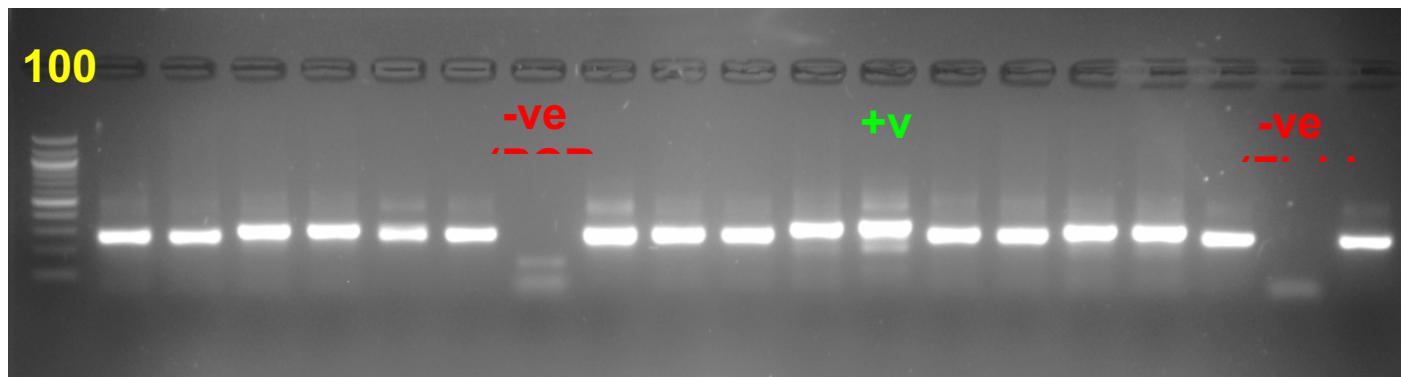


6.3. Visualising the Gel

As seen in the gel image below, DNA samples that successfully amplified during the PCR process show up as white bands.

By comparing the bands to a ‘DNA ladder’ which itself consists of known DNA fragment lengths, we can roughly determine the length of our PCR amplicons. In the example below a 100bp ladder was used, starting from the bottom there is a band every 100 bases with a brighter band at 500bp and 1000bp.

The intensity of the band is proportional to the quantity of DNA loaded into the gel. A lane without a band tells us no DNA was amplified, as in the negative control.



6.4 Genomics Laboratory Facility Tour

The Genomics Laboratory Facility has over 20 years experience in a wide range of genomic techniques including long & short read sequencing, SNP genotyping, qPCR, single-cell analysis and NGS library preparation.



It is also home to ISO 5 and ISO 7 inspired clean rooms. Both rooms use HEPA filters to remove airborne particles or contaminants larger than 0.3 µm.

The rooms have a positive air pressure system to prevent unfiltered air from entering the workspace and each room uses a UVC tower system to disinfect and inactivate any DNA or RNA contaminants present on surfaces or in aerosols.



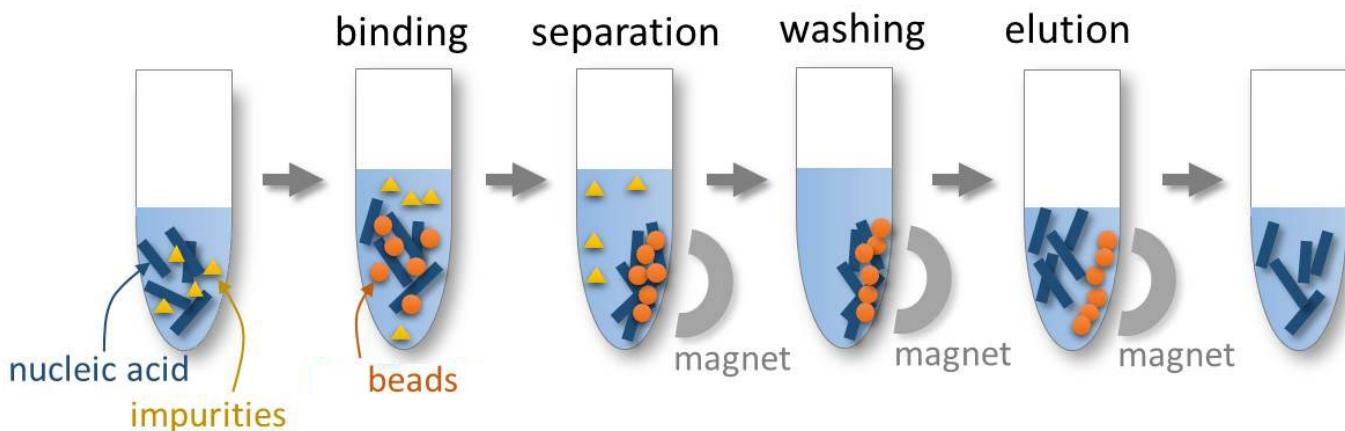
This high standard of clean room is essential for projects that use very low amounts of DNA or RNA, such as eDNA metabarcoding.

7. Bead Cleaning

After the PCR reaction, samples will consist not only of our desired amplicons, but also excess reagents and extraneous DNA fragments.

Bead cleaning allows us to remove waste products and size select our desired amplicons using a specific ratio of magnetic beads to PCR product.

The beads more readily bind to longer fragments of DNA, therefore we select a ratio that will capture our amplicons whilst removing shorter extraneous fragments.



1. Clean the bench and equipment with 10% bleach.
2. Add your remaining PCR product (~6 µl) to a new labelled 1.5 mL tube and add **24 µl ddH₂O**.
3. **Vortex and quickly spin down (1 sec)** the Pronex Beads.
4. Add **45 µl Pronex Beads** to the tube containing the PCR product, this will make a 1.5 x bead clean.
5. Mix well by pipetting up and down 10 times and incubate at **RT for at least 5 minutes**.
6. Place on a magnetic rack to separate beads from the solution.
7. When the liquid is completely clear (~1 minute), aspirate the supernatant and discard.
8. With the samples still on the magnetic rack, add **500 µl of Wash Buffer** and incubate at **RT for 30-60 seconds**. Discard the wash buffer.
9. Repeat step 8.
10. Allow the beads to dry at **RT for at least 5 minutes**, while on the magnetic rack.
11. Remove samples from the magnetic rack and elute with **25 µl of AE Buffer**.
12. Mix well by pipetting up and down 10 times and incubate at **RT for at least 5 minutes**.
13. Place on a magnetic rack to separate beads from the solution.
14. When the liquid is completely clear, aspirate the solution containing your DNA into a fresh tube.

8. PCR2 Indexing

During PCR1, our primers (Blue) incorporate an overhang adapter sequence (Orange) at the ends of our amplicons.

F:5'ACACTCTTCCCTACACGACGCTCTCCGATCTNNNNNGTGCAGCMGCCGGTAA3'

R:5'GTGACTGGAGTCAGACGTGCTCTCCGATCTGGACTACHVGGGTWTCTAAT3'

In PCR2 we attach IDT Unique Dual Index primers (Integrated DNA Technologies), also known as i5 and i7 identifier sequences, to these overhang sequences. These identifier sequences are added to each sample in a unique combination (Red).

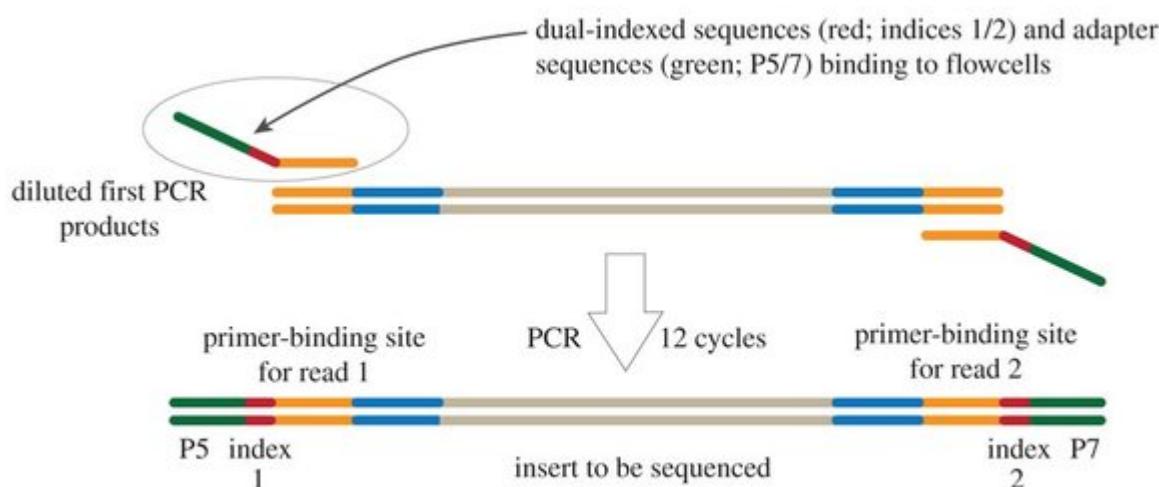
i5 index seq

5'AATGATAACGGCGACCACCGAGATCTACACATGCTTACTGACACTCTTCCCTACACGACGCTCTCCGATCT3'

i7 index seq

5'CAAGCAGAACGGCATACGAGATACTGGTCCGGTGACTGGAGTTAGACGTGTGCTCTCCGATCT3'

The final region of the unique identifier sequences is used in binding to the flowcell (Green) used in the Illumina sequencing. Once bound, all the tagged PCR products can be combined into one pool, and separated bioinformatically post-sequencing.



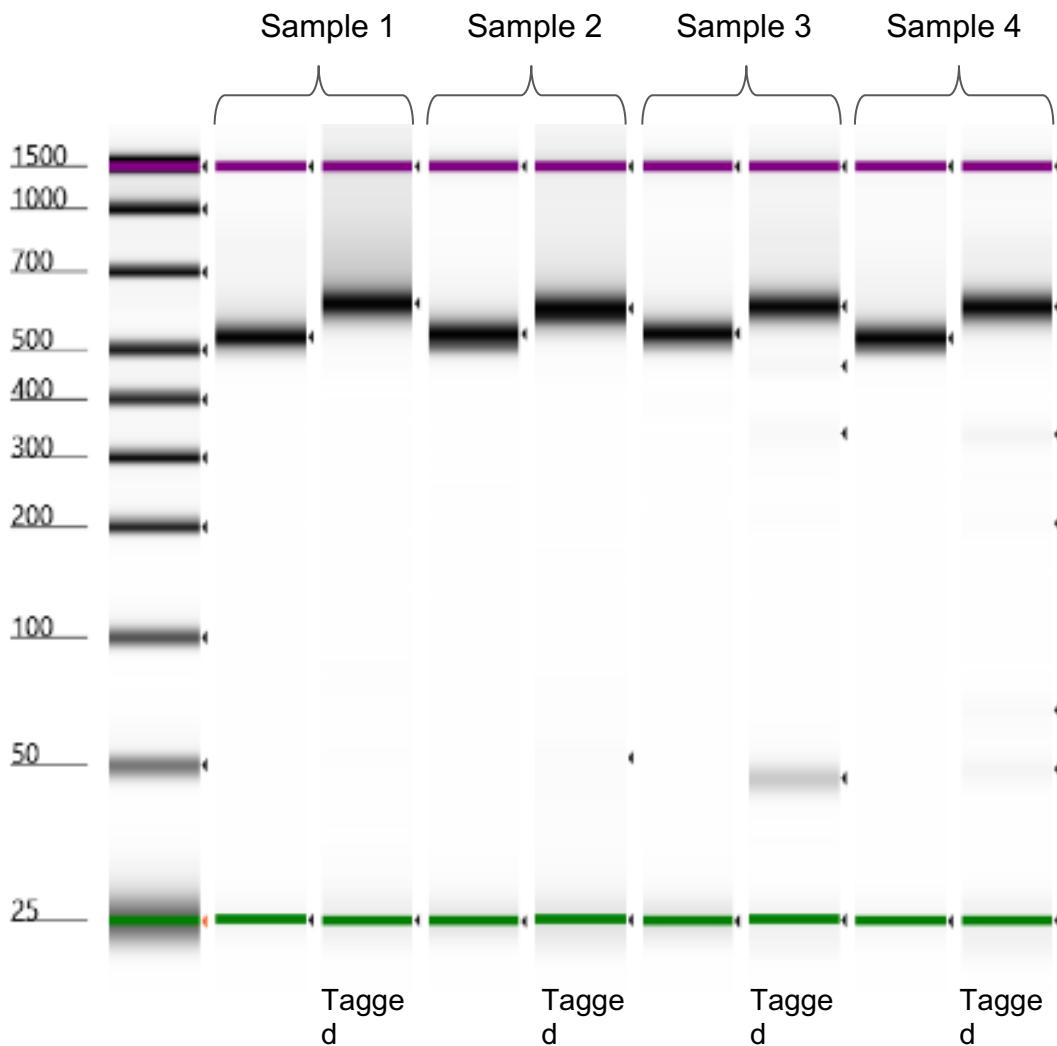
9. Size Determination

9.1 TapeStation

The TapeStation (Agilent Technologies) is a quick, automated electrophoresis platform for nucleic acid sample quality control.

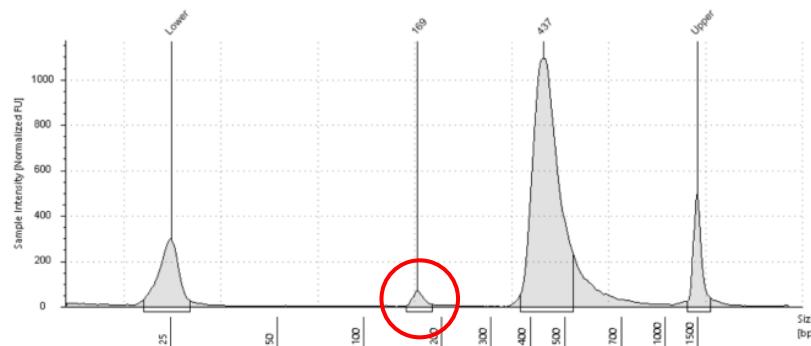
Samples are loaded in the top of the cassette and a voltage is supplied to them which causes the sample to diffuse through a gel matrix thus separating out the DNA by size.

In this case we use it to compare pre- and post- PCR2 products to determine if the unique identifier sequences have been successfully added to the amplicon fragments, and therefore increased in size.



9.2 BluePippin

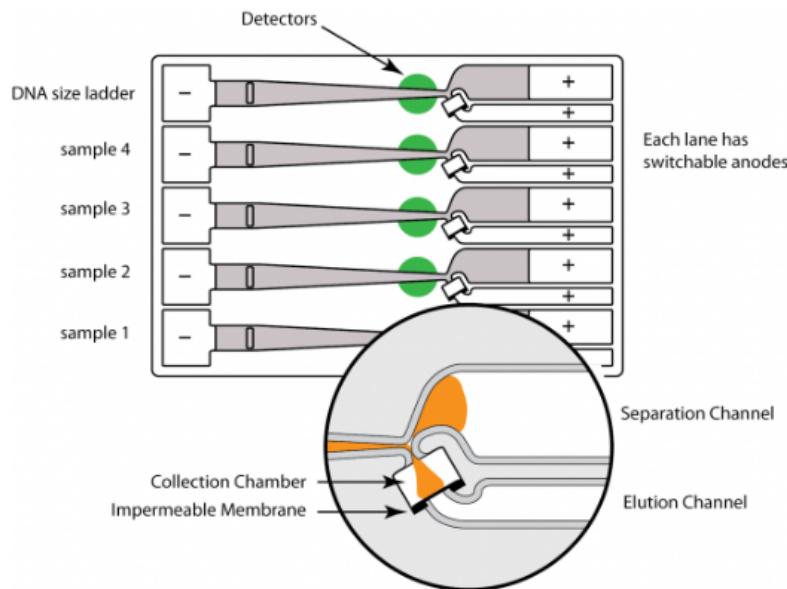
After PCR2 the samples require another clean up. Unfortunately, bead cleaning can't always remove all primer dimer or PCR artefacts. In this case we use the **BluePippin** (Sage Science) which is an automated DNA size-selection and cleanup system.



The BluePippin uses a disposable gel cassette that automates the separation and collection of DNA fragments based on size. It utilizes pulsed-field electrophoresis to resolve and collect DNA fragments ranging from 100 base pairs to 50 kilobases. The system uses pre-cast agarose gels and allows for the collection of up to five samples per cassette, with each sample lane physically separated to prevent contamination.



The system is programmed with specific size ranges that the user inputs. As DNA fragments migrate, they reach a branch point in the gel channel. The instrument then switches the active electrode to divert the DNA fragments within the desired size range into a separate buffer-filled well for collection.

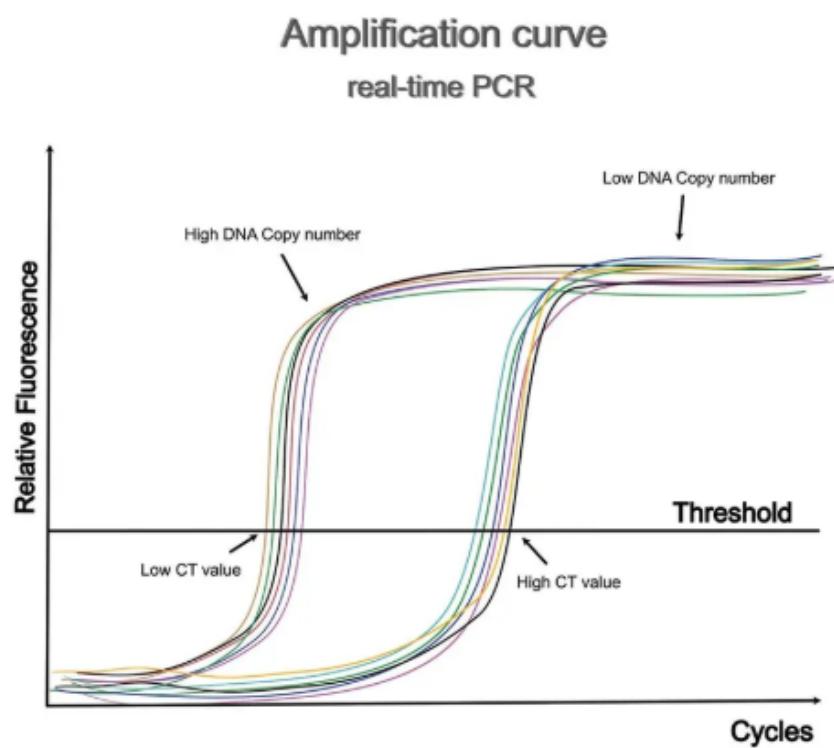
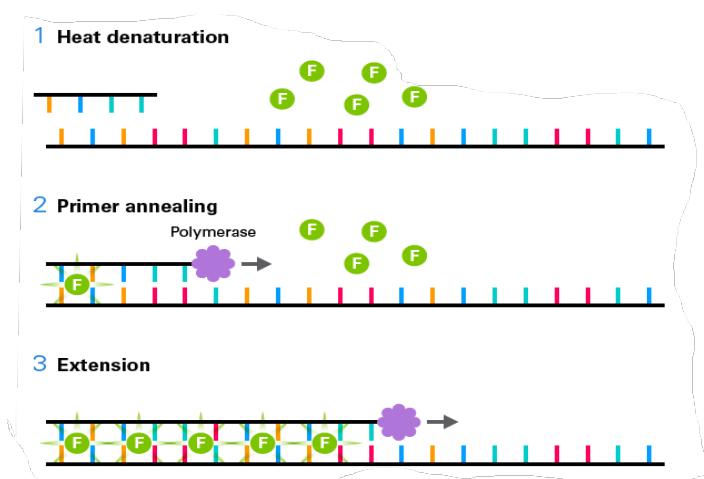


10. Library Quantification & Pooling

10.1 qPCR

SYBR Green quantitative PCR (qPCR) is a commonly used technique to quantify DNA in real-time. It uses a fluorescent dye, SYBR Green (F), which binds specifically to double-stranded DNA (dsDNA).

As the DNA is amplified during PCR, the fluorescence increases proportionally, allowing real-time measurement of DNA concentration.



The diagram above shows a typical graphical output from a qPCR run. The CT value refers to the cycle number at which the fluorescent signal of a target DNA sequence crosses the instrument's baseline threshold in real-time PCR, indicating the amount of target DNA present in the sample. A higher starting concentration of the target DNA results in fewer cycles required to reach this threshold and therefore a lower CT value. A lower starting concentration means more cycles are needed to reach the threshold and therefore has a higher CT value.

10.2 Quantification using qPCR

Due to the presence of unlabelled DNA from PCR2 we must quantify products by qPCR. Using a commercial kit we can quantify products for Illumina sequencing on the qPCR machine that only target DNA with i5 and i7 indexes (identifier sequences) attached.

Samples and standards are run in triplicate to obtain averages that should cancel out any minor deviations of pipetting error. From these results we can pool the PCR2 products together as each sample now has unique identifiers.



10.3 Pooling

Pooling helps normalise the coverage of reads per sample and reduces the number of samples needing to be quantified using qPCR. Using the formula below we can calculate the correct volumes of each sample so that each sample is evenly represented when pooling together.

$$\frac{(\text{Red Bar} \times \text{Purple Bar})}{(\text{Green Bar} \times \text{Cyan Bar})} = \text{Yellow Bar}$$

Sample ID	Final Volume (μl)	Initial Concentration ($\text{ng}/\mu\text{l}$)	Final Concentration ($\text{ng}/\mu\text{l}$)	Number of Samples	Volume to Pool (μl)
X01	50	6.90	4	6	4.83
X02	50	2.49	4	6	
X03	50	4.85	4	6	
X04	50	4.01	4	6	
X05	50	9.12	4	6	
X06	50	5.74	4	6	

Total Volume to Pool (μl) = _____

We then make up the pool with water to reach the final volume:

Volume of H₂O (μl) =

12. Next Generation Sequencing



Next-generation sequencing (NGS), also known as high-throughput sequencing, is the term used to describe a number of different modern sequencing technologies. These technologies allow for sequencing of DNA and RNA much more quickly and cheaply than the previously used Sanger sequencing (capillary electrophoresis).

There are a number of different companies that provide NGS platforms, the main being Illumina, Oxford Nanopore Technologies (ONT) and Pacific Biosystems (PacBio). For our short amplicon metabarcoding libraries, we usually sequence using an Illumina platform.



Platform	Maximum Read Length	Maximum Output	Estimated Cost (2024)*
iSeq 100	2 x 150bp	1.2 Gb	£600
MiniSeq	2 X 150bp	7.5 Gb	£700 - £1,600
MiSeq	2 x 300bp	15 Gb	£1,200 - £1,700
NovaSeq X	2 x 150bp	8 Tb	£2,300 - £3,800

*Based on quotes from companies in February 2024.

The iSeq, MiniSeq and NovaSeq only allow for up to 150bp paired end read, a 300bp total amplicon length. The MiSeq, however, is a bit more versatile as it is able to read 150bp, 250bp and 300bp paired end read.

We typically aim for ~50 bp overlap in order to stitch together forward and reverse reads. Sequencing larger products is possible, but stitching forward and reverse reads together will probably not be feasible, so this will need to be taken into account during project design. It's also worth bearing in mind that amplicons which have a wide range of product sizes from different species, such as ITS, if we use a stitched-read approach for analysis, this will exclude any species generating products >550 bp.



Nano and micro kits for some of these platforms are also available for doing small tests on which allows for sequencing a small amount of products to make sure your libraries are giving you what you expect.

12.1 Loading the Sequencer

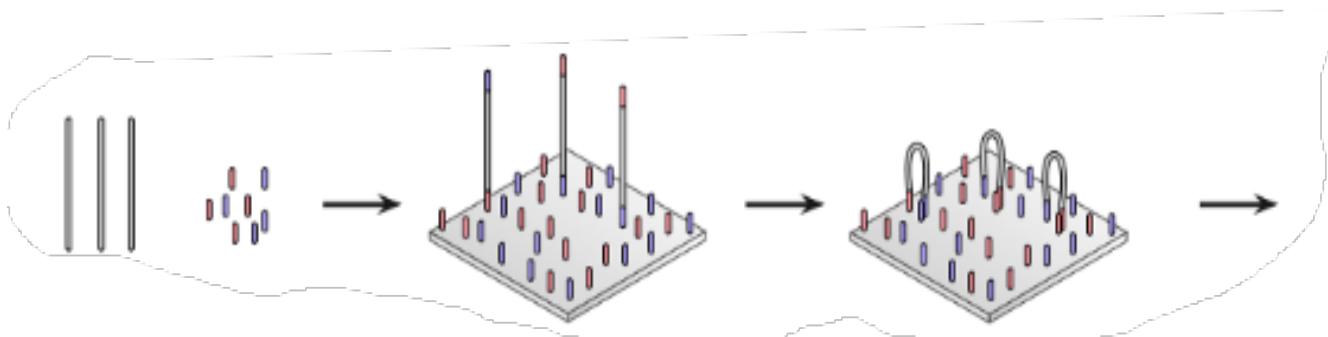
The amplicon library is first denatured and diluted with an Illumina buffer. It is then loaded into the cartridge which contains all the necessary components needed to perform the sequencing. The flow cell is also added either to the cartridge or the sequencing platform.



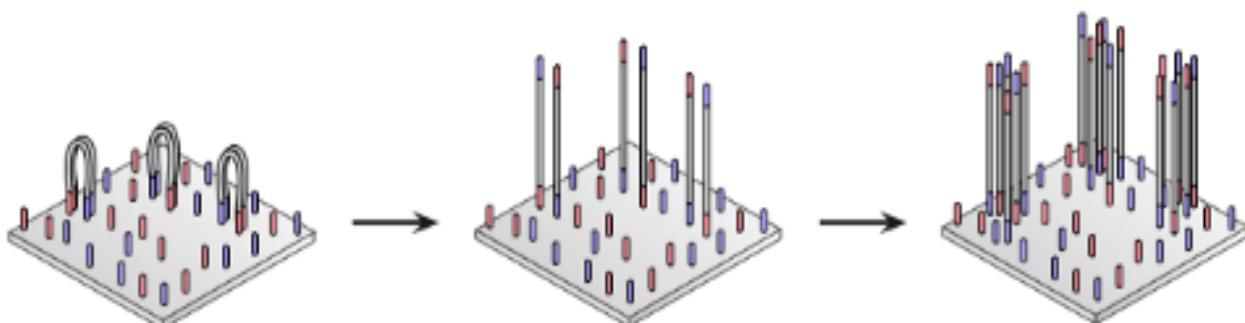
12.2 Sequencing by Synthesis (SBS)

Illumina's Sequencing by Synthesis (SBS) is a next-generation sequencing (NGS) technology that determines DNA sequences by detecting the sequential incorporation of fluorescently labeled nucleotides into a growing DNA strand. This process occurs in parallel across millions of DNA fragments attached to a solid surface, enabling high-throughput sequencing.

1st Stage - Cluster Generation (Bridge Amplification)



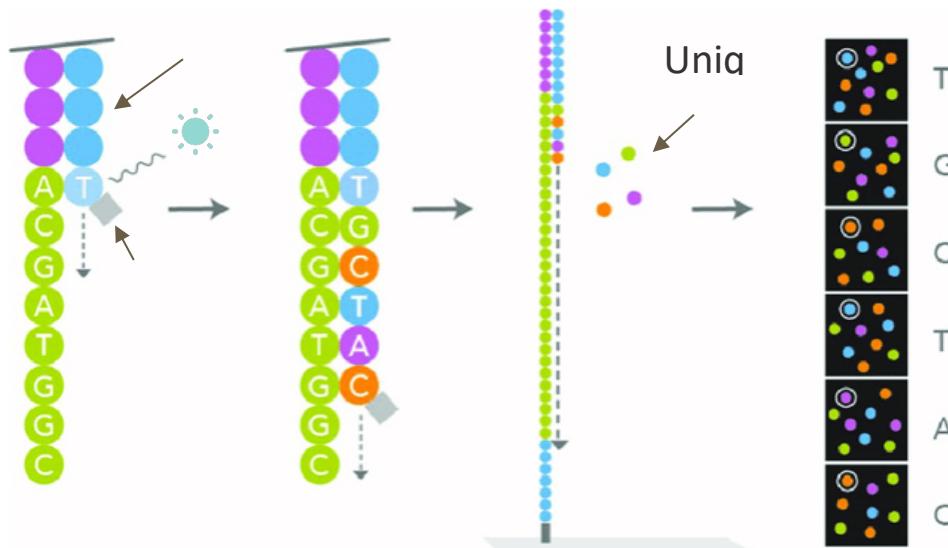
The glass flow cell is coated with two types of oligonucleotides and the DNA is denatured to form single strands. DNA fragments are then added to the flow cell and hybridize with one of the oligonucleotides on the flow cell surface. The bound fragments bend over and hybridize to nearby oligos.



DNA polymerase synthesizes the complementary strand. Denaturation forms two separate DNA

fragments. Repetition forms clusters of identical strands.

2nd Stage - Sequencing

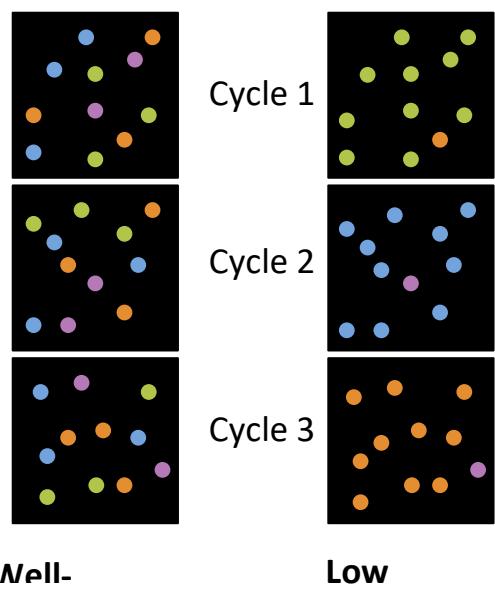


Fluorescently labeled, **reversible terminator nucleotides** (A, T, C, G) and DNA polymerase are added. One complementary nucleotide is incorporated per cycle. A camera detects the fluorescent signal which is unique to each nucleotide. No further elongation of the strand occurs as the incorporated base is blocked at its 3'OH. Unincorporated bases are washed away. Before a new cycle, the terminator and dye are cleaved allowing the incorporation of the next base.

12.3 PhiX

PhiX is bacteriophage DNA that is added to Illumina sequencing runs to provide nucleotide diversity, which is crucial for accurate cluster identification and data analysis, especially when working with libraries that have low base diversity. Libraries will have low diversity if only one or two different primers are used in the library.

If only one primer is used to make the library the beginning of every run would be identical which would overwhelm the nucleotide detection process during sequencing and could result in the sequencing failing. 5% to 20% PhiX is added to the amplicon library depending on the level of diversity.



12.4 Data output

The sequence data (reads) are generated on the sequencing platform. An index file is used to demultiplex the sequence data. The index file contains the sample information such as sample name and i5 and i7 indexes associated with that sample. Output is in the form of FASTQ files containing the sequence and quality scores. Sequences go through a bioinformatic pipeline which cleans the sequences and then identifies the sequences by comparing to known sequences deposited in databases.

12.5 Previous eDNA Study

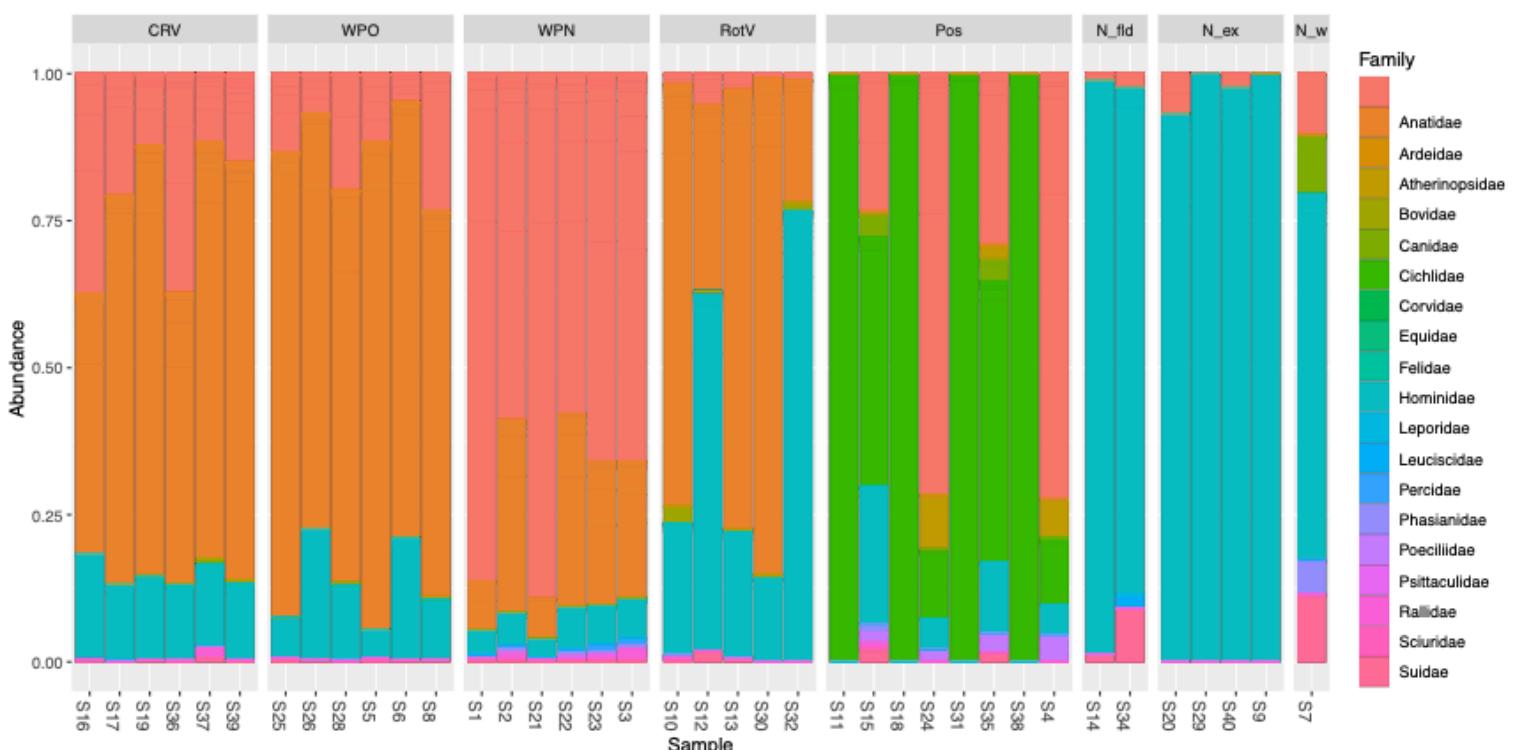
In Autumn 2022 we carried out an eDNA metabarcoding experiment to test what species can be found in three local ponds, including Weston Park which was sampled on Monday.

Samples:

- Weston Park (**WPO/WPN**)
- Rother Valley (**RotV**)
- Crookes Valley (**CRV**)
- Positive controls - Mexican river samples, African Cichlid stomachs (**Pos**)
- Negatives
 - Field (**N_fld**)
 - Extraction (**N_ex**)
 - PCR (**N_w**)



The “Riaz 12S-V5” primer set (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3241669/>) was used to target vertebrates.



Glossary

- **Agarose** - A polysaccharide, which when dissolved in a buffer can set into a 3D gel. The gel consists of tiny pores that allow DNA fragments to pass through when an electrical current is present
- **Amplicons** - The source / product of amplification. Interchangeable with PCR product within a molecular laboratory
- **Amplification** - Replicating a DNA sequence multiple times; a key element of PCR
- **Annealing** - When complementary strands of DNA bind to each other e.g. when a primer binds to a single stranded DNA molecule
- **Aspirate** - Draw up liquid
- **Barcode library** - Genetic sequences that are highly conserved within a species, but are variable between species
- **Bioinformatics** - Software tools designed to interpret large amounts of biological data e.g. data generated by genetic sequencing
- **Blocking primer** - Primers that preferentially bind to the DNA of which amplification is to be avoided, e.g. the host DNA in a diet study
- **Buffer** - A solution where the pH does not change significantly, used to preserve the stability and activity of nucleic acids
- **Centrifuge** - Spins tubes (typically containing liquids) so that the contents under centrifugal forces migrate to the base of the tube. Also a way of separating constituent elements in solution by mass. We might say 'spin down' when referring to the centrifuge
- **Conserved** - Identical or similar sequences in nucleic acids or proteins across species, or within a genome. The most highly conserved genes are those that can be found in all organisms
- **Denature** - Break apart, especially of a double stranded DNA molecule to two single stranded DNA molecules
- **eDNA** - Environmental DNA, collected from environmental sample types such as water, soil and air, rather than directly from the tissue of an organism
- **Elution** - Dissolving DNA into suspension
- **Flow cell** - An Illumina flow cell is a hollow glass slide with one or more channels ("lanes"), coated with short nucleotide sequences that are complementary to the sequencing adapters ligated to your (single-stranded) target DNA, allowing attachment through hybridization. Sequencing takes place on the flow cell
- **Fluorometry** - Dye based method for determining DNA concentration in solution
- **Fragment** - A length of DNA sequence smaller than the 'in-tact' genomic DNA contained within a cell. DNA can be fragmented by enzymes or by environmental actions e.g. UV, vibration
- **Gel electrophoresis** - Uses an electrical current to separate DNA fragments by size within an agarose gel submerged in a buffer. Useful for determining the length of your DNA / amplicon in base pairs, and as a quality control method for determining the integrity of your DNA samples
- **Gene** - A defined genetic sequence that codes for a particular protein
- **Genome** - The entire set of DNA instructions found in a cell
- **Illumina** - Manufacturer of Illumina next generation sequencing platforms. Produce sequencing equipment based on a sequencing-by-synthesis approach

- **Ladder** - A product containing a range of DNA fragment lengths of known size. A ladder can be run alongside DNA / PCR products during electrophoresis. You can estimate the length of your DNA fragments by comparing their position to those of the ladder's bands using the manufacturer's key
- **Ligation** - The joining of two DNA molecules using the enzyme DNA ligase
- **Lysis solution** - Used to break open cells in order to release the DNA contents
- **Mutation** - A change in the expected sequence of nucleotides. Types of mutation include point mutations (substitution, insertion, deletion), and chromosomal mutations (deletion, duplication, inversion, and translocation)
- **Normalisation** - In our case, normalisation refers to bringing our DNA samples to the same concentration, typically through dilution
- **Nuclease** - Enzyme that degrades nucleic acids e.g. DNases and RNases
- **PCR** - Polymerase Chain Reaction, the amplification of a target sequence to a scale that can be used for downstream sequencing
- **PCR product** - Once DNA has been through the polymerase chain reaction, we refer to it as PCR product, because it has been altered in some way, typically through amplification of specific gene region to the point where it outweighs all other genetic content
- **PhiX** - Control DNA with diverse base composition, added to Illumina sequencing runs, providing the balanced fluorescent signals that low diversity samples lack. Aids cluster identification and QC calculations. It's derived from the small, well-characterised bacteriophage genome, PhiX
- **Phylogeny** - Study of the evolutionary history and relationships among or within groups of organisms
- **Primer** - A short nucleic acid sequence that provides a starting point for DNA synthesis.
- **Qubit** - A model of fluorometer
- “**RT**” - room temperature
- **Reads (sequencing)** - The sequence of a single DNA fragment.
- **Reference sequence** - A representative example of a specific sequence, based on previously submitted data
- **Sequence** - Particular order of nucleic acids in the DNA strand, which may or may not code for amino acids
- **Sequencing** - Determining the order of the nucleic acid bases in a genetic sequence
- **Single Nucleotide Polymorphisms (SNP)** - Variant at a single base position in the DNA sequence. Locations in the genome known to have variable bases when compared to other individuals and populations of the same species. SNPs are spread across the genome. SNPs may or may not alter the protein coded by the gene in which they're found
- “**Spin down**” - Centrifuge
- **Supernatant** - Liquid phase of sample, post-centrifugation
- **Taxonomic** - Naming, defining and classifying groups of biological organisms based on shared characteristics
- **Thermocycler** - Equipment typically used to carry out PCR, by cycling through a preset programme of temperatures and timings
- **Vortex** - Small lab equipment used to mix liquids rapidly

Metabarcoding eDNA Course Risk Assessment:



Health and Safety in the Field - eDNA workshop

The first part of the course takes us to **Weston park**, where we will be carrying out water **sampling of the pond**. Please be prepared to approach the edge of the pond by wearing sturdy footwear, and be aware of your surroundings. Please give due consideration to the uneven/slippery nature of the ground, members of the public and their pets, and potential exposure to the sun. Follow the instructions of the course leaders.

Health and Safety in the Laboratory - eDNA workshop

PPE is provided in the lab to protect you from all elements of the course deemed hazardous or harmful to health. As such, **please wear the lab coat and nitrile gloves provided at all times when in the lab**.

Additionally, **goggles** are provided to protect your eyes when cracking open water filters. Full instructions will be provided by the course leaders throughout the course.

Finally, **eating and drinking is not permitted** in the lab at any time. Refreshments are available periodically throughout the day, or feel free to step outside of the lab to take a drink.

Precautionary advice

Spillages and splashes

- Wash spillages with water and soak up immediately
- Splashes in the eye
 - Flush the affected eye with running water for at least 10 minutes. Gently hold the eyelids apart to ensure the water bathes the eyeball.
 - Always seek medical advice. Advise medical staff which chemical is responsible for the eye injury.
- Waste disposal - wash liquids down the sink with copious amounts of water
- Blue roll disposed of in hazardous waste bins
- Examples
 - 10% bleach - Used to destroy bacterial and DNA contamination. Reacts with alcohols to form chloroform - do not mix.
 - Buffers and enzymes - DNA extraction kit
 - QuantiFluor® dsDNA Dye - fluorometry

Using sharp blades to cut filter paper

- A needlestick or ‘sharps’ injury is an incident, which causes a hypodermic needle, blade (such as scalpel) or other sharp instrument to penetrate the skin.
- Immediate first aid:
 - a. Encourage the wound to gently bleed, ideally under running water – do not suck the wound.
 - b. Wash the site liberally with water and plenty of soap (do not scrub the wound).
 - c. Dry the wound and cover it with a waterproof plaster or dressing.
 - d. For eye/face splash injuries, irrigate the affected area with large quantities of water.
 - e. Report the incident without delay to your course leader.
- Precautionary measures - hold filter at inlet end

Use of laboratory equipment

- All electrical equipment is PAT tested to meet safety standards
- Do not use equipment without supervision from course leaders

Burns

- Caution must be taken when working in the vicinity of the heating blocks, which contain a hot surface.
- Heat resistant gloves are provided when boiling agarose solution in the microwave
- Burns must be placed under cold running water for 10 minutes
- In the case of a fire, alert the course leader and evacuate the building, following emergency exit signs

Making gels

- Wear heat proof gloves and goggles - mixture can become superheated, but regular gentle swirling will mitigate this
- Always point beaker away from yourself and other people

Use of centrifuge

- Centrifuges operate at high speeds and must be balanced to prevent them from becoming unstable
- This requires an equal volume (mass) being loaded in the opposing position to the sample being centrifuged
- Unstable centrifuges must be switched off at the plug where safe to do so

Safety for pregnant people

The majority of the course will be safe for pregnant people to attend. Only one element of the lab work (running agarose gels) is seen to pose a minimal risk to the unborn child, due to the use of TBE and ethidium bromide. We would ask anyone pregnant to observe this step from a distance. Appropriate PPE (lab coat and gloves) will be provided. These mitigating factors are sufficient to make attending the course safe for pregnant people and their unborn children.

Notes