



# Metabarcoding and Environmental DNA Workshop:

*Assessing Biodiversity of a Local Pond using DNA Metabarcoding*



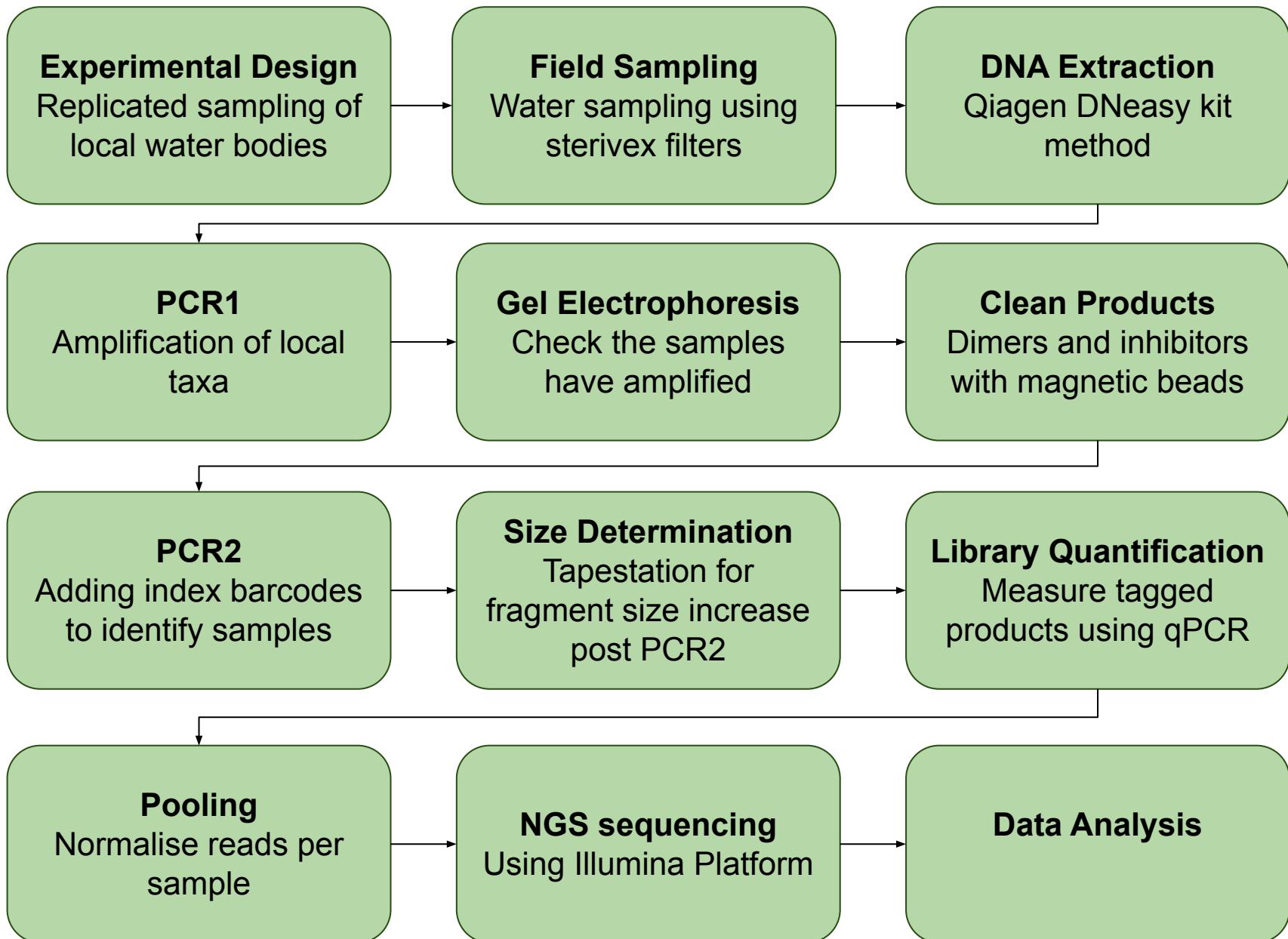
The University  
Of Sheffield.

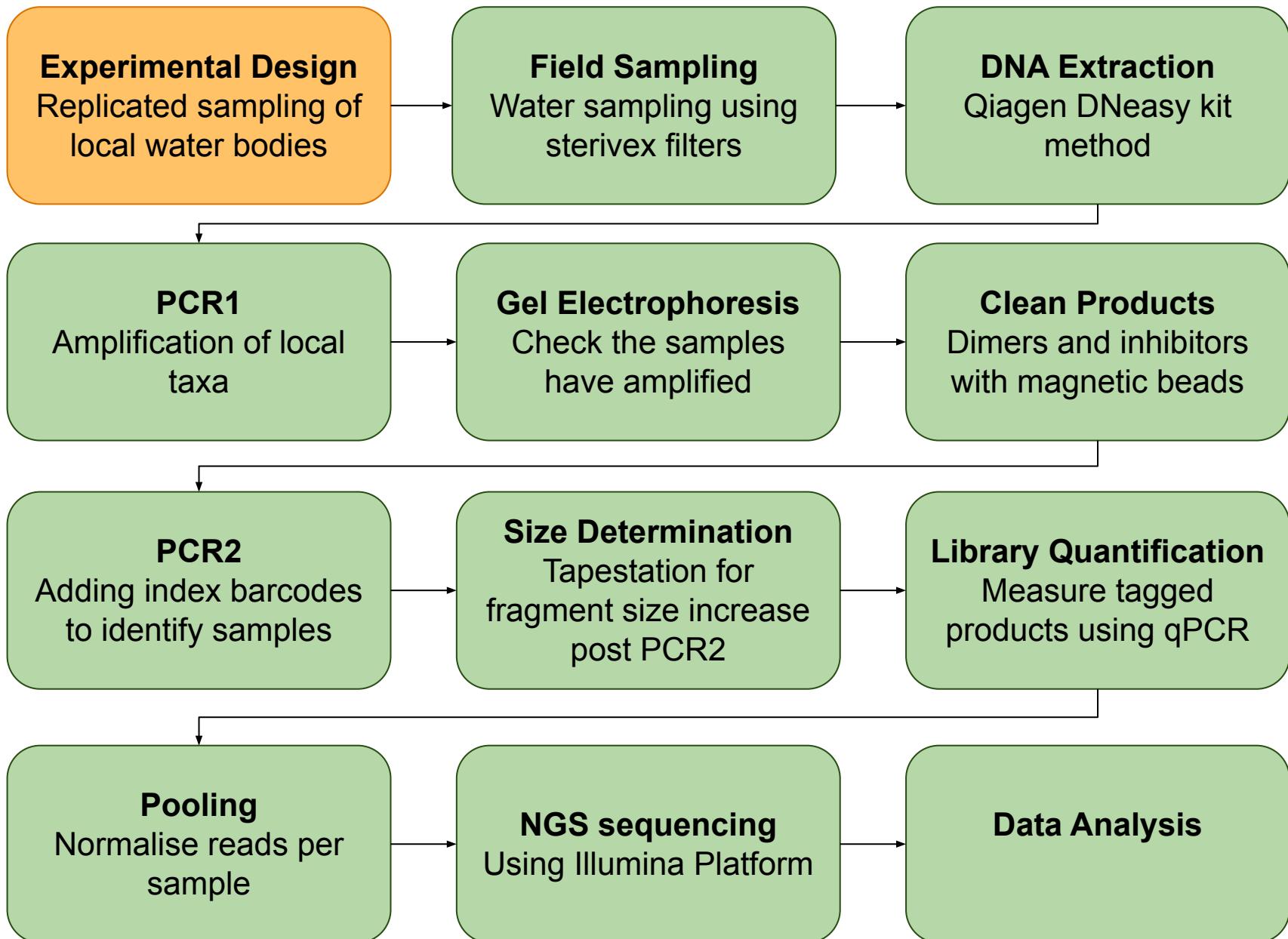


Natural  
Environment  
Research Council



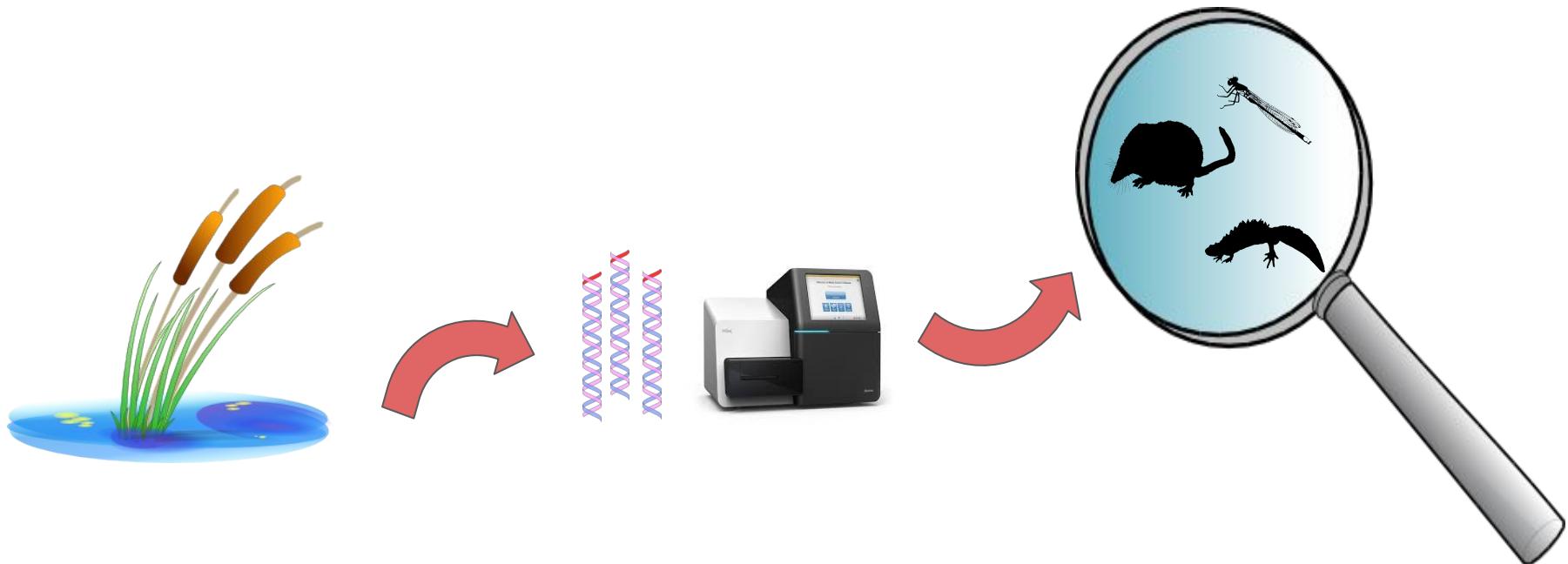
**NEOF**  
NERC ENVIRONMENTAL  
OMICS FACILITY





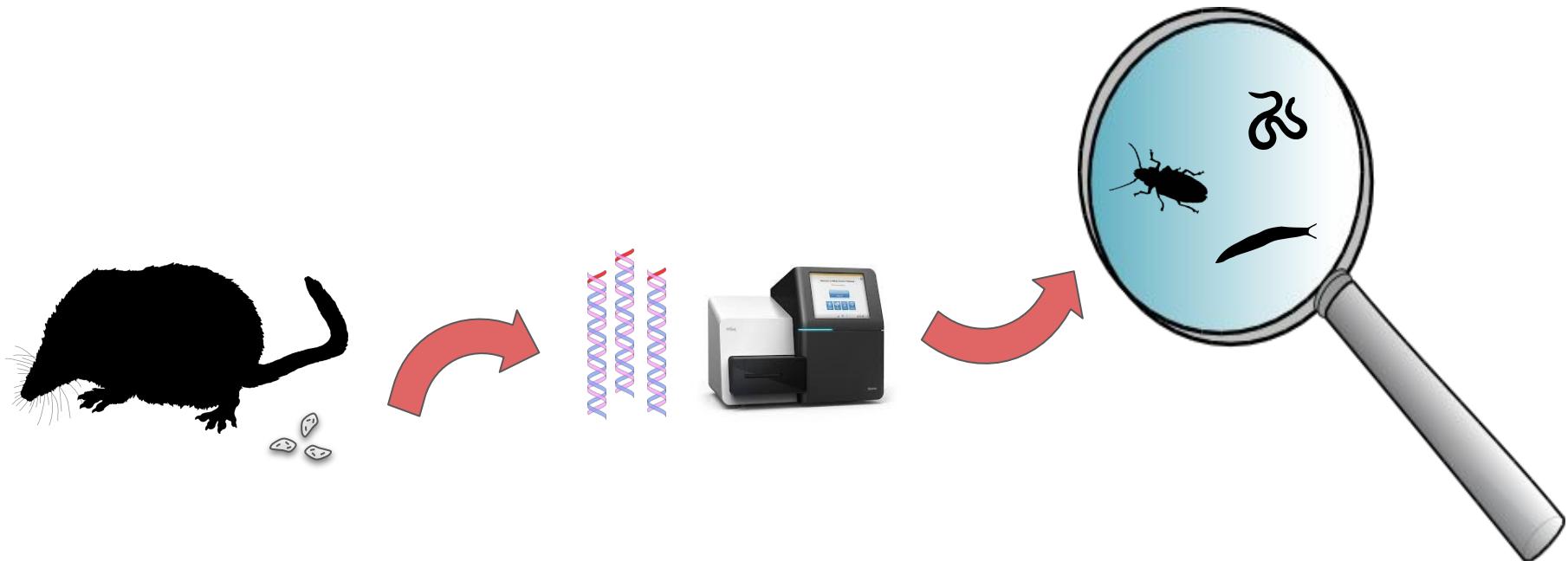
# Why metabarcoding?

Identify a range of species in an environmental sample simultaneously, through genetic sequencing.



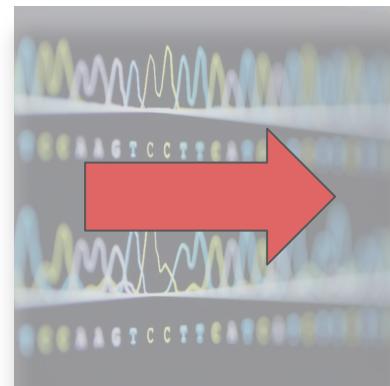
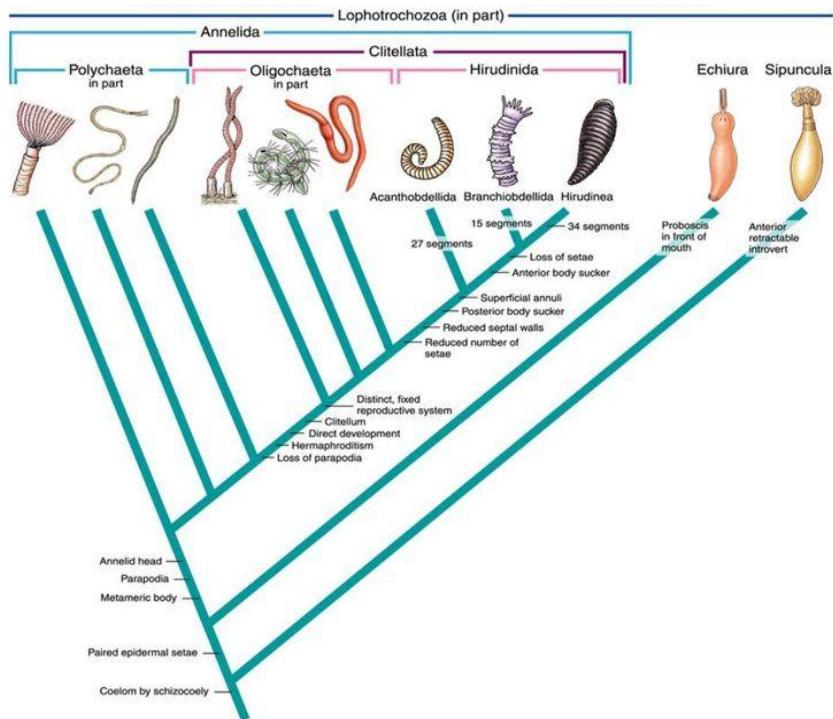
# Why metabarcoding?

This allows scientists to carry out a range of studies including diet analysis and population surveys



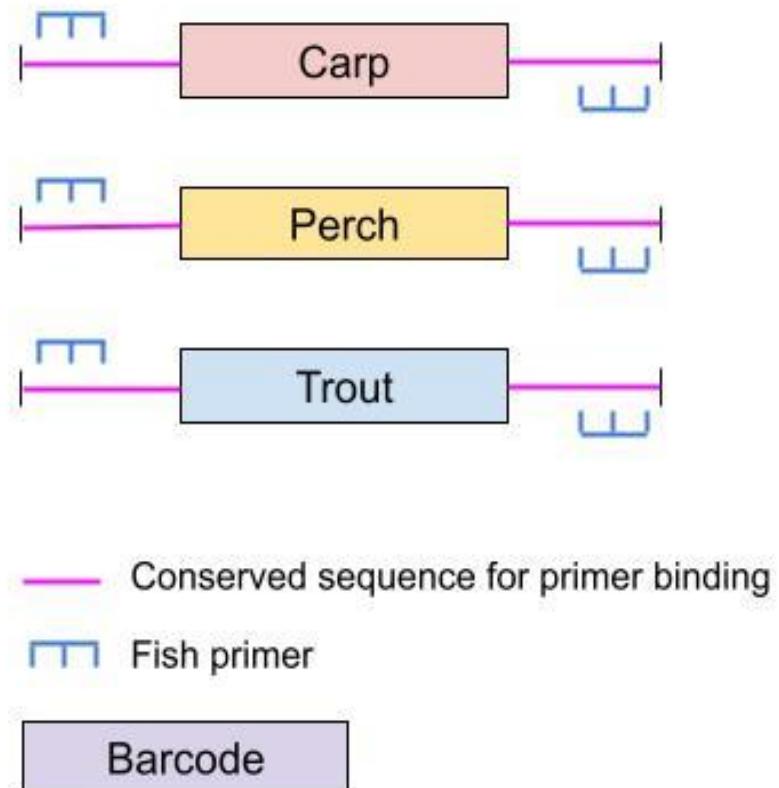
# How does metabarcoding work?

Generic primers are used to target whole taxonomic groups, whilst bioinformatic analysis allows species identification

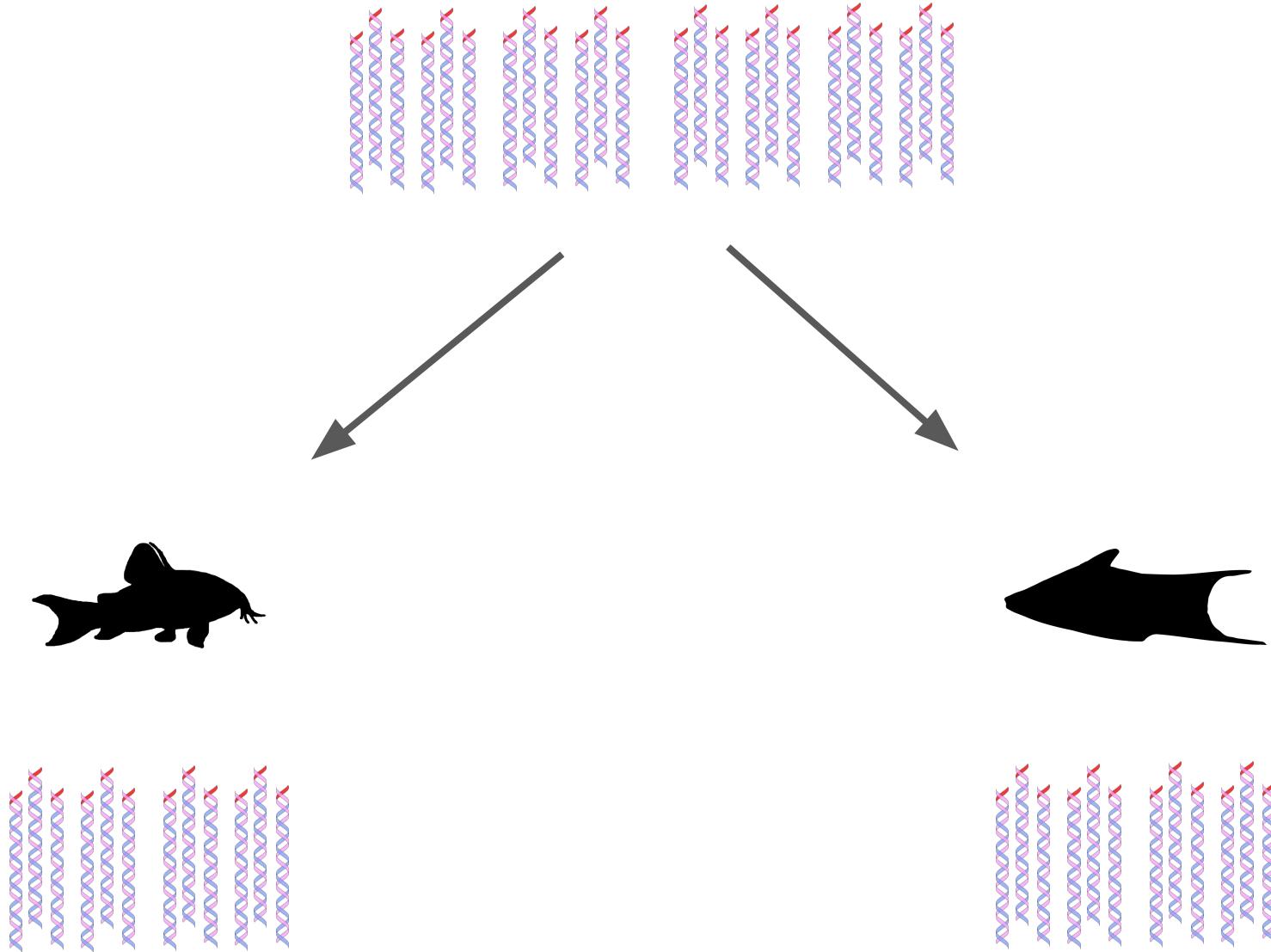


# Metabarcoding genes allow species identification

- Barcoding genes are highly conserved within a species, but variable between species
- The barcoding gene should be flanked by sequences that are conserved within the target phylogeny
- Primers are designed to target the conserved sequences
- Resulting sequence data can be compared to existing barcode libraries for species identification

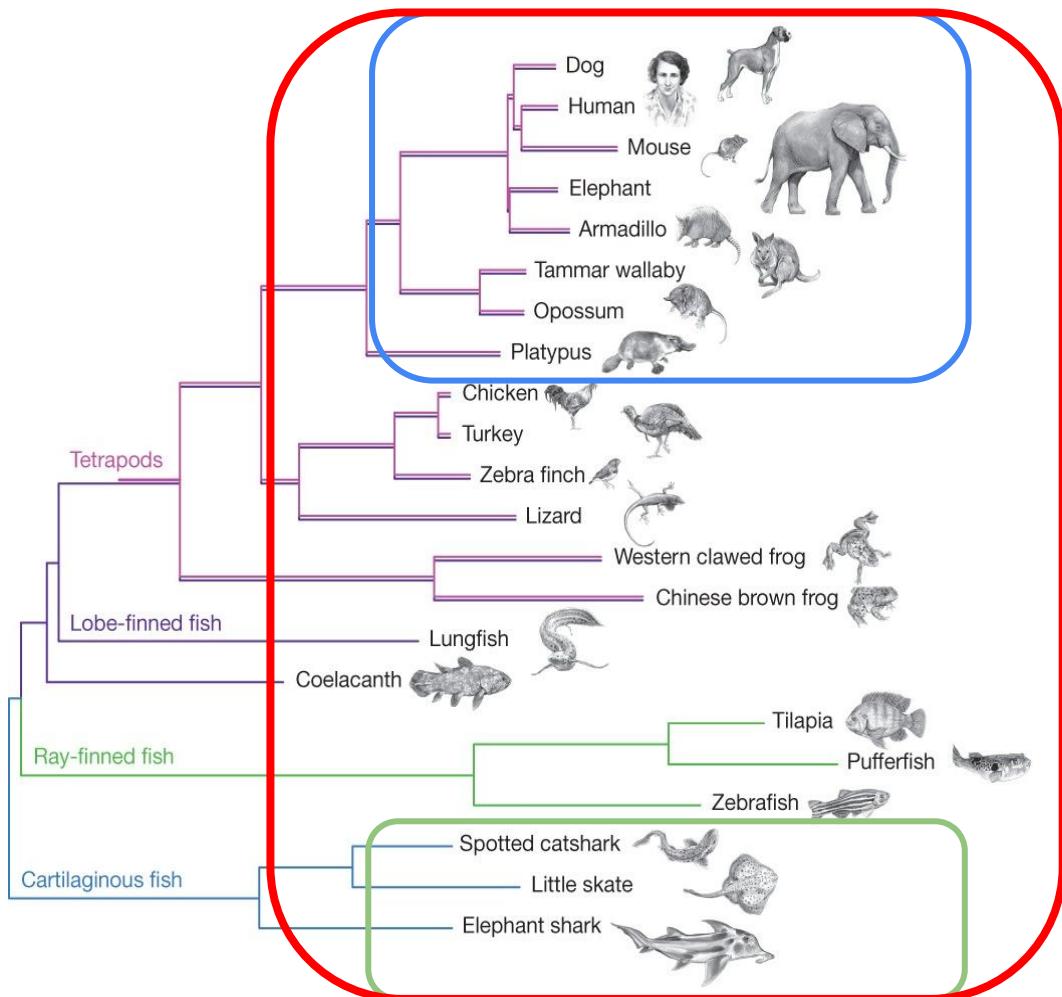


# Total number of sequencing reads to be shared between target species





# PCR 1 - Primer Selection



LCO1490

MiMammal-U

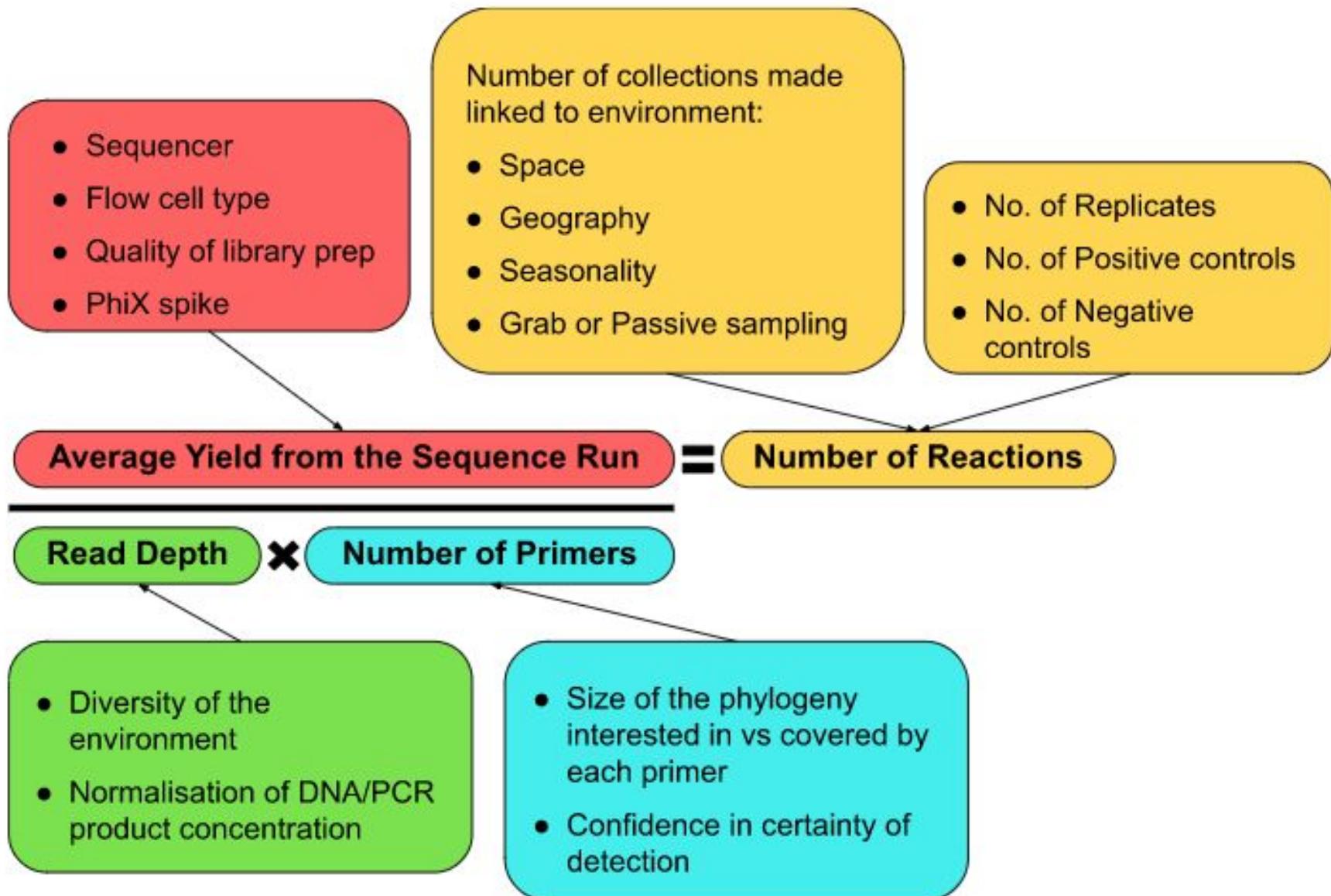
Elas02

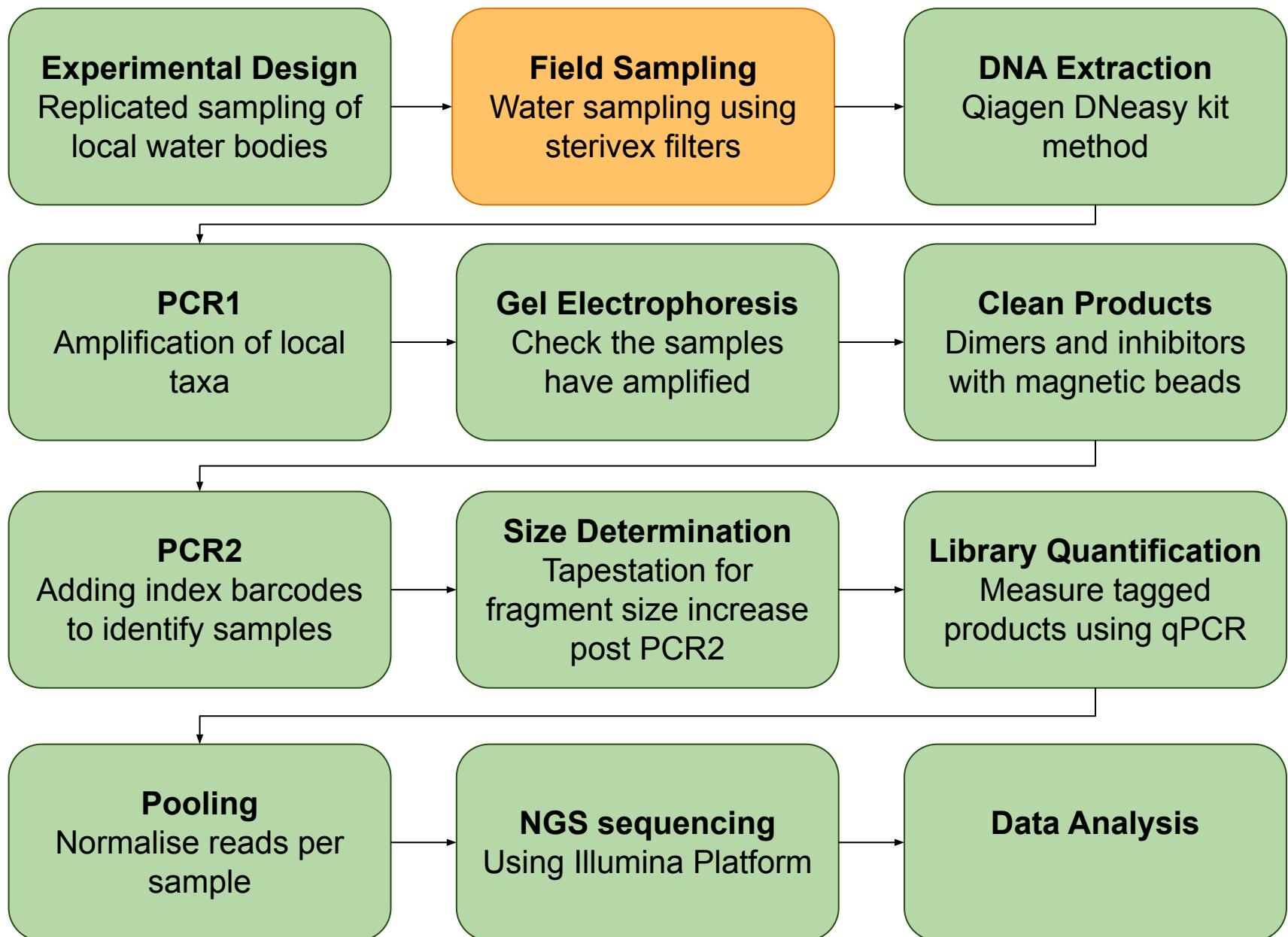
When choosing (or designing) primers it's important to have a clear focus on the scale of the phylogeny that is required to answer your question

## The Diversity vs. Precision trade off

The more conserved the target region is across a phylogeny, the greater the number of potential species but the less read depth (and therefore accuracy) for each you will yield

# Experimental Design

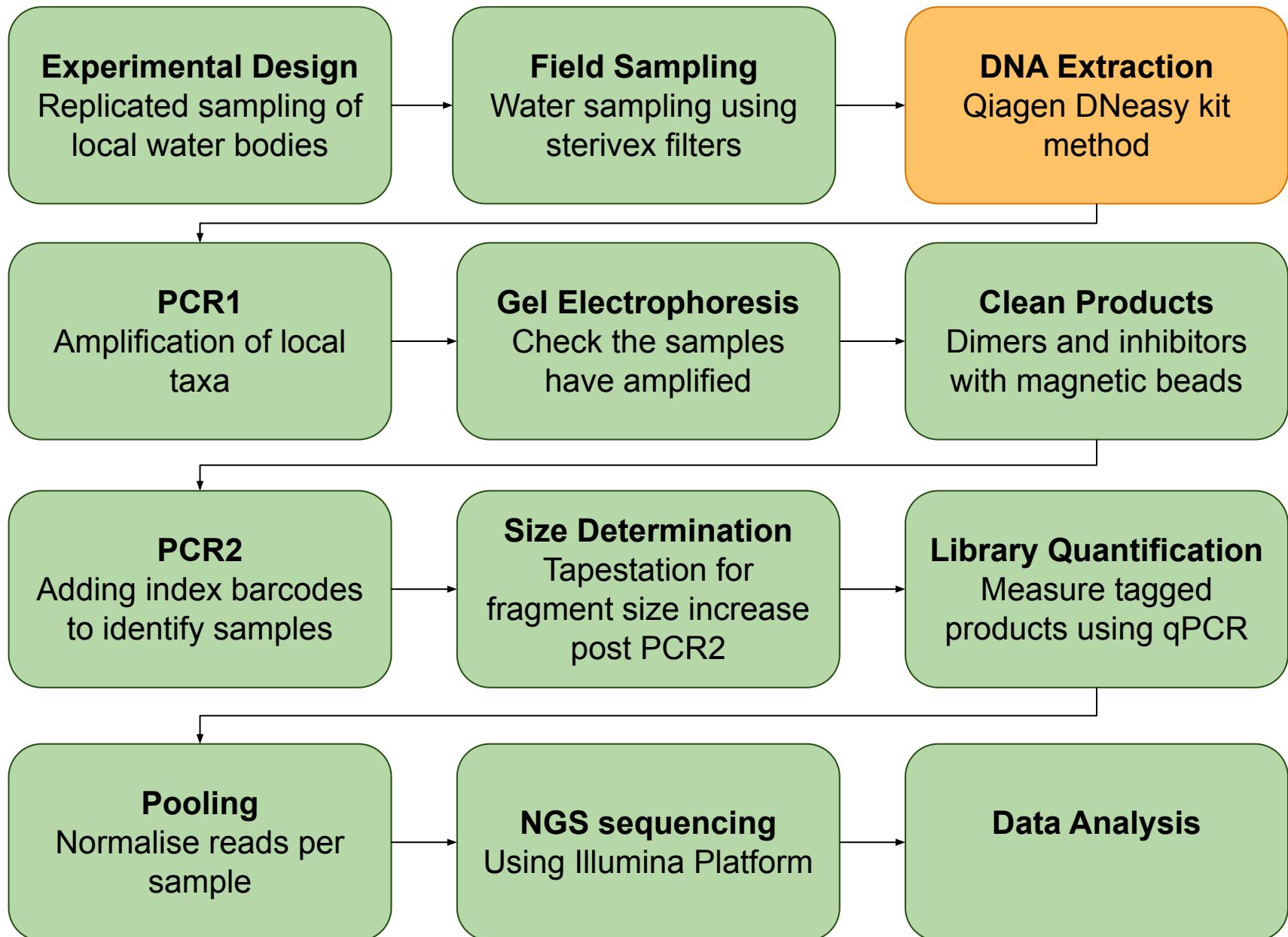




# Sample Collection: Field Equipment

- Dipper
- Sterivex filter unit
- Syringe
- 250ml bottle
- Field negative bottle
- Caulking gun
- Box of ice
- 10% Bleach
- Tissue paper roll
- Gloves
- Permanent marker

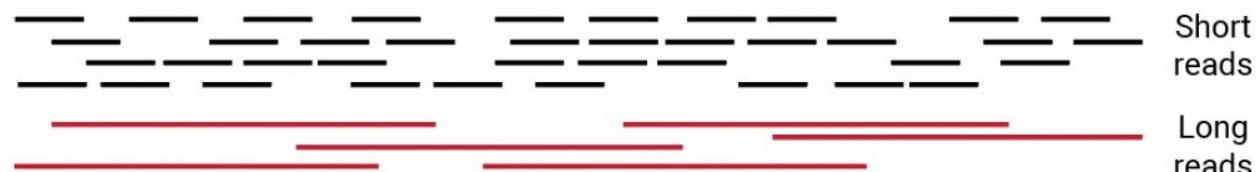




## Sample Type



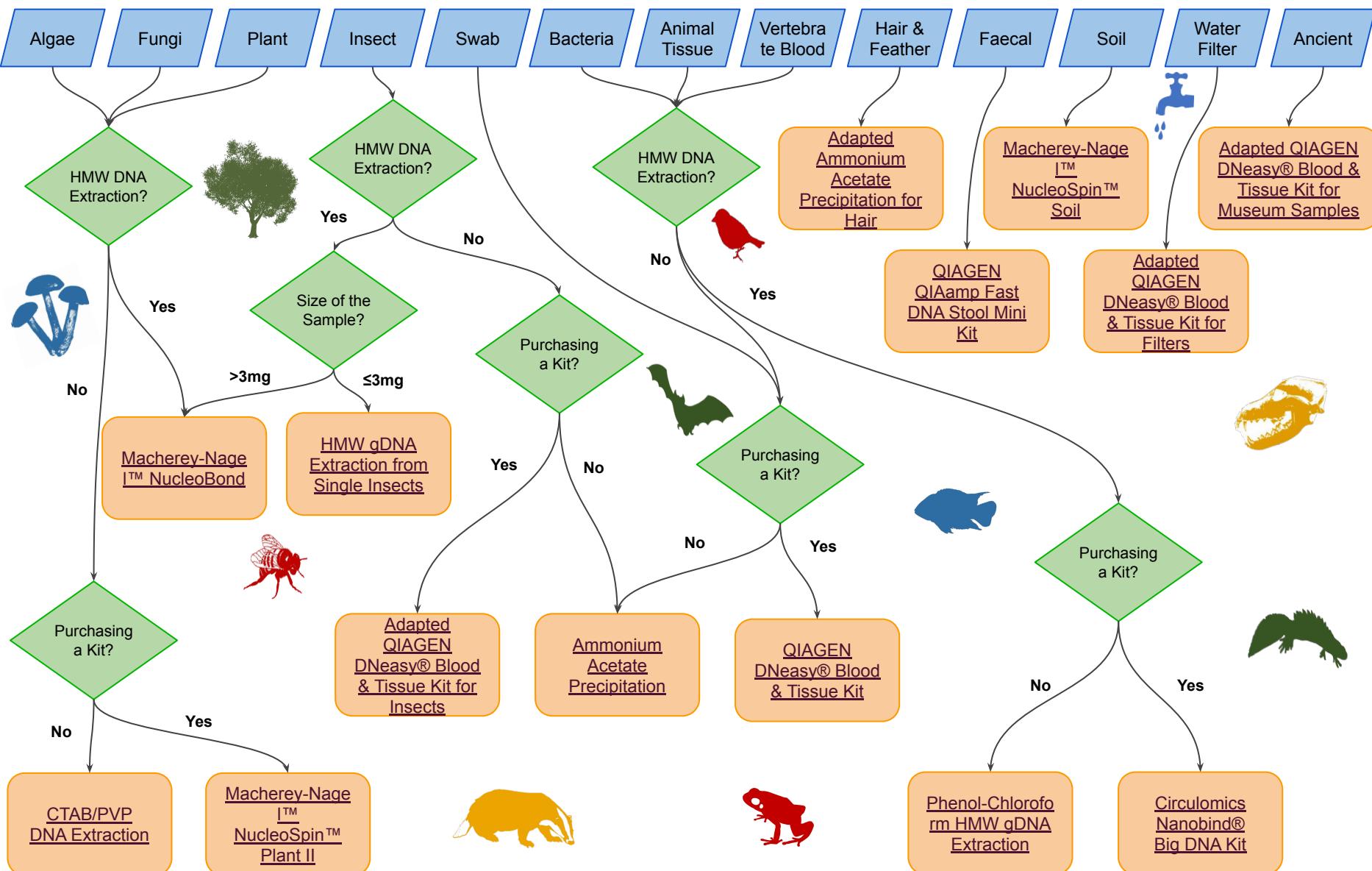
## Project Aim



## Budget



# Recommended DNA Extraction Methods



# Time to start the extractions

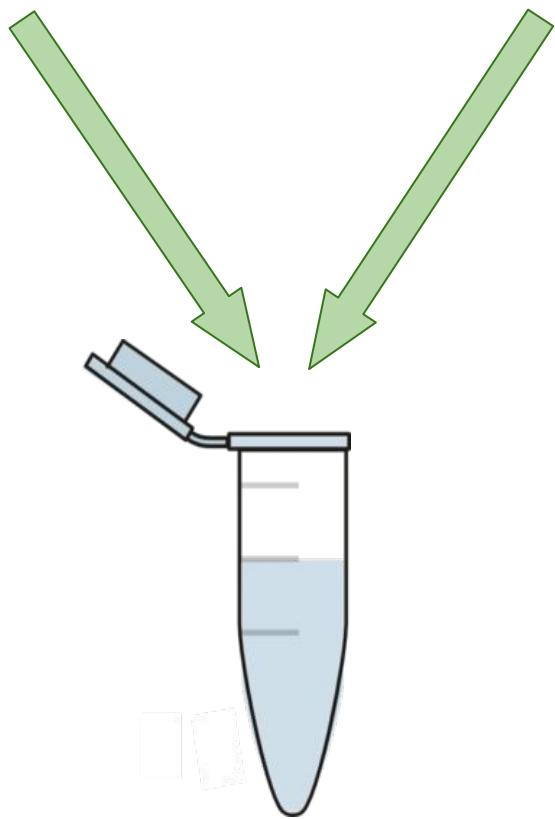




# DNA Extraction: Set-up

**400 µl Buffer ATL**

**10 µl Proteinase K**



**Remember to label your tubes!**

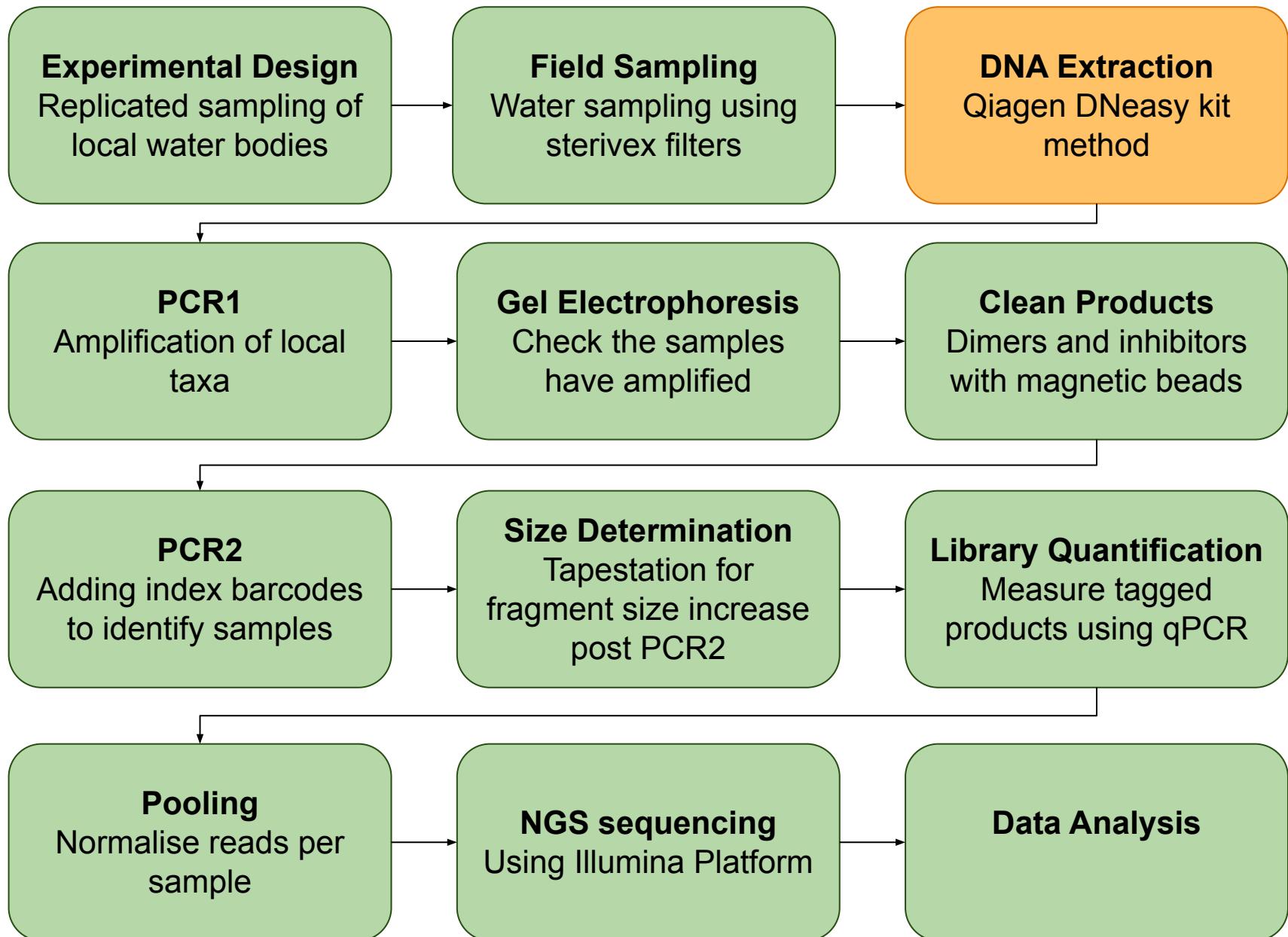
# DNA Extraction: Remove Filter Paper from Unit



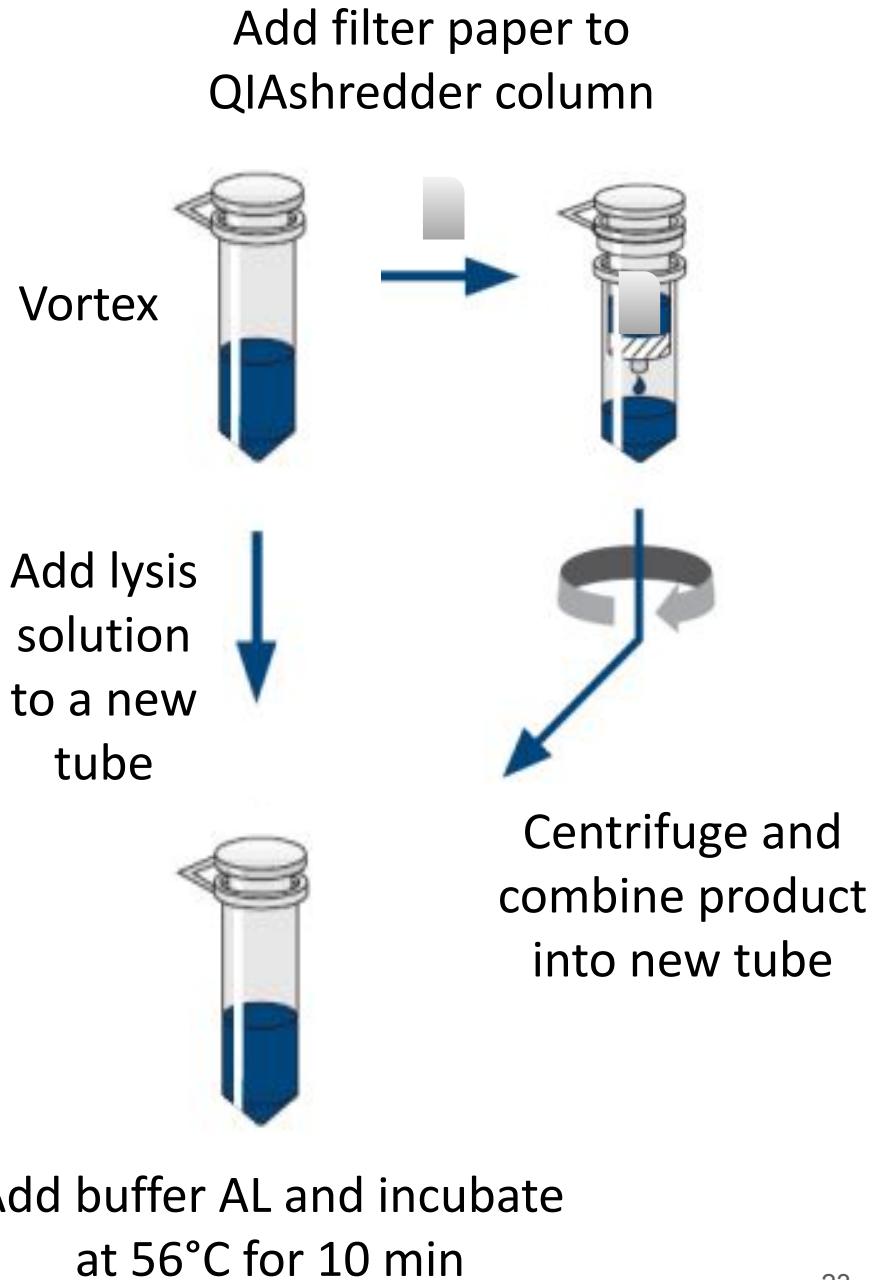
# DNA Extraction: Overnight Incubation



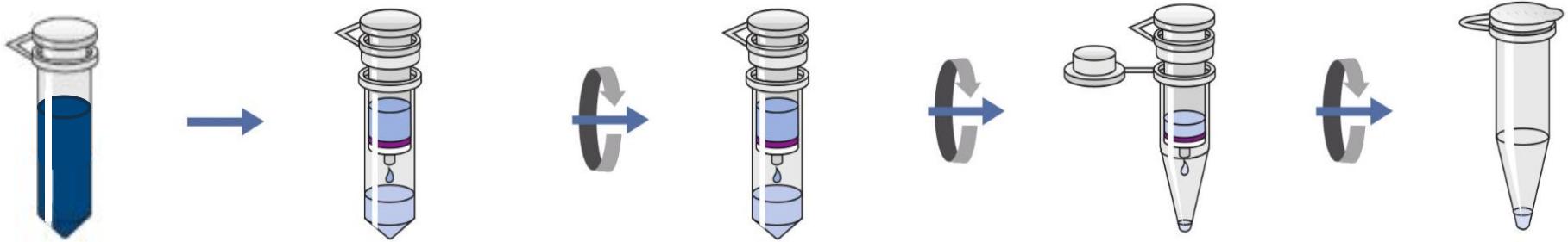
Incubate the samples at 56°C overnight



# DNA Extraction: Qiagen DNA Purification Kit



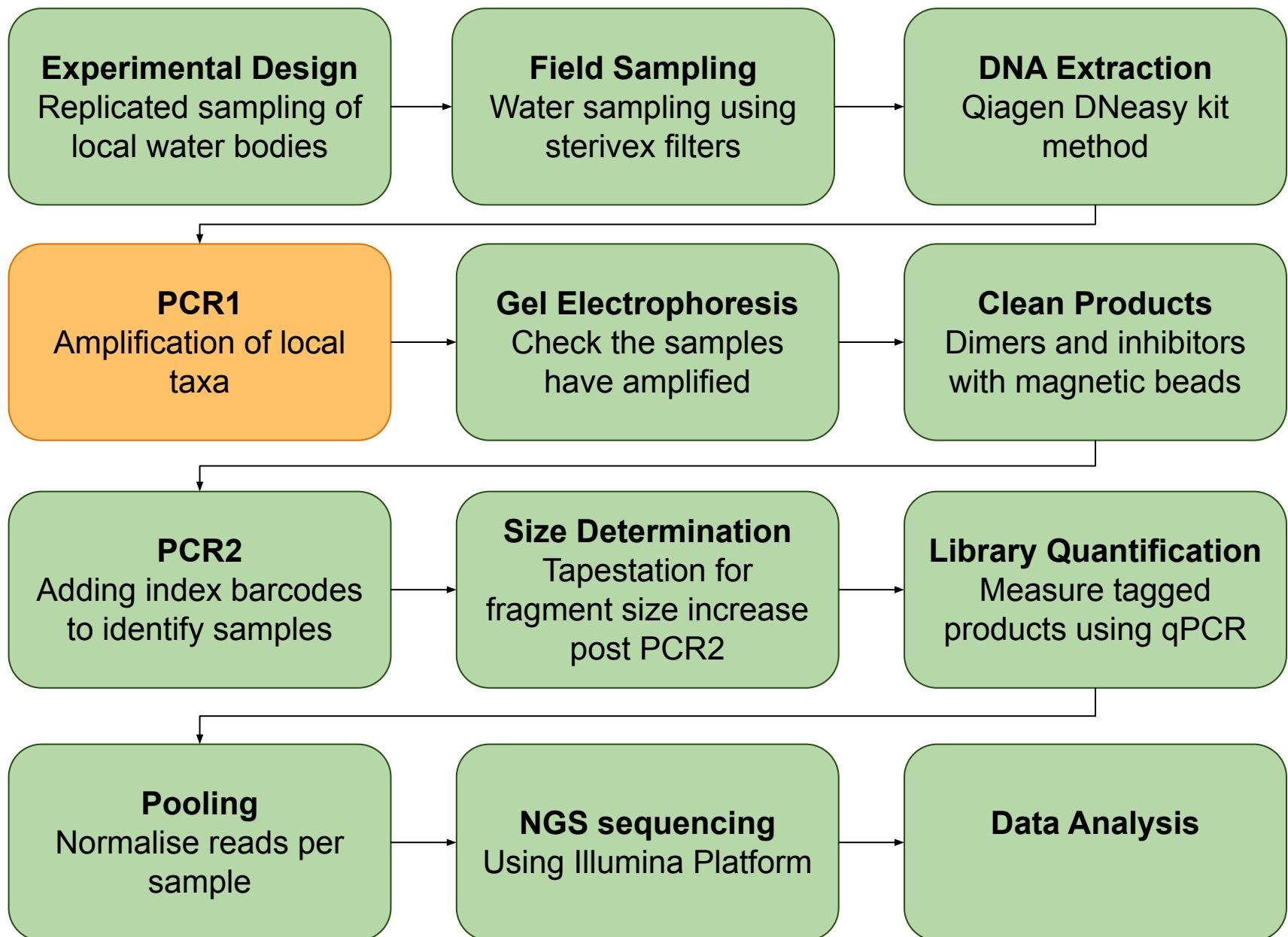
Load sample onto  
spin column



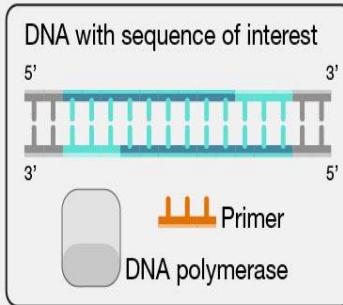
Add 100%  
ethanol and  
vortex

Wash the spin  
column x2

Clean DNA

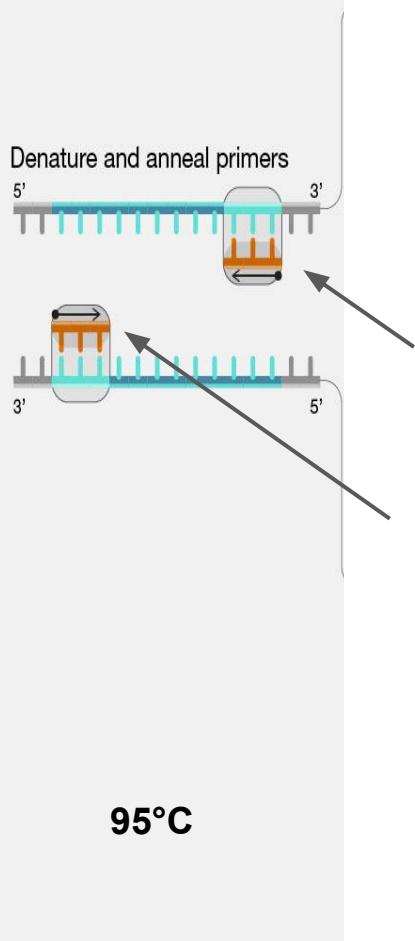


# Polymerase Chain Reaction (PCR)

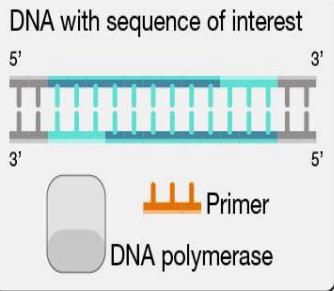


## Step 1 - Denaturation

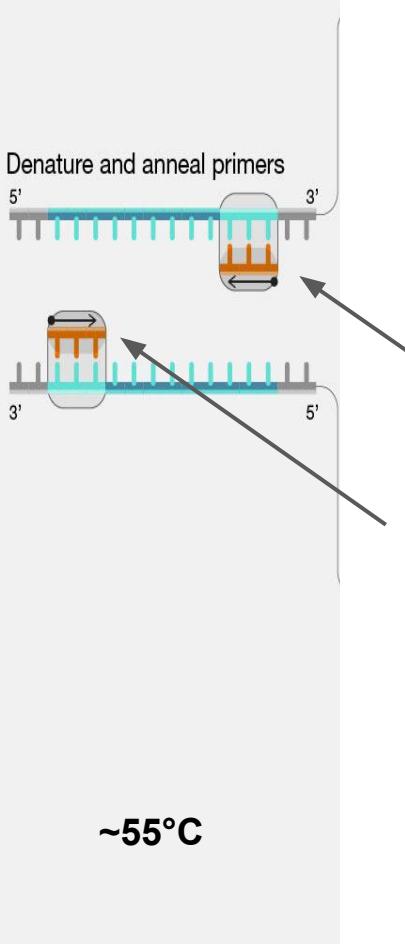
The double-stranded DNA is separated into single strands when heated



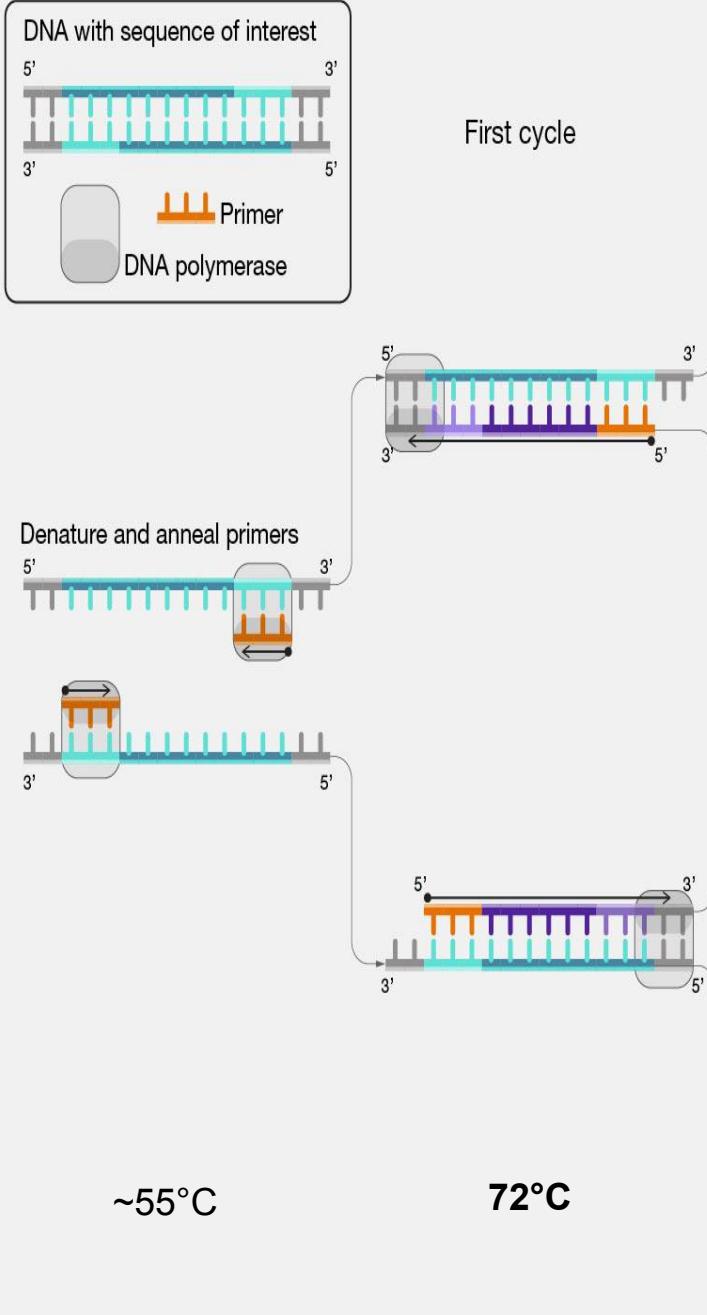
## Step 2 - Annealing



- Primers bind
- Determine melting temperature

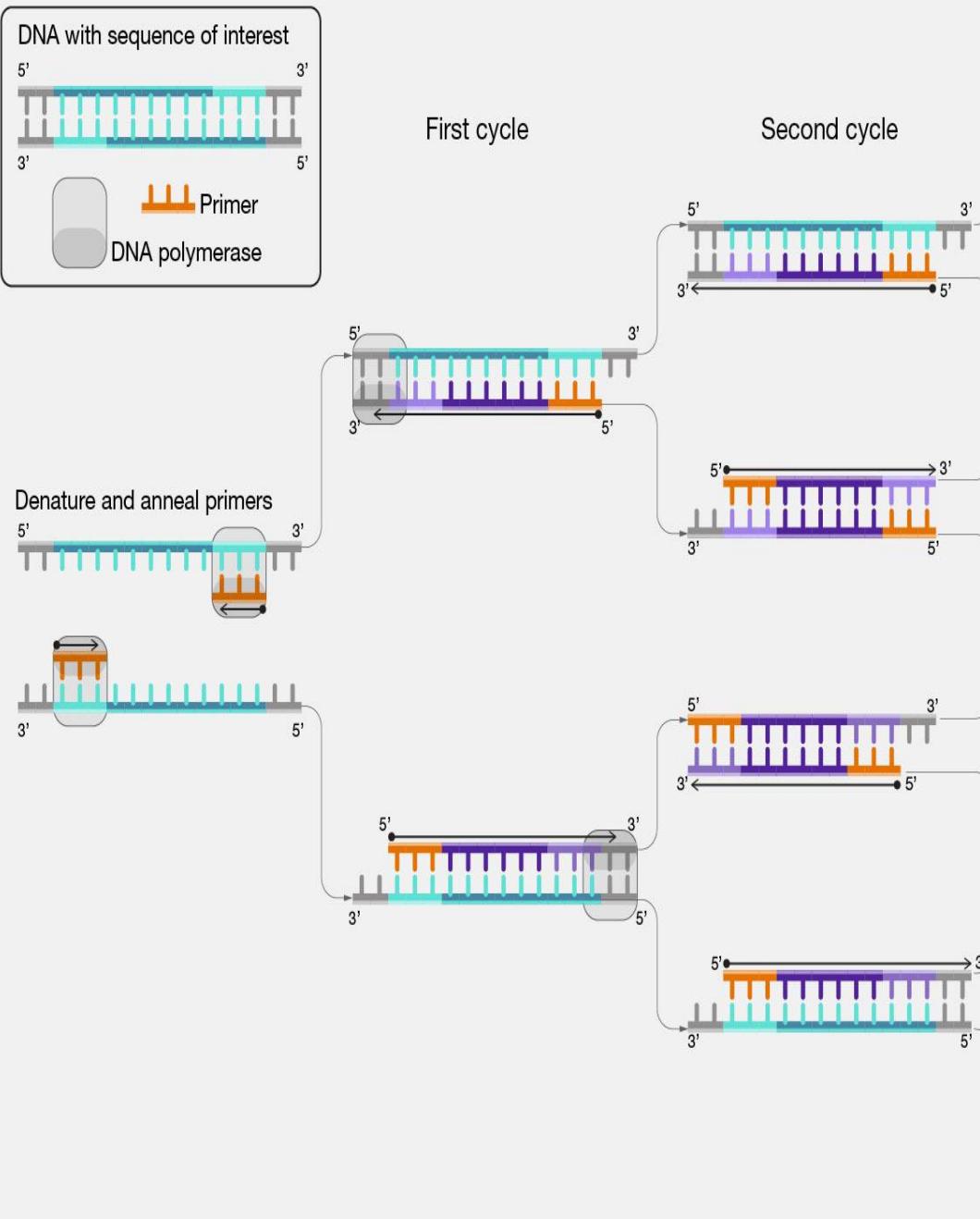


## Step 3 - Extension

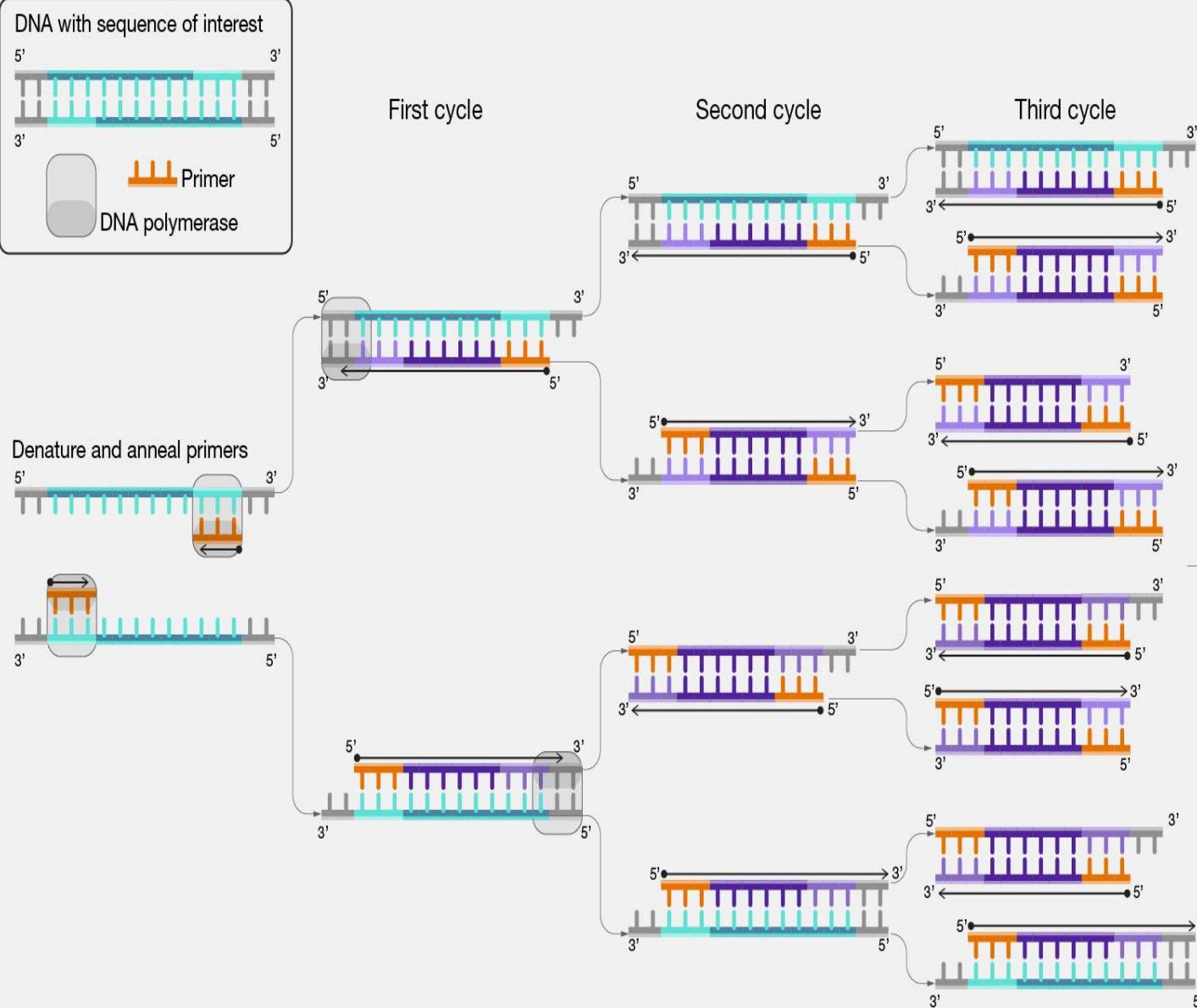


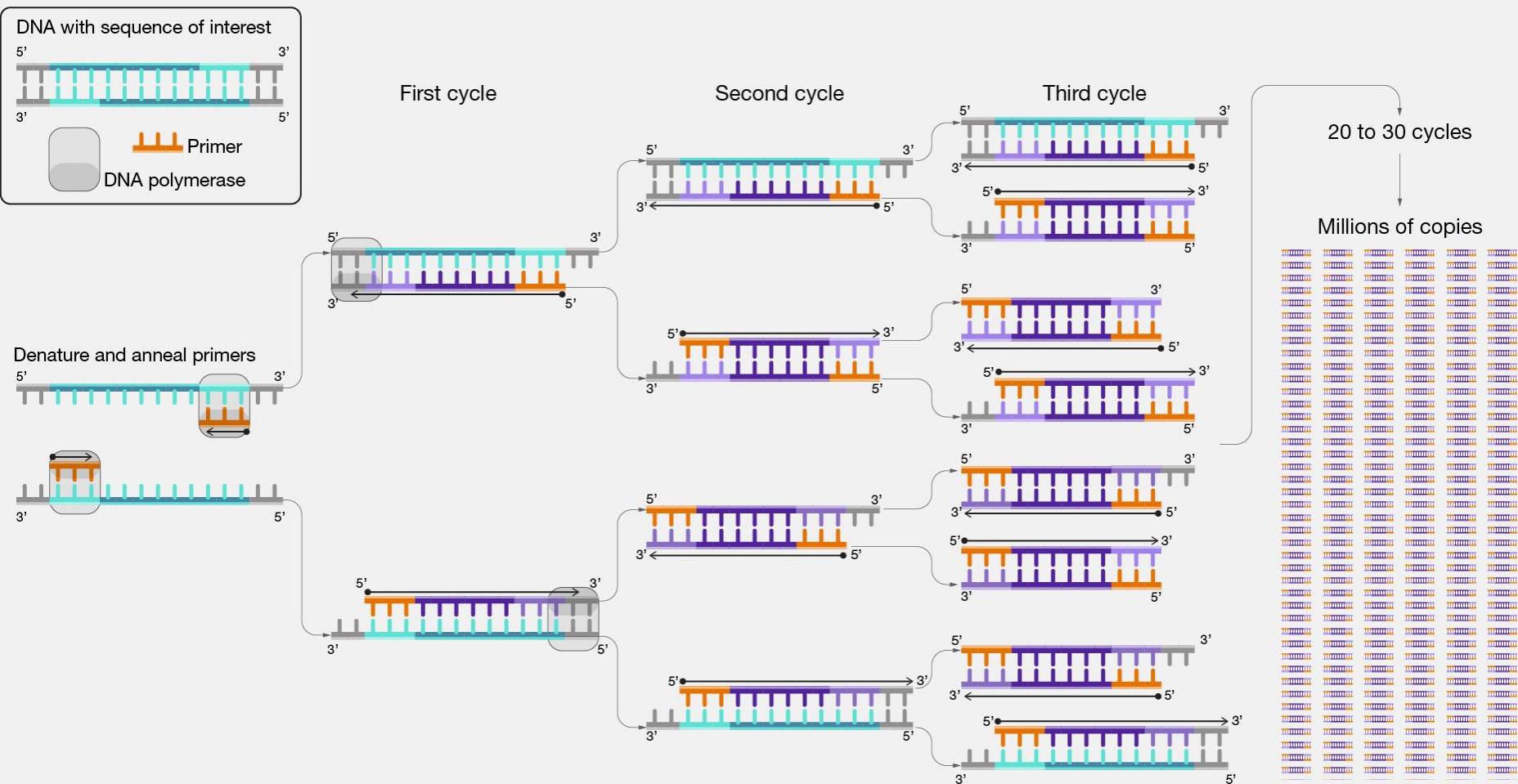
- Temperature increases
- Polymerase - amplicon extension
- Extension time

The PCR process then replicates this sequence repeatedly over 20 to 30 cycles



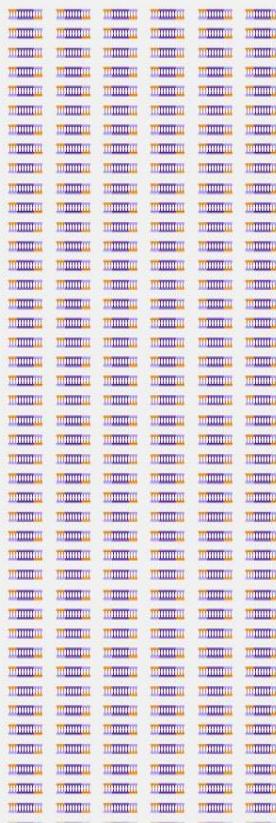
The number  
of replicates  
grows  
exponentially





20 to 30 cycles

Millions of copies



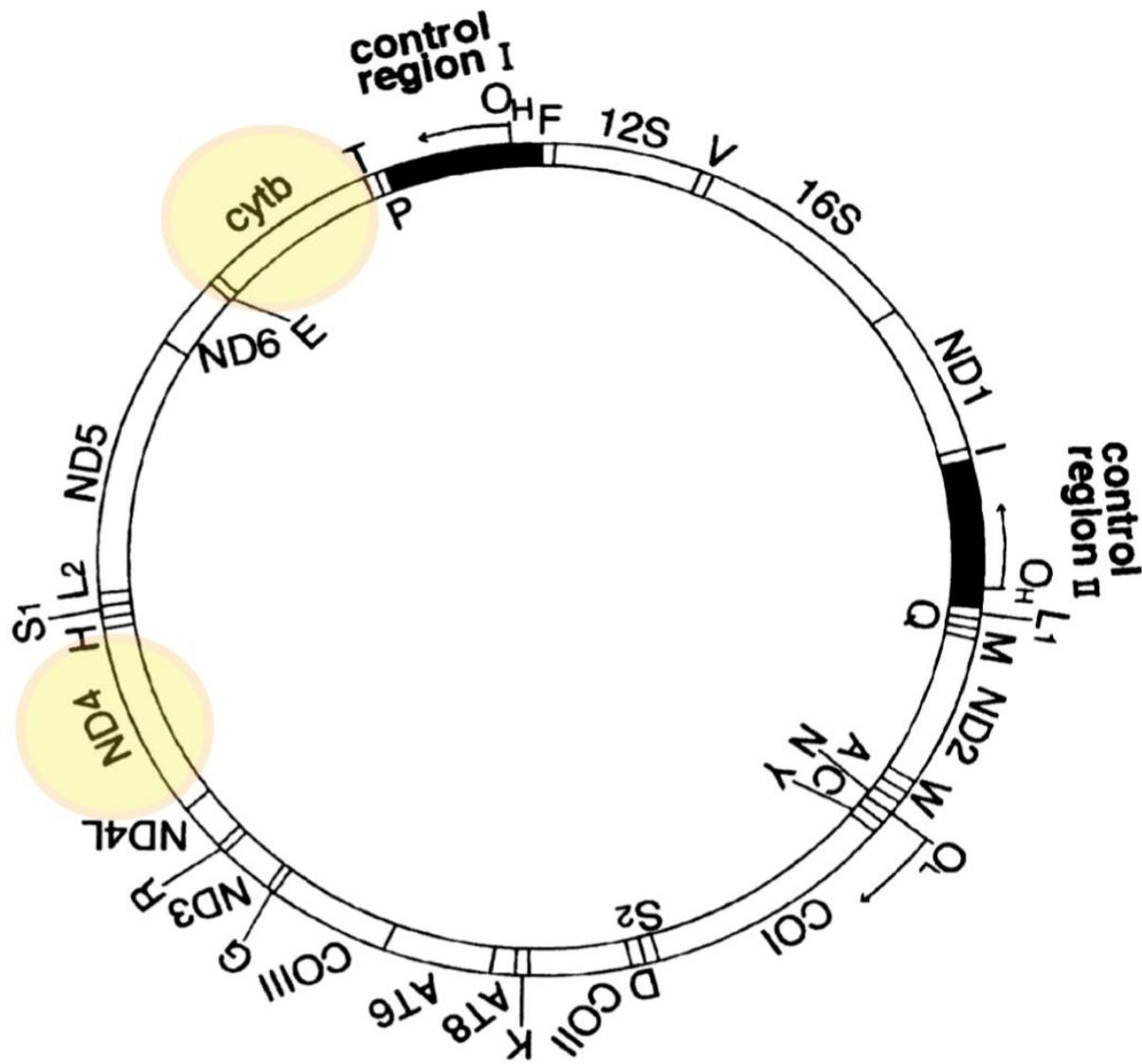
The result - millions of exact copies of your genetic sequence, known as amplicons

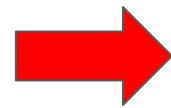
# The Thermocycler



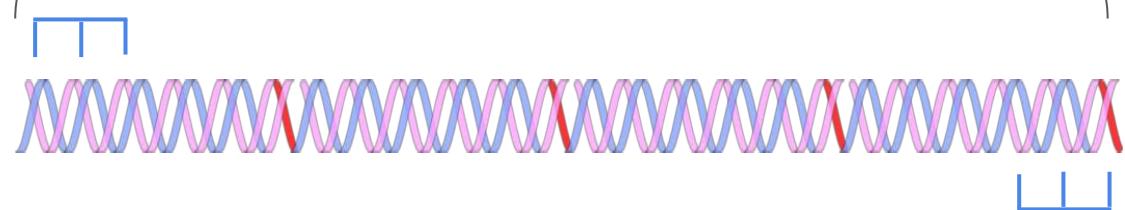
- Easily programmable
- Timings, temperature, cycles
- Tailor conditions to your sequence

# PCR 1 - Primer Selection

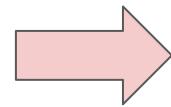


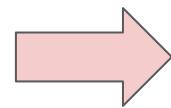


Amplicon

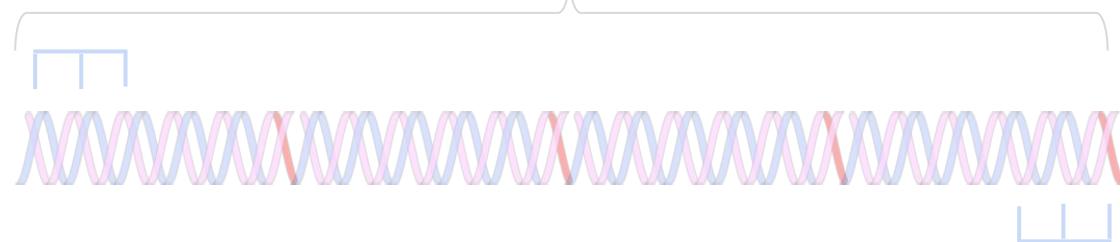


Amplicon length

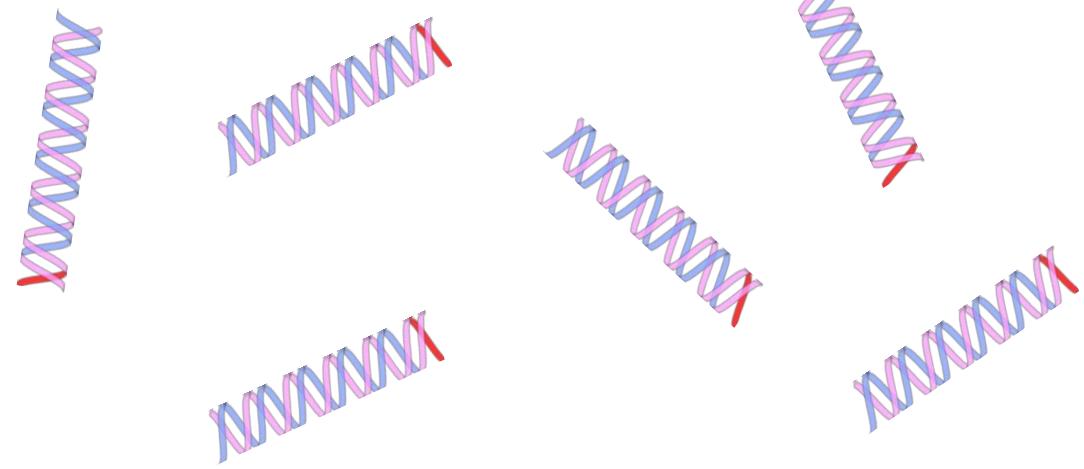
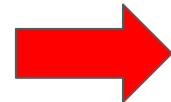




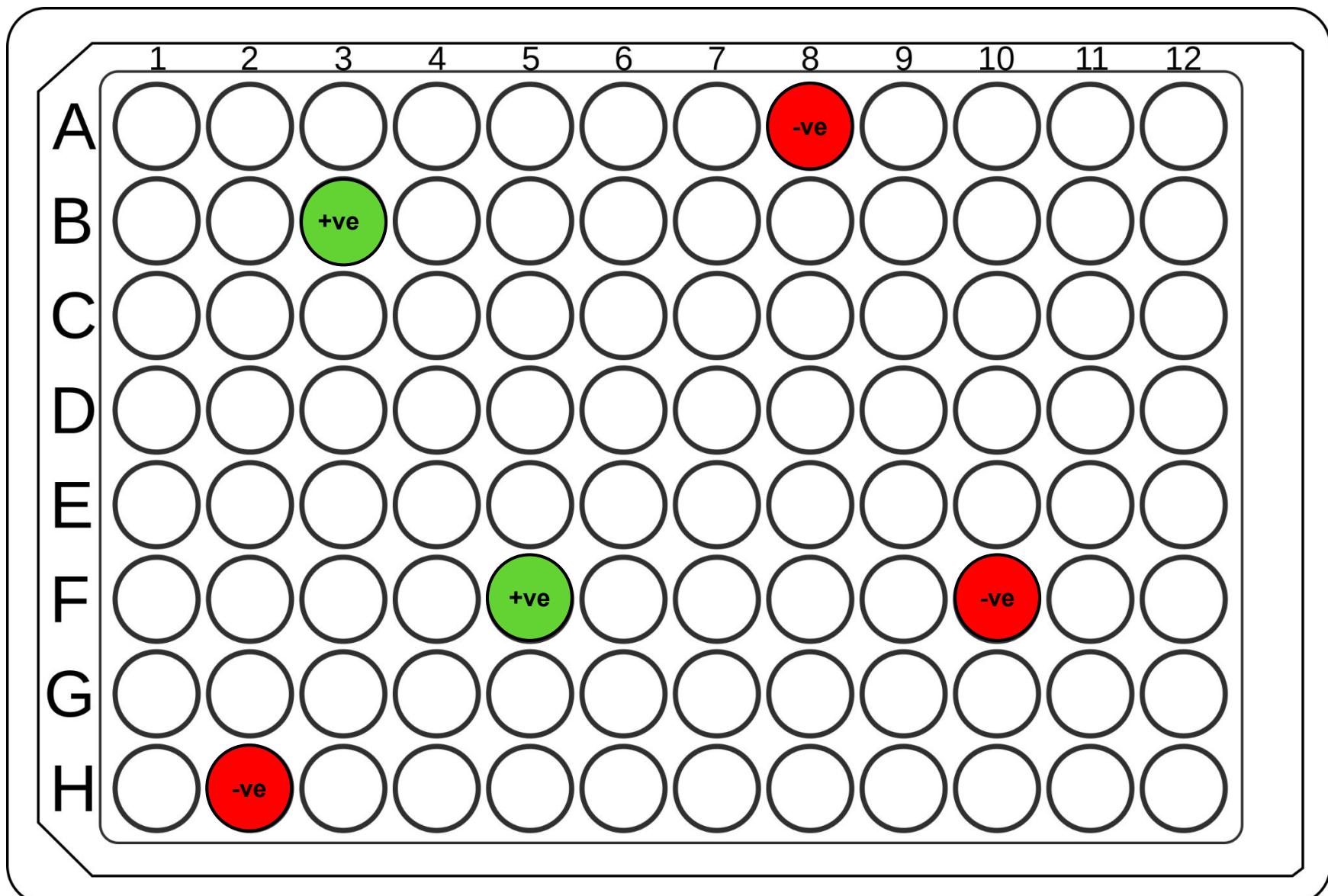
Amplicon



**Amplicon length**

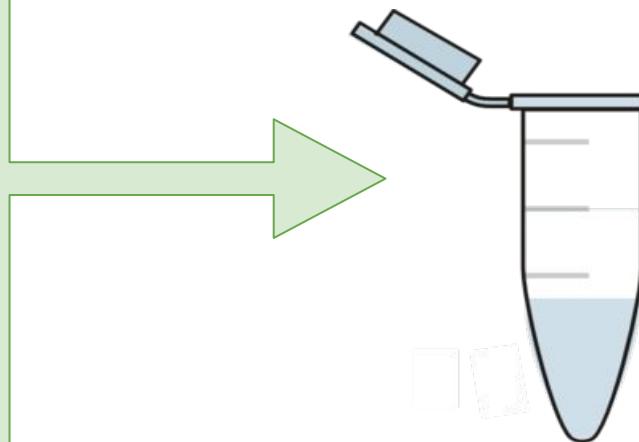


# PCR 1 - Quality control

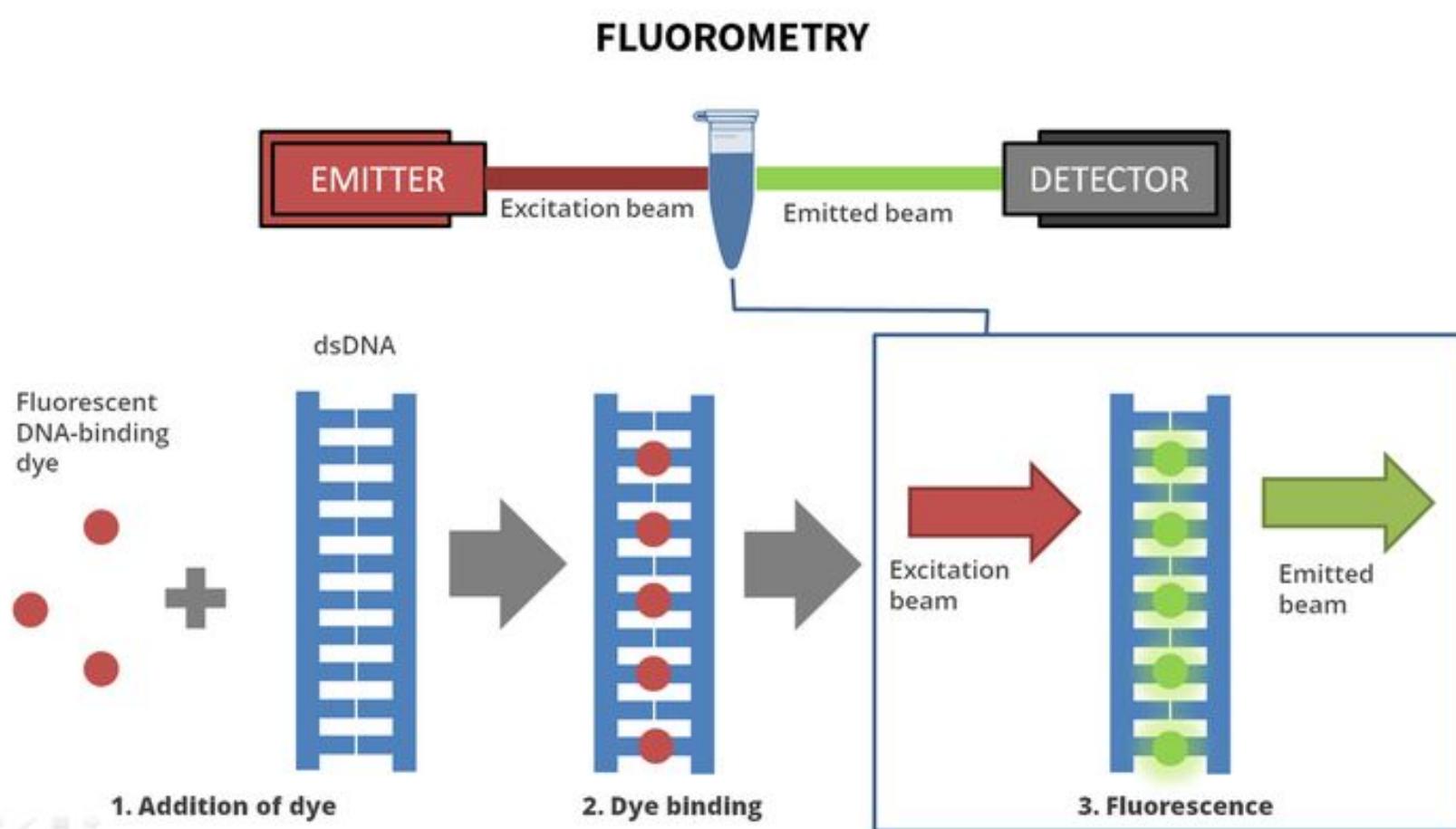


# Setting up a PCR

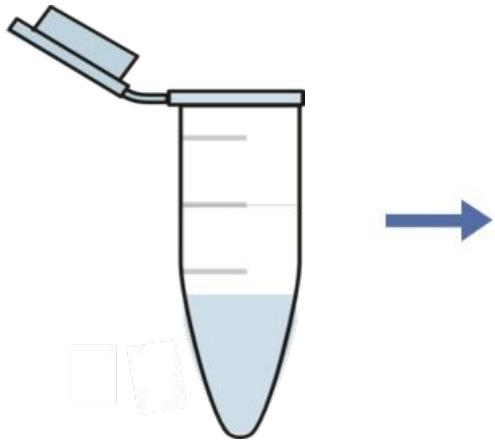
2 µl diluted DNA sample  
1 µl 12s Reverse Primer  
1 µl 12s Forward Primer  
2 µl ddH<sub>2</sub>O  
4 µl MyTaq HS Mix



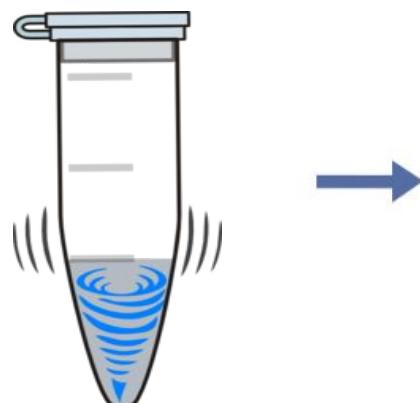
# Quantification



# Quantification

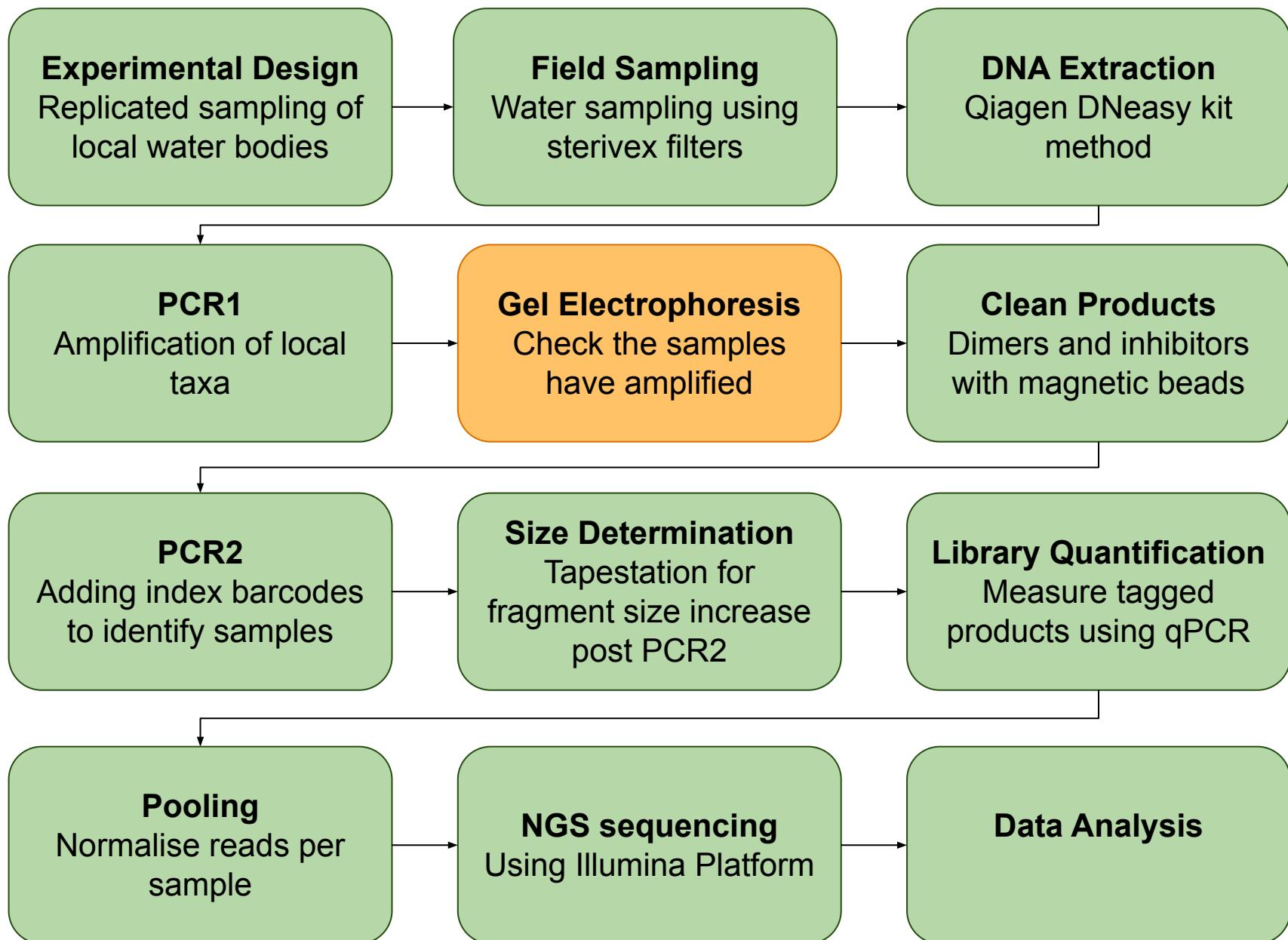


Add **1 µl** of DNA  
to Qubit Buffer

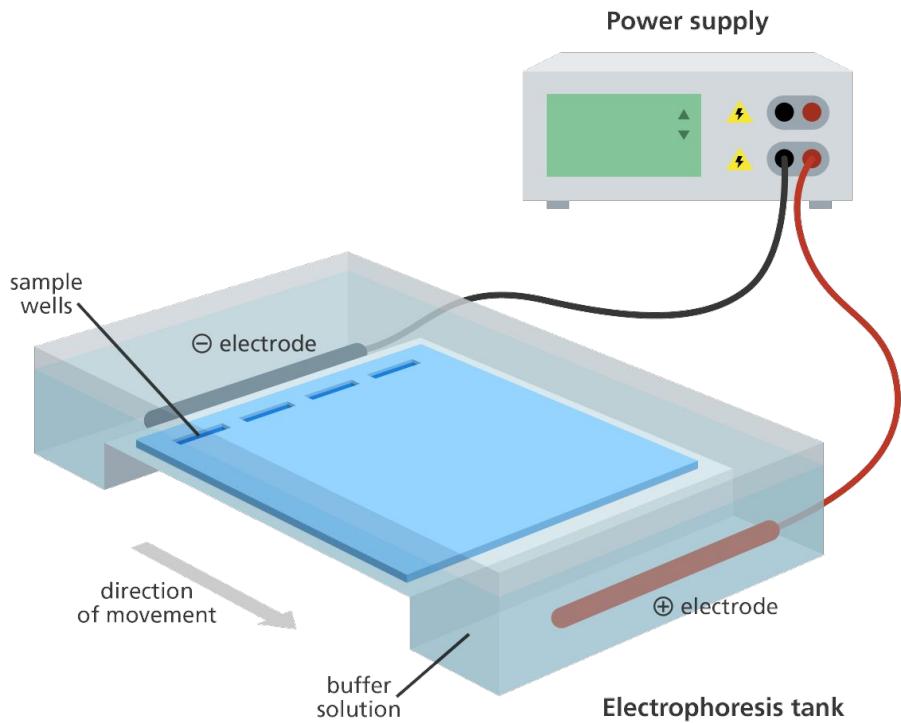


Vortex





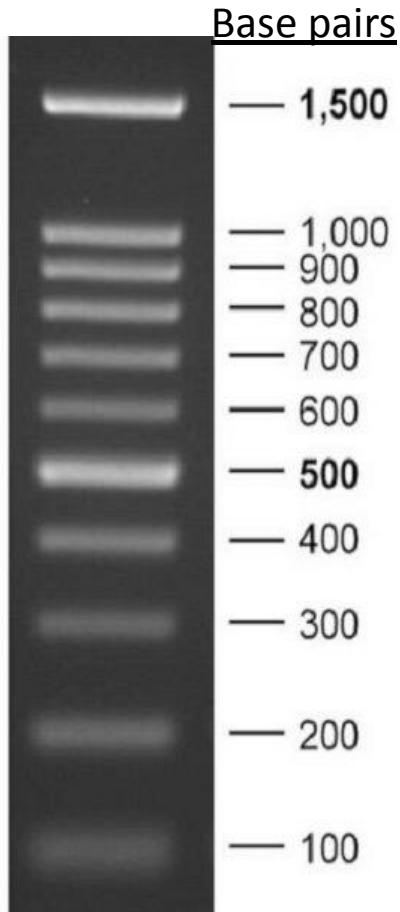
# Gel Electrophoresis



- 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

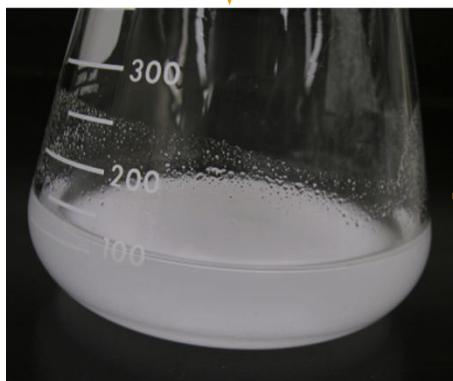
# Gel Electrophoresis

## 100bp ladder

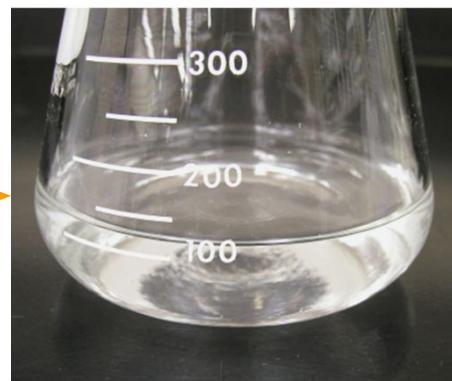


- A ladder is usually added to the first lane in order to visualise the length of your PCR products, measured in base pairs
- We already have an idea of what the length our PCR products should be, based on the primers we designed

# Making a 1% Agarose Gel



Microwave



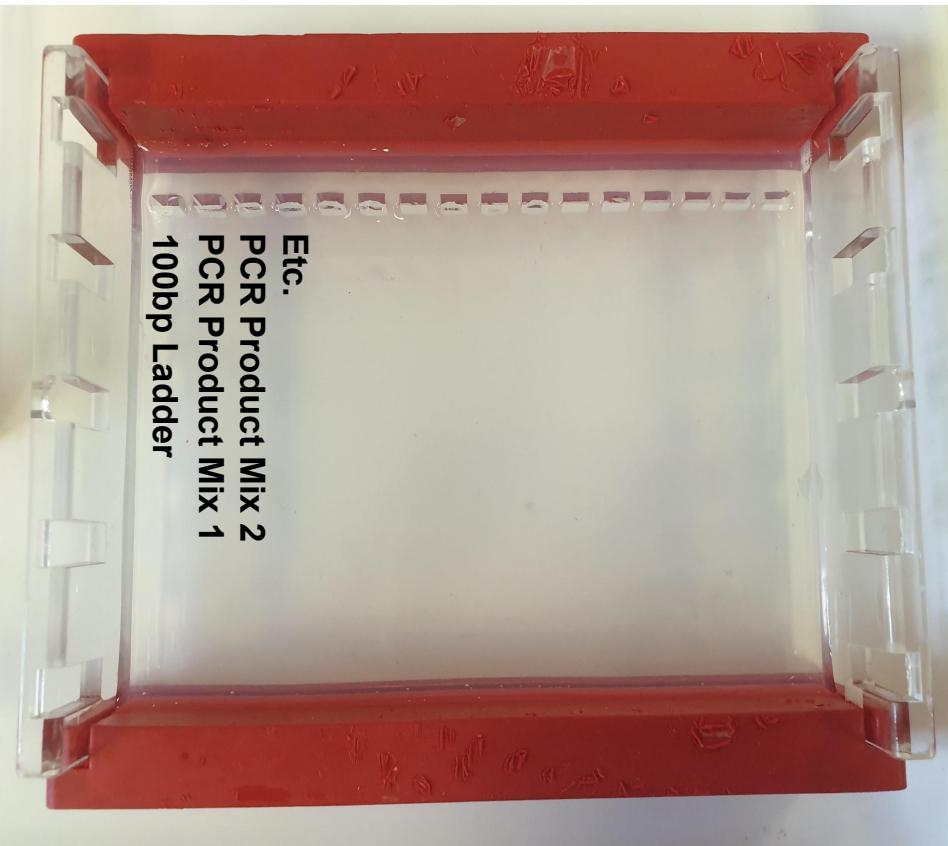
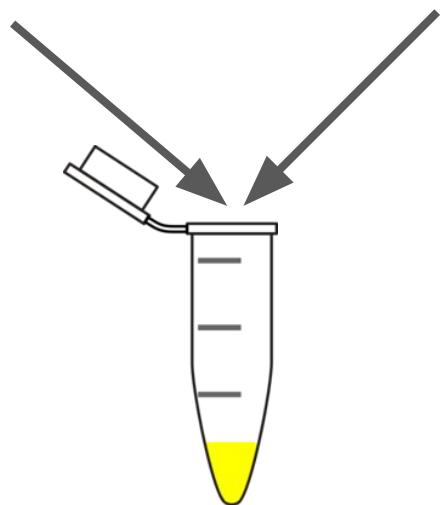
Add 4  $\mu$ l  
Ethidium  
Bromide



# Loading the Gel

4 $\mu$ l PCR  
product

6 $\mu$ l Orange G



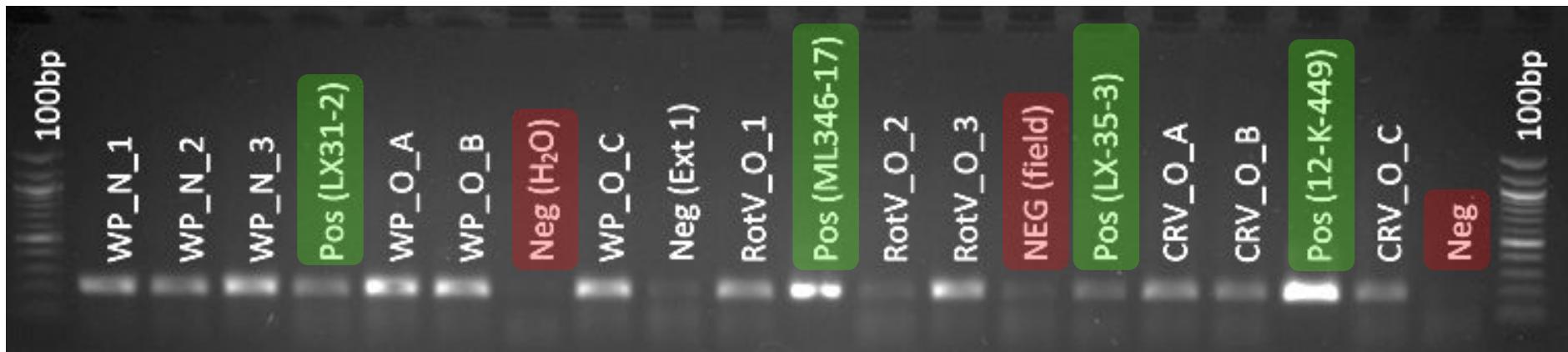
Be careful not to pierce the gel

# PCR 1 - Visualising Products

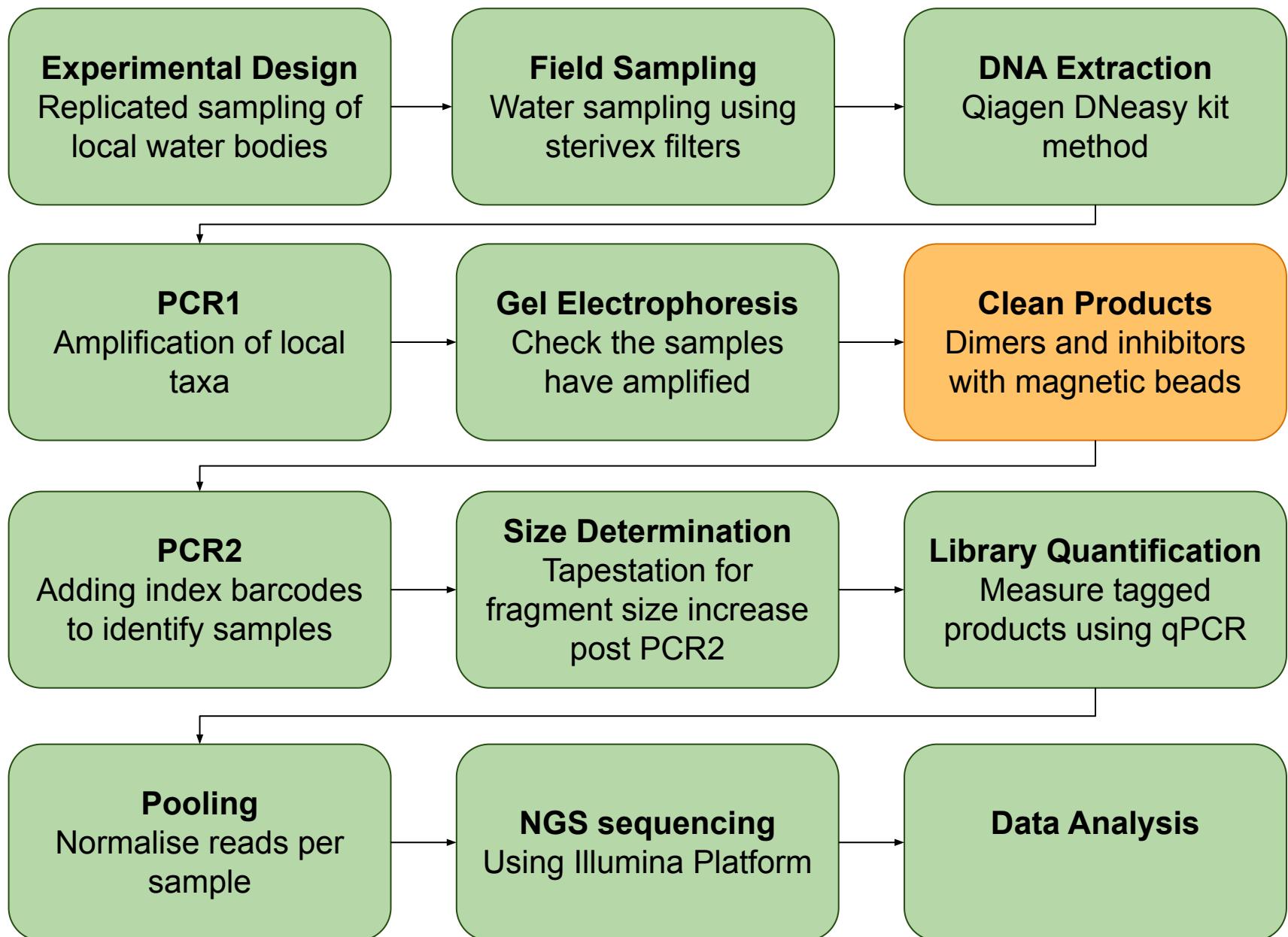


A gel imaging dock like this uses UV light to illuminate the PCR products in the gel

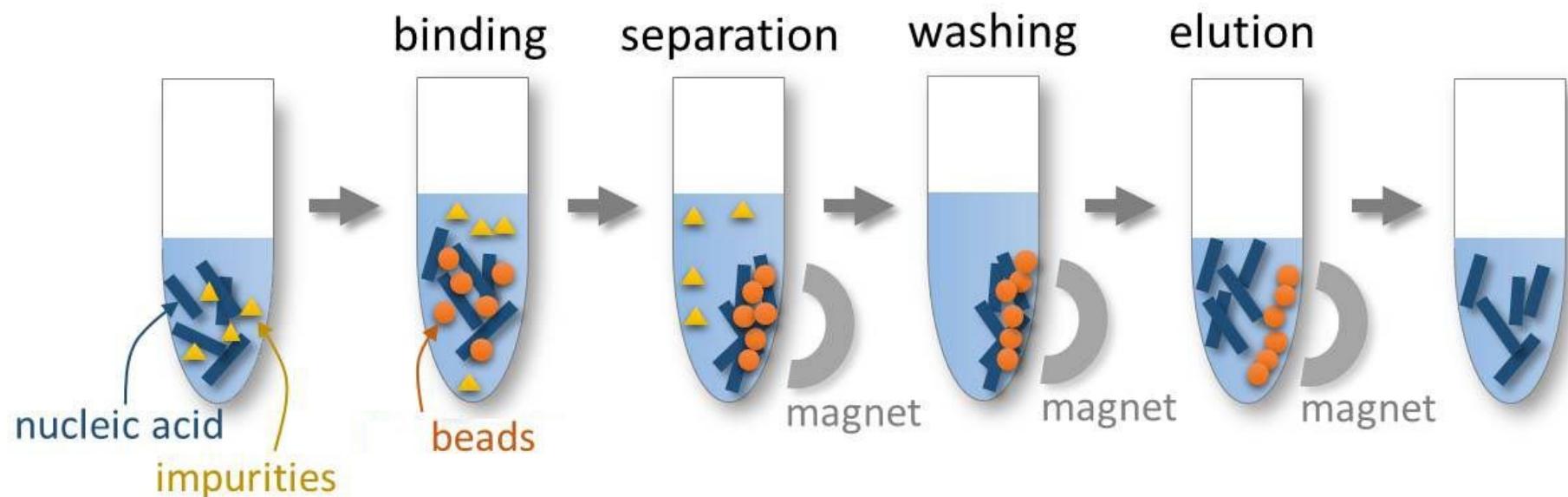
# PCR 1 - Visualising Products



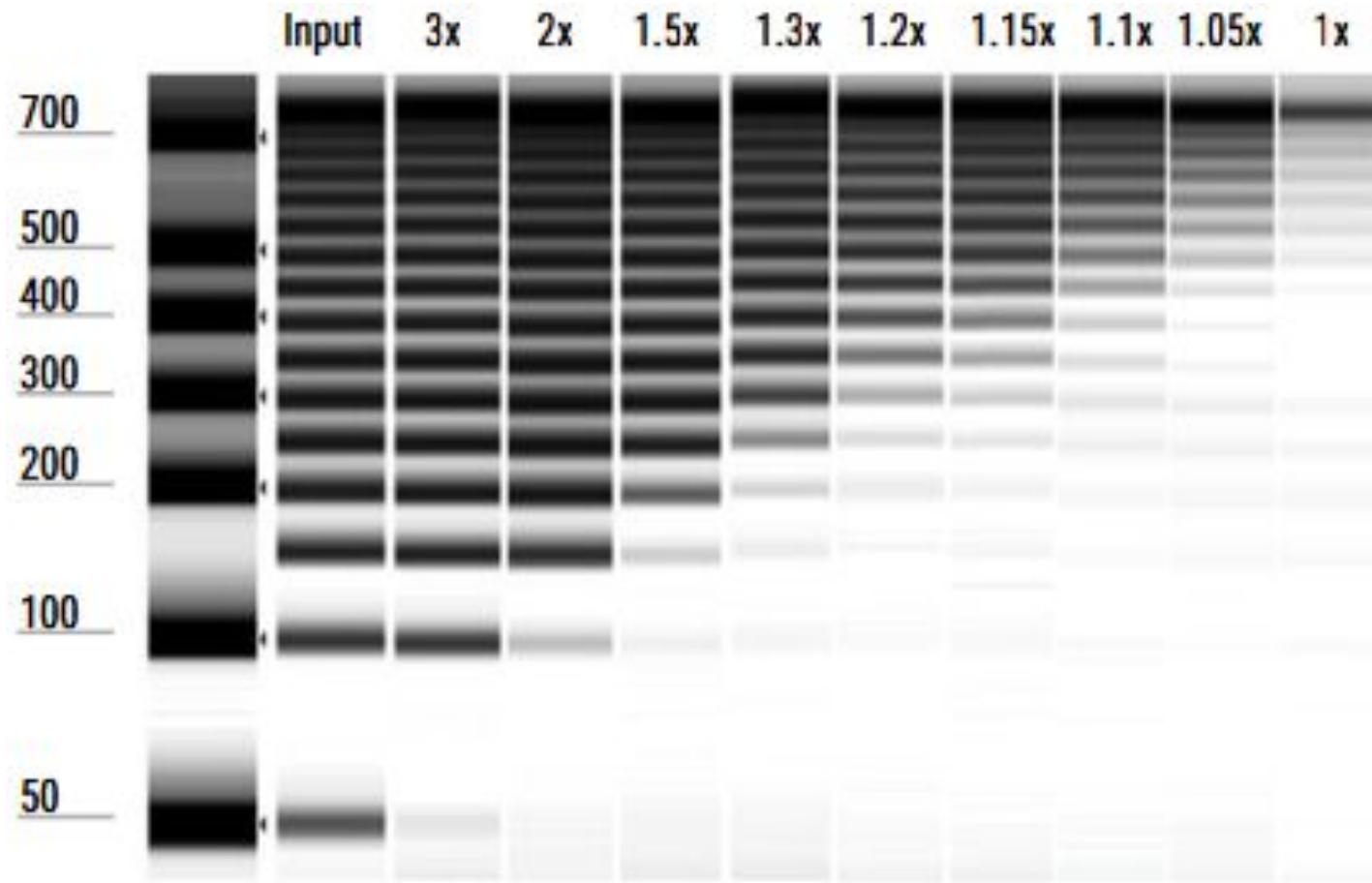
Post-PCR1 amplicons from local of sampling sites



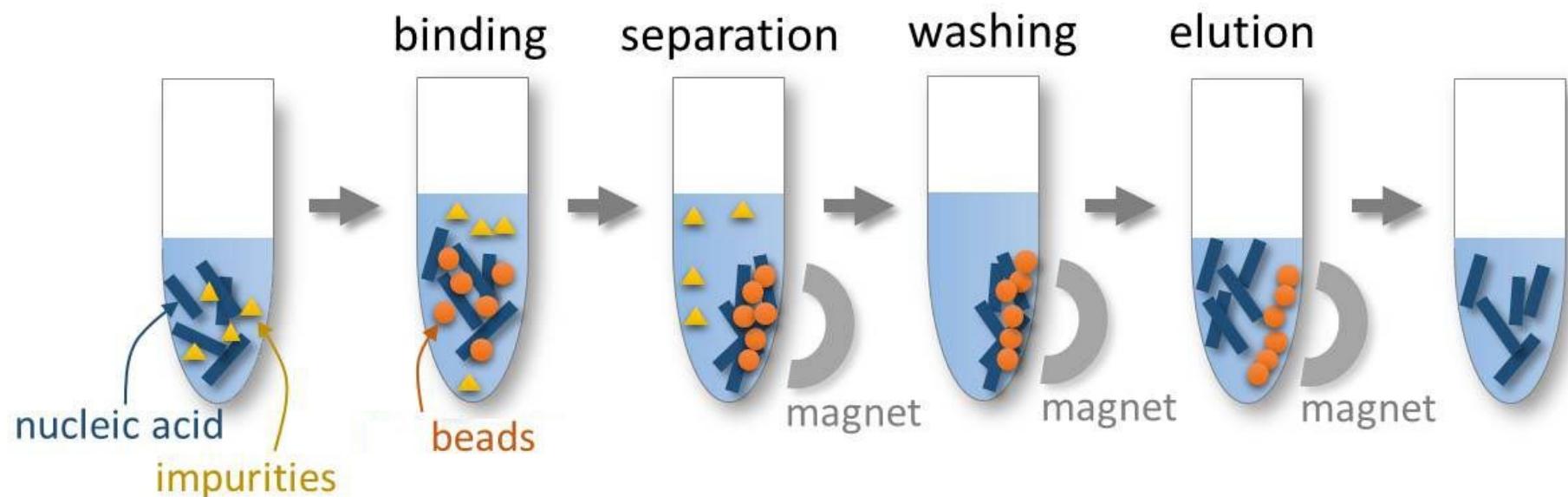
# PCR 1 - Bead cleaning

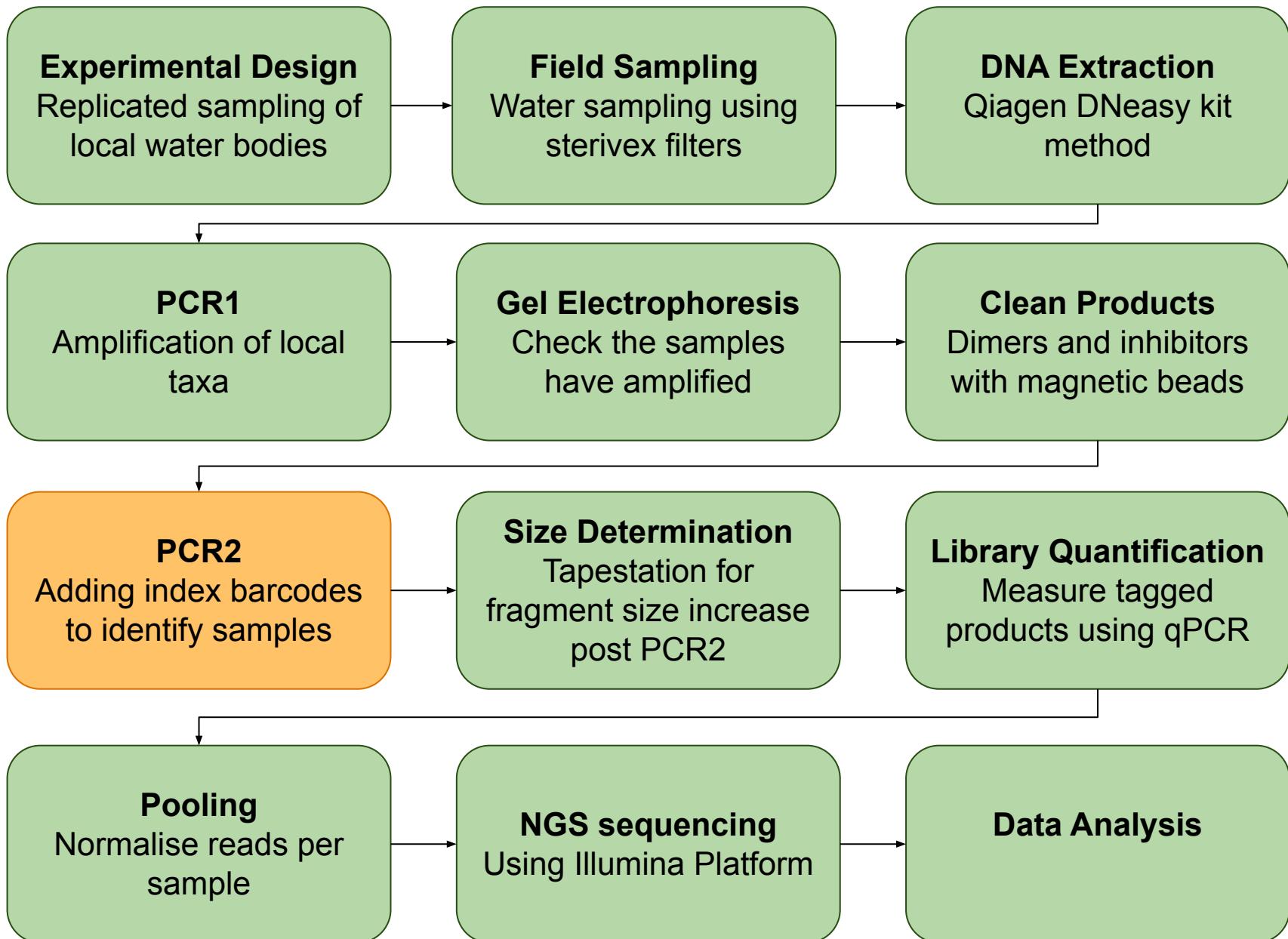


# PCR 1 - Bead cleaning



# PCR 1 - Bead cleaning

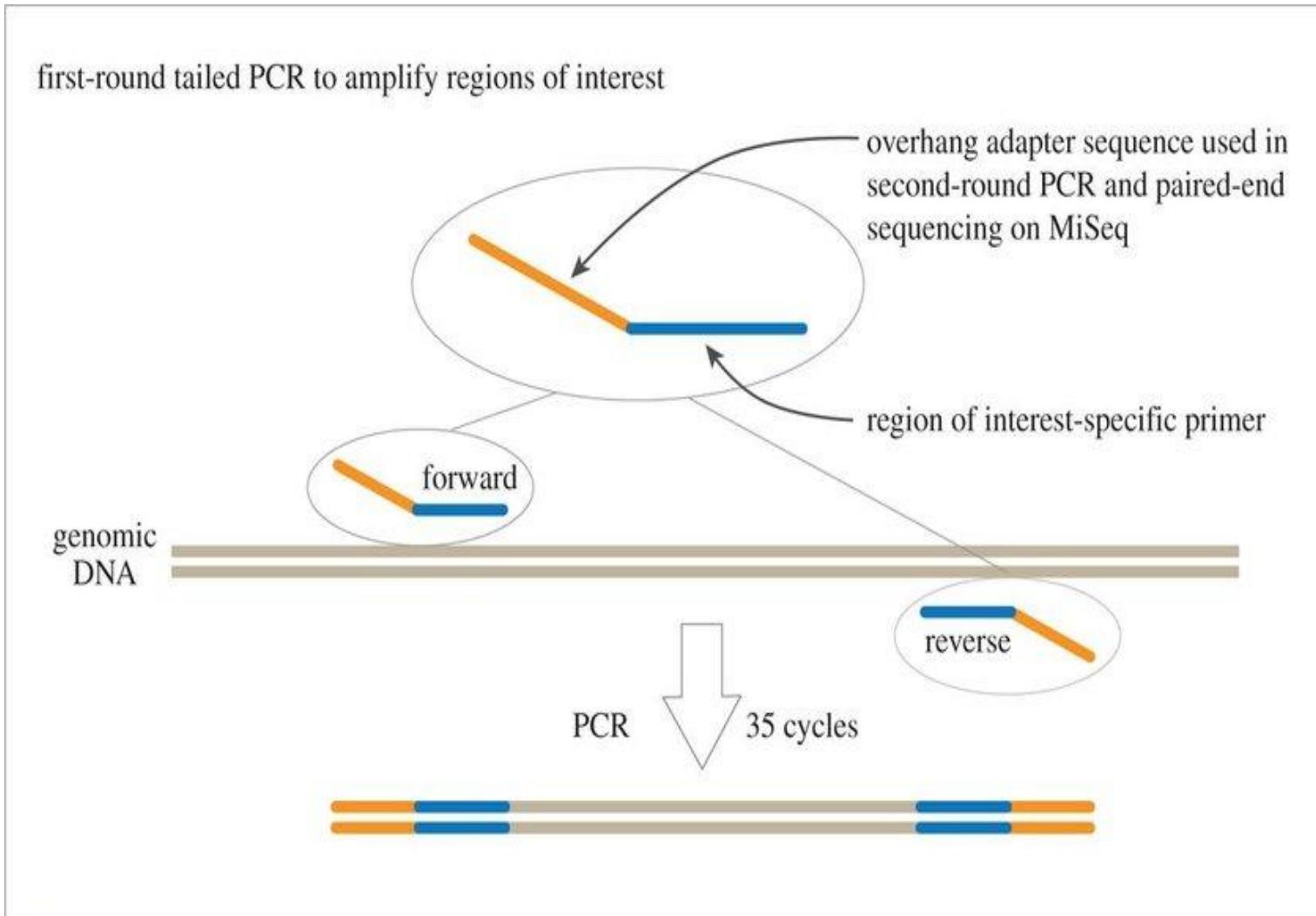




- During PCR1, our primers (**Blue**) incorporate an overhang adapter sequence (**Orange**) at the ends of our amplicons

F: 5'**ACACTTTCCCTACACGACGCTTCCGATCTNNNNNNGTGCCAGCMGCCGCGGTAA**3'

R: 5'**GTGACTGGAGTTCAGACGTGTGCTCTCCGATCTGGACTACHVGGGTWTCTAAT**3'

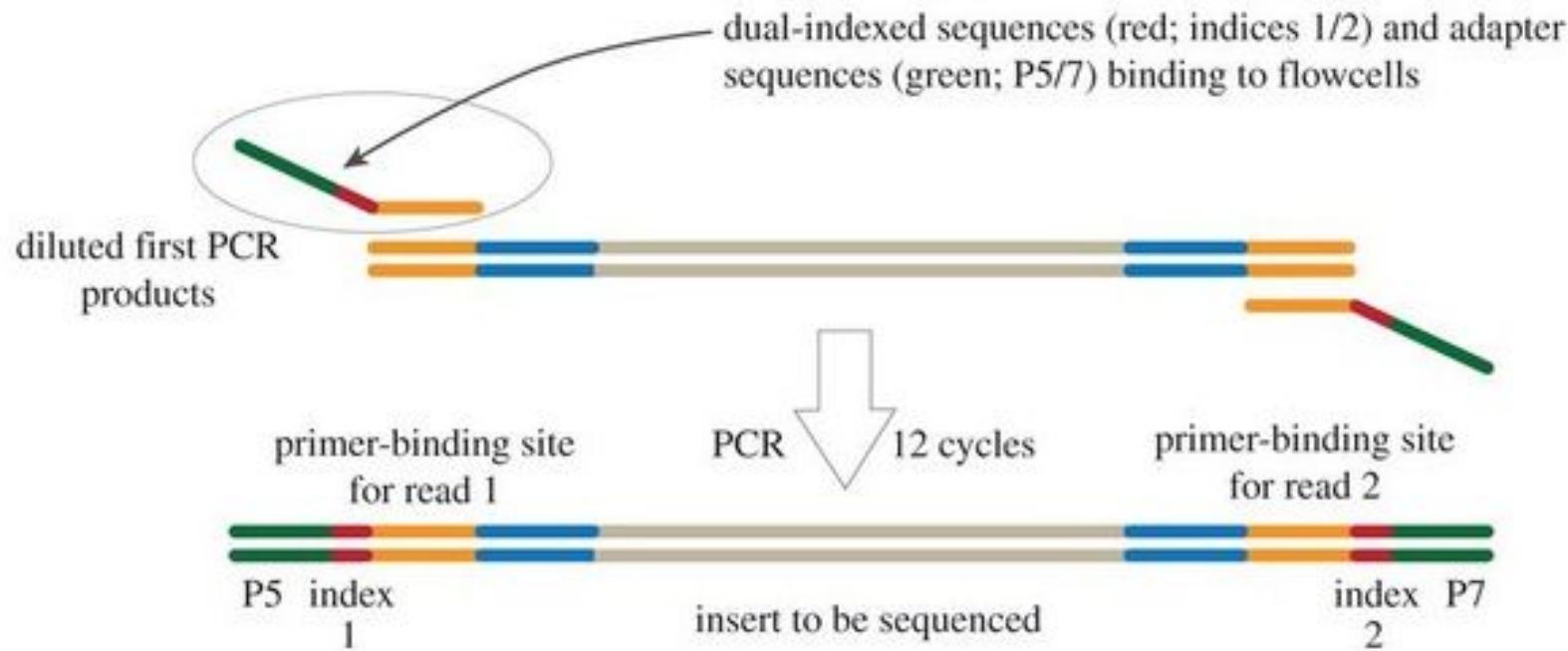


# PCR 2

- In this second PCR reaction, unique identifier sequences are attached to our PCR 1 products

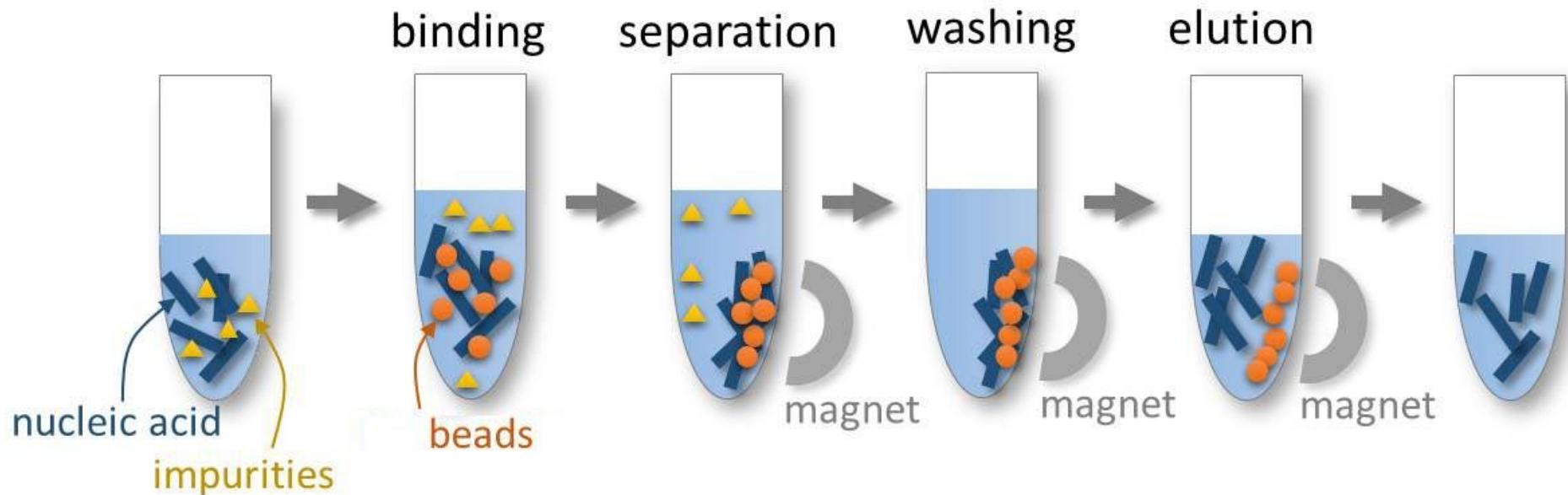
i5 index seq AATGATAACGGCGACCACCGAGATCTACACATGCTTACTGACACTCTTCCCTACACGACGCTTCGATCT

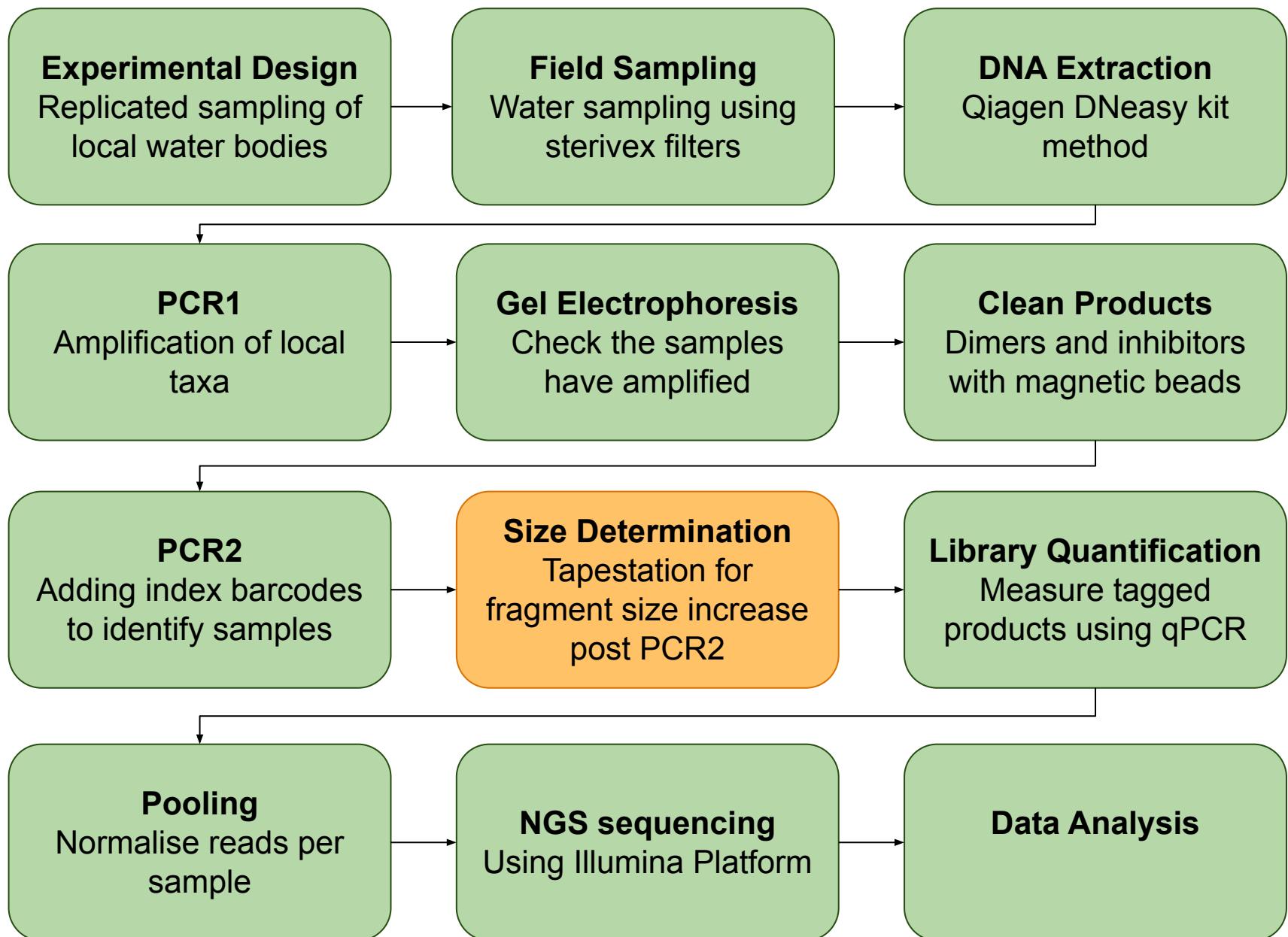
i7 index seq CAAGCAGAAGACGGCATACGAGATACTGGTCCGGTACTGGAGTTAGACGTGTGCTTCGATCT



- These sequences allow individual samples to be identified in the bioinformatic data produced post-sequencing

# PCR 2 - Another Bead Clean



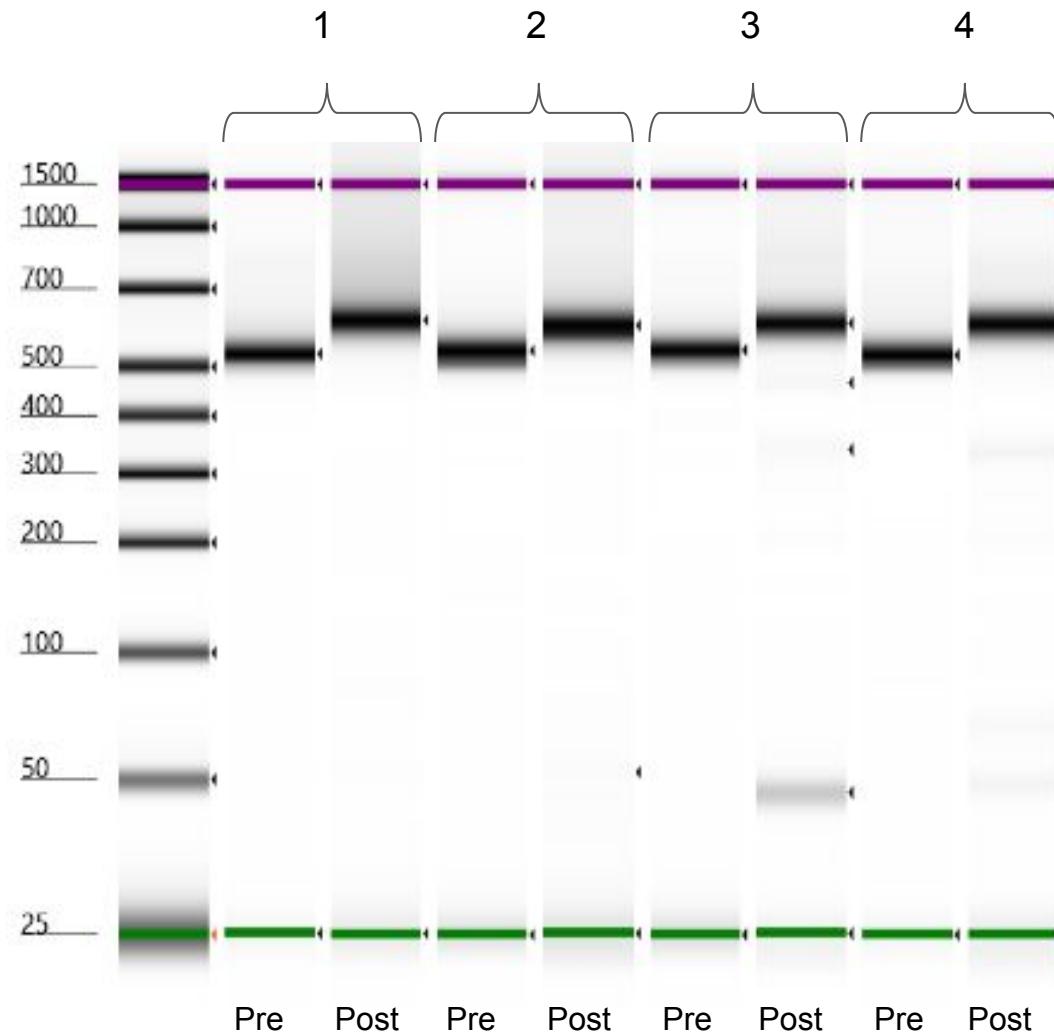


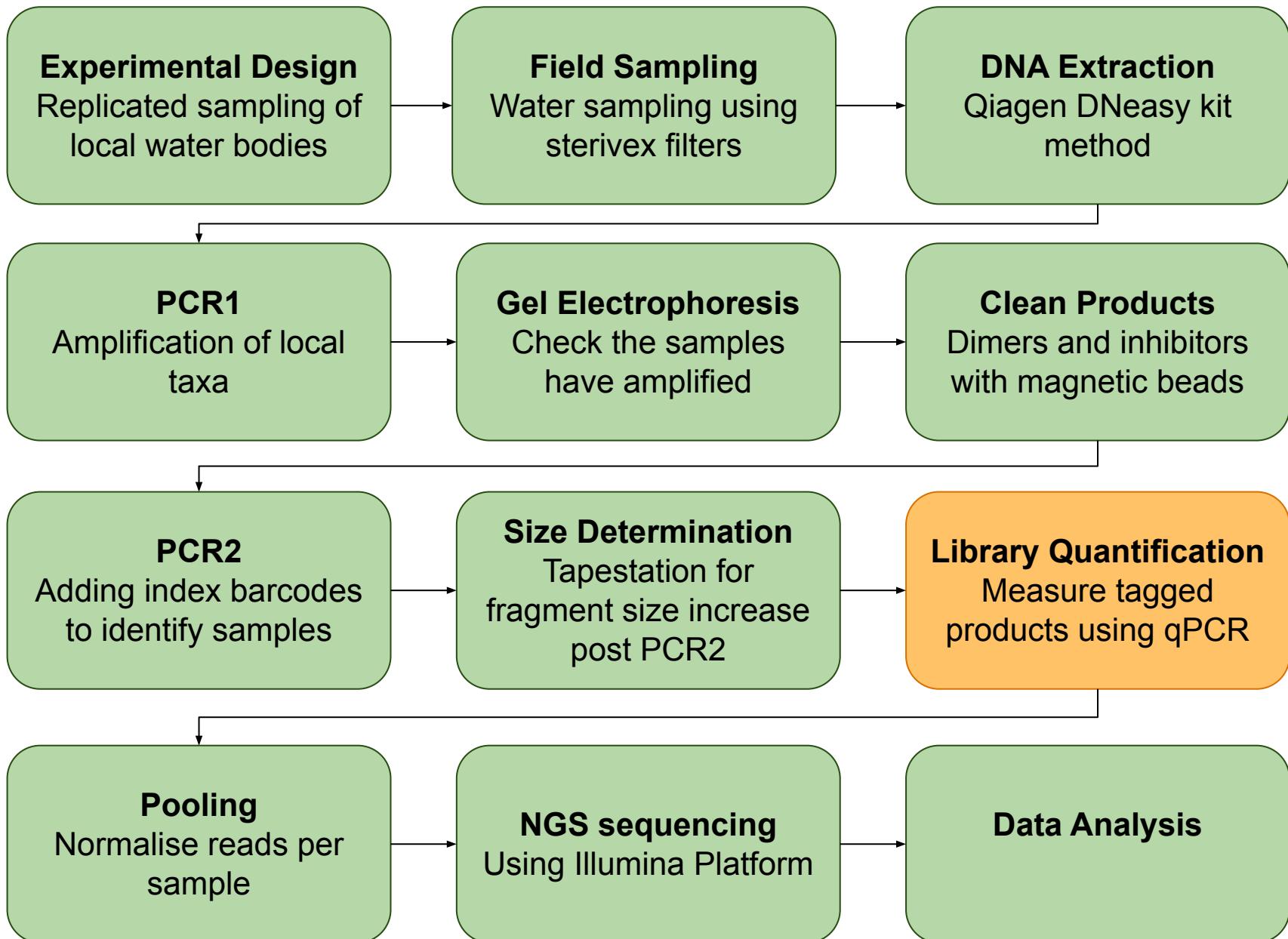
# PCR 2 - Tapestation



- Works like a miniature gel
- Compare samples pre and post-PCR 2
- Amplicons increase in size by length of unique identifier

# PCR 2 - Tapestation



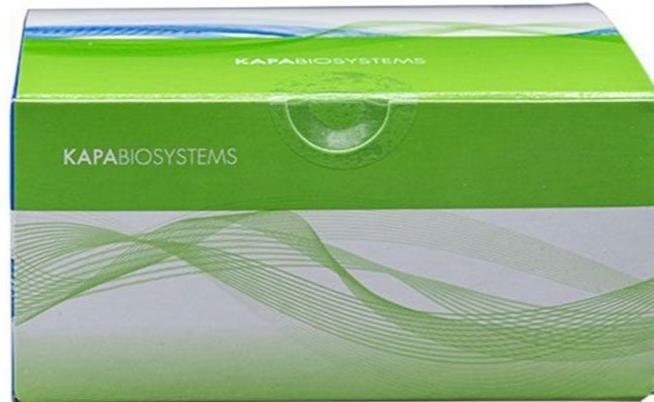


# Quantification - qPCR

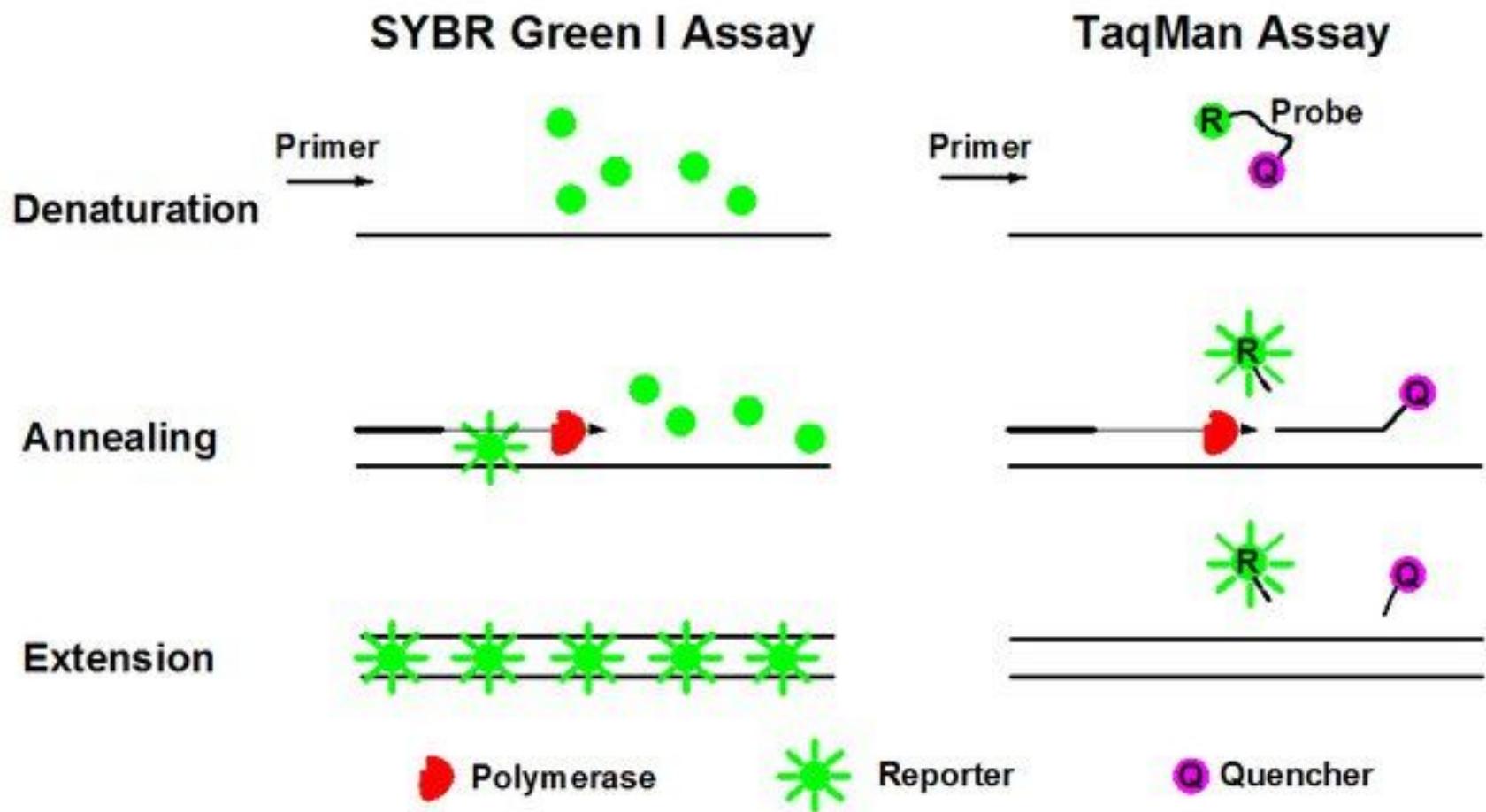


A commercial kit can quantify products that only have i5 and i7 indexes (identifier sequences) attached

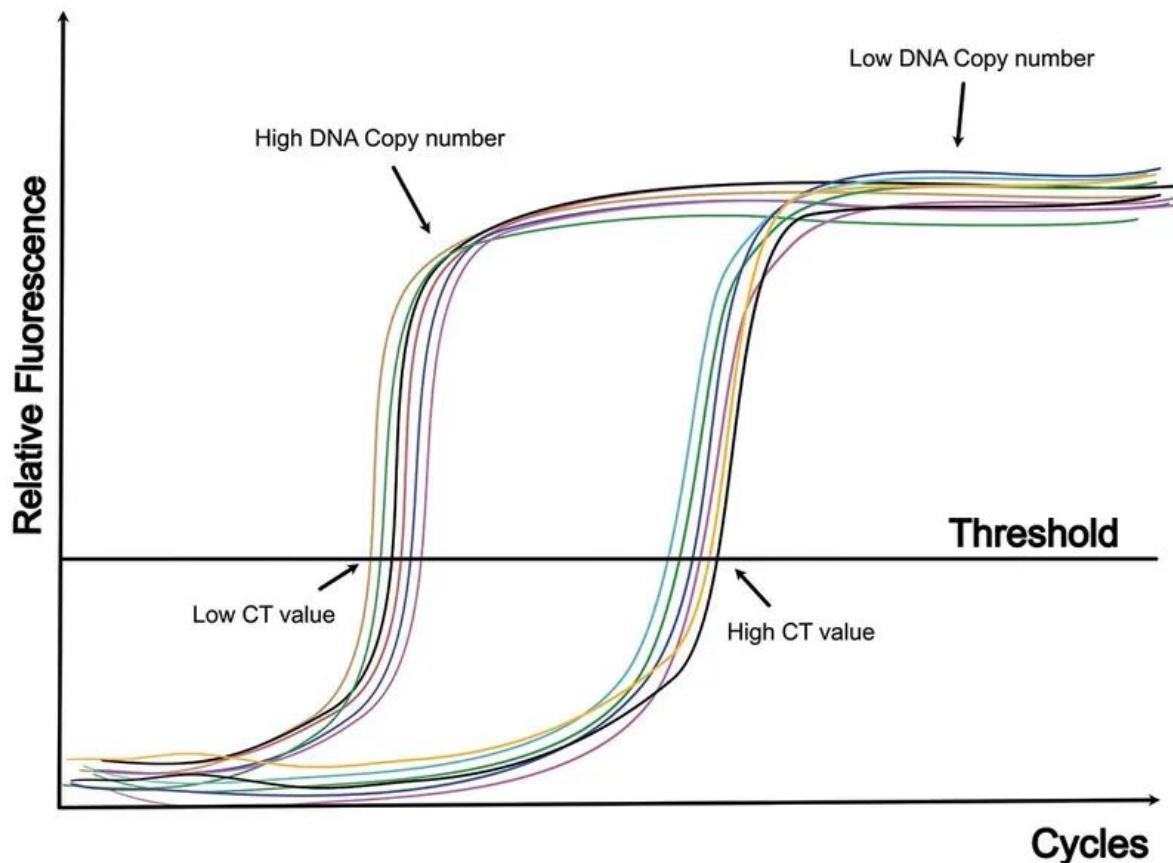
Due to the presence of unlabelled DNA from PCR2 we must quantify products by qPCR

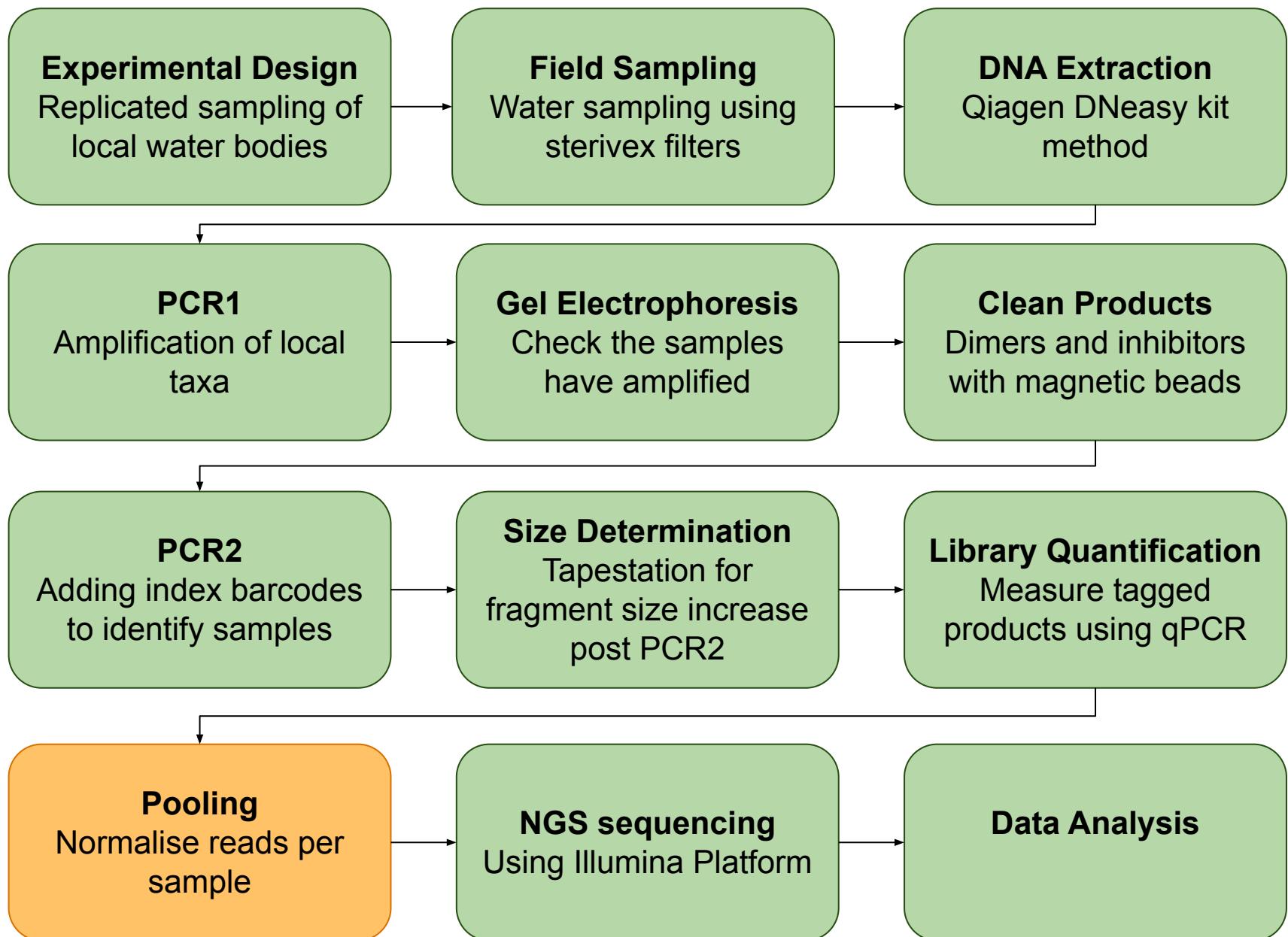


# qPCR?



## Amplification curve real-time PCR





**Pooling:** For an even representation of all samples in the final data

$$\frac{(\text{Final Volume } (\mu\text{l}) \times \text{Final Concentration } (\text{ng}/\mu\text{l}))}{(\text{Number of Samples} \times \text{Initial Concentration } (\text{ng}/\mu\text{l}))} = \text{Volume to Pool } (\mu\text{l})$$

$$X02 = 13.39$$

$$X03 = 6.87$$

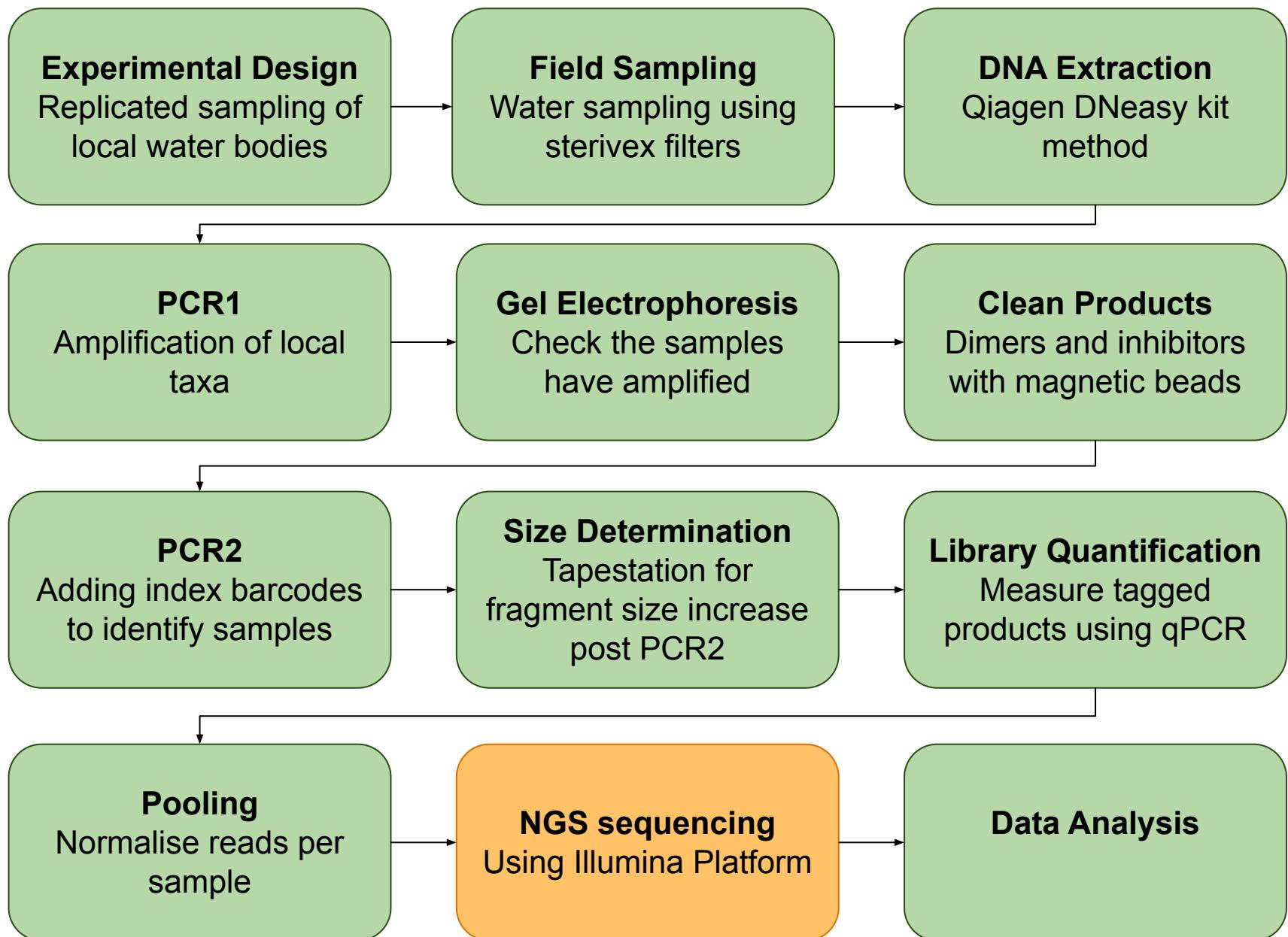
$$X04 = 8.31$$

$$X05 = 3.66$$

$$X06 = 5.81$$

$$\text{Total Volume to Pool } (\mu\text{l}) = 42.87$$

$$\text{Volume of H}_2\text{O } (\mu\text{l}) = 7.13$$



# Sequencing - Next Generation Sequencing

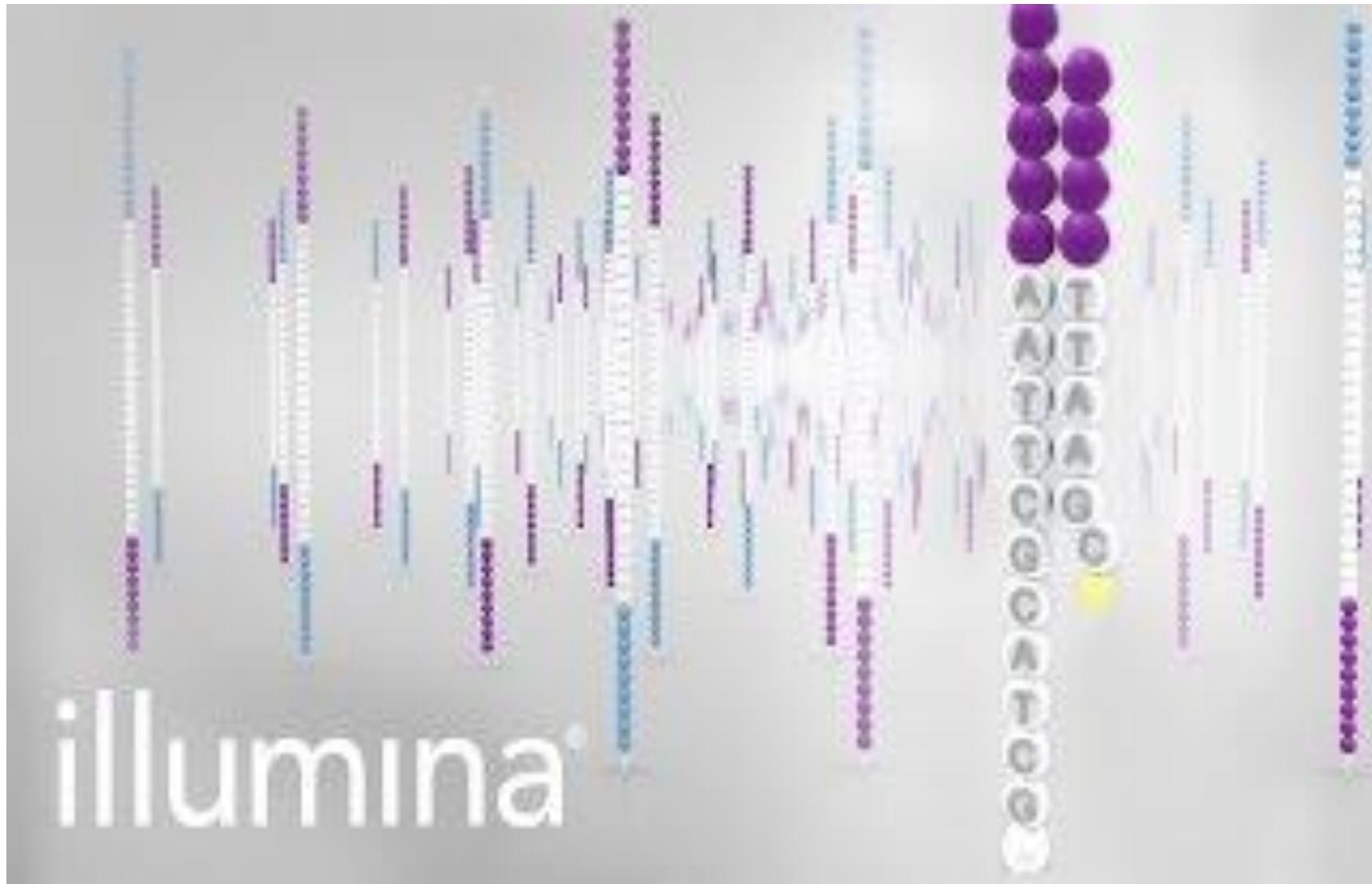


PacBio

# Sequencing - illumina®



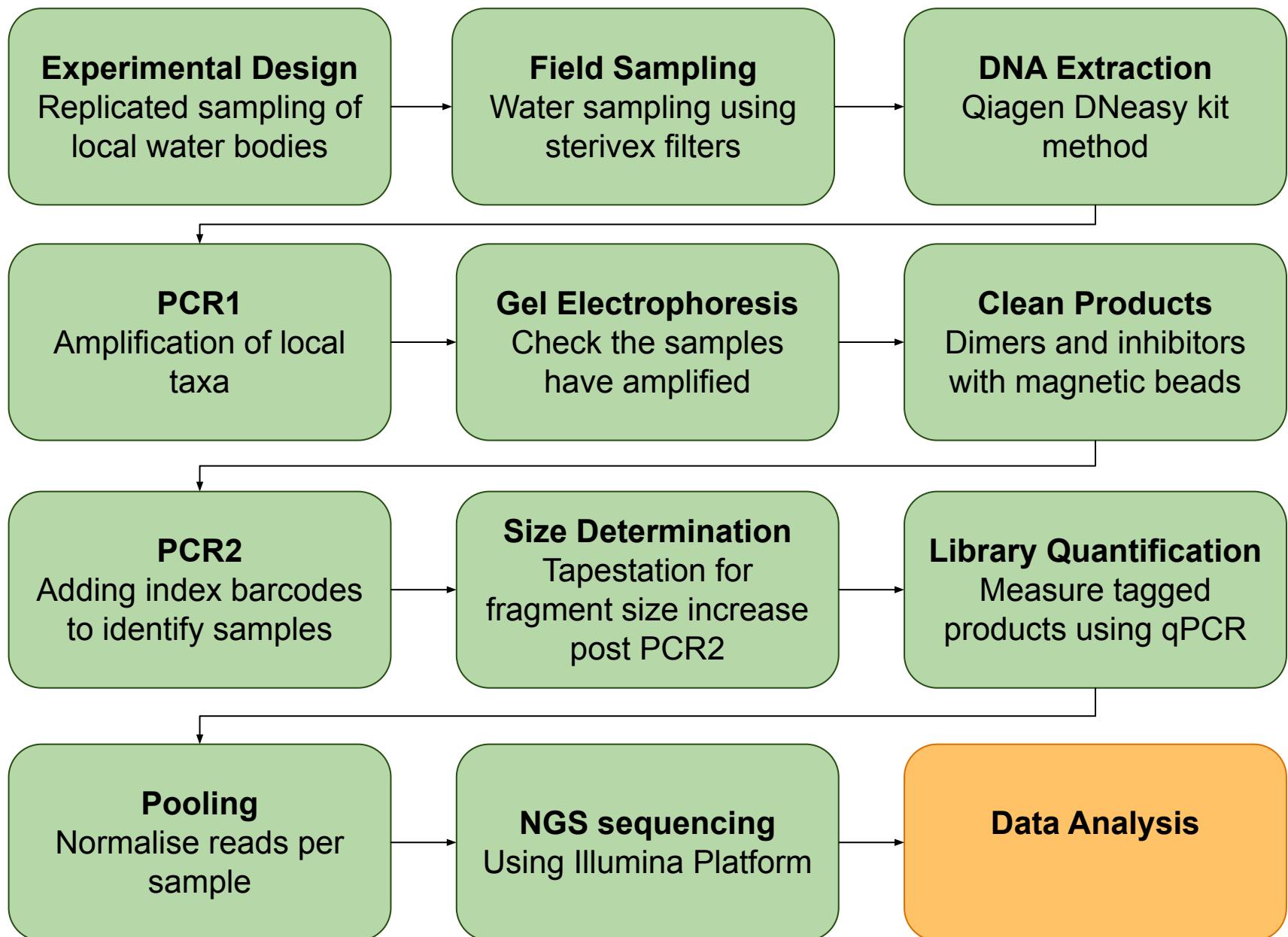
# Sequencing - illumina®



# Sequencing - illumina®

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.



# Previous eDNA study

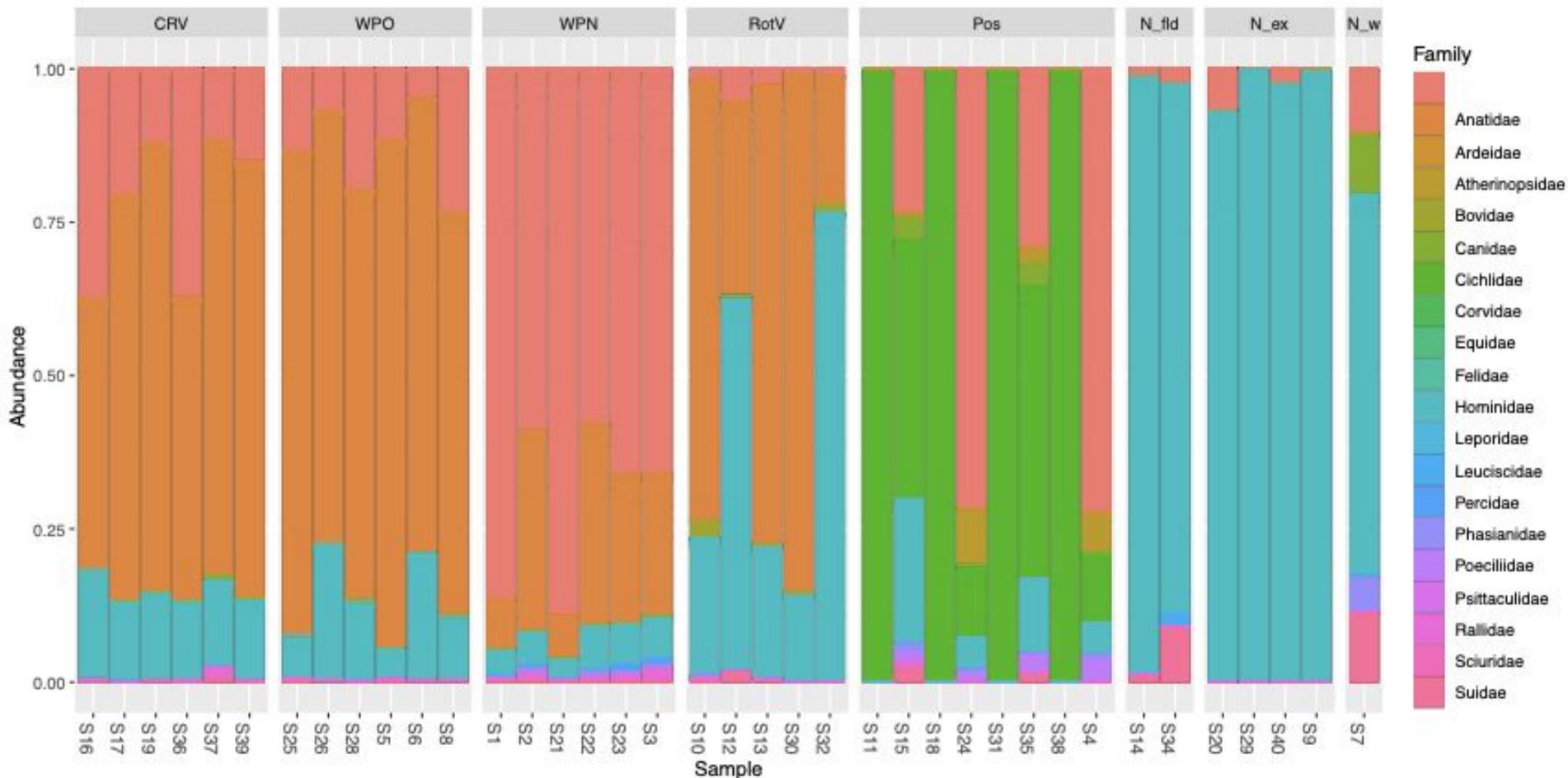
- Samples:
  - Weston Park November (**WP\_N**)
  - Weston Park October (**WP\_O**)
  - Rother Valley (**RotV\_O**)
  - Crookes Valley (**CRV\_O**)
  - Positive controls - Mexican river samples,  
African Cichlid stomachs (**Pos**)
  - Negatives
    - Field (**N\_fld**)
    - Extraction (**N\_ex**)
    - PCR (**N\_w**)
- Riaz 12S-V5 primer set  
(<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3241669/>)



# What did we find?



# What did we find?



## Take home message...

Consider this equation when planning your metabarcoding projects

$$\frac{\text{Average Yield from the Sequence Run}}{\text{Read Depth} \times \text{Number of Primers}} = \text{Number of Reactions}$$