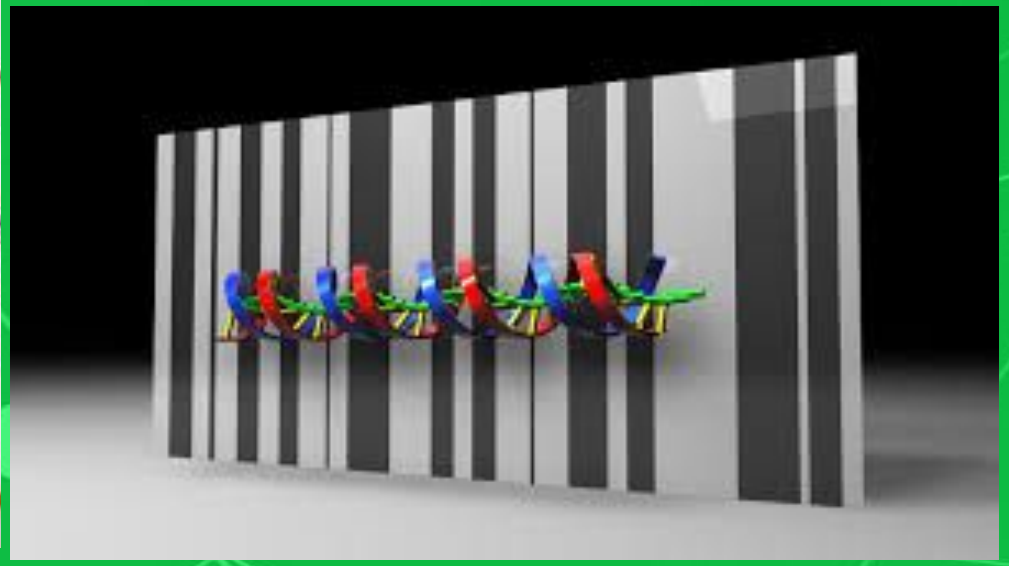
Centre for Genomic Research

UKRI Natural Environment Research Council

NEOF NERC ENVIRONMENTAL OMICS FACILITY

Metabarcoding for diet analysis and environmental DNA



UNIVERSITY OF LIVERPOOL

The University Of Sheffield.

Websites
NEOF: https://neof.org.uk/
NERC: https://nerc.ukri.org/
CGR:
https://www.liverpool.ac.uk/genomic-research/

Twitter
NEOF: @NERC_EOF
NERC: @NERCscience
CGR: @CGR_UoL

Upcoming workshops

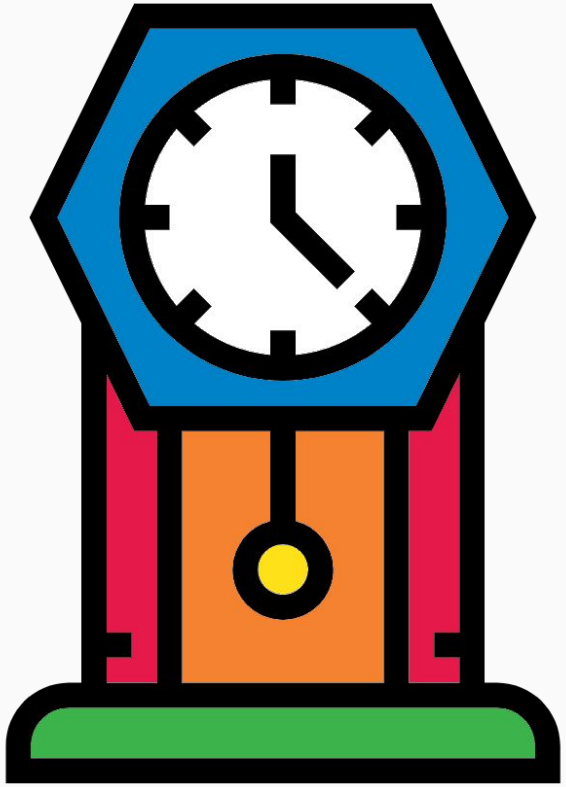
https://neof.org.uk/training/

- Microbial shotgun metagenomics
  - 21st & 23rd March 2023
- Eukaryote genome assembly - NEW!
  - 18th & 20th April 2023
- Community analysis in R - NEW!
  - 9th & 11th May 2023
- RNA-seq gene expression & pathway analysis
  - 6th & 8th June 2023
- More dates to be announced soon!

Format & Schedule

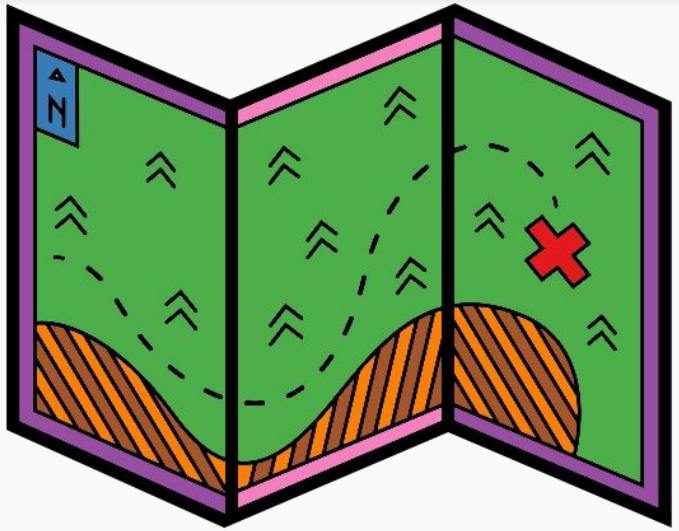
This intro
Bookdown
Theory
Practice
Exercises
Optional materials

Work at your own pace on your own time



Outline

- What is metabarcoding?
- Environmental DNA
- Diet analysis
- Lab methods & sequencing data
- Analysis pipeline
- Further analysis
- Using R
- Example data set

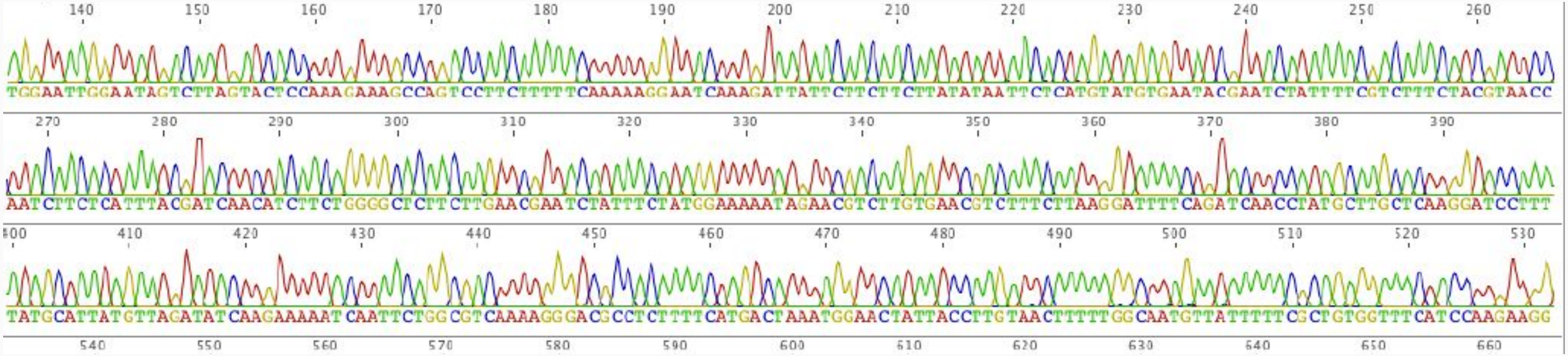


‘Ideal’ DNA barcode

- **High discrimination ability**: Low intraspecific divergence but high interspecific divergence (‘barcoding gap’)
- **Universality**: highly conserved priming sites and highly reliable DNA amplifications
- **Standardized**, with the same DNA region used for different taxonomic groups
- **Informative**: should contain enough phylogenetic information

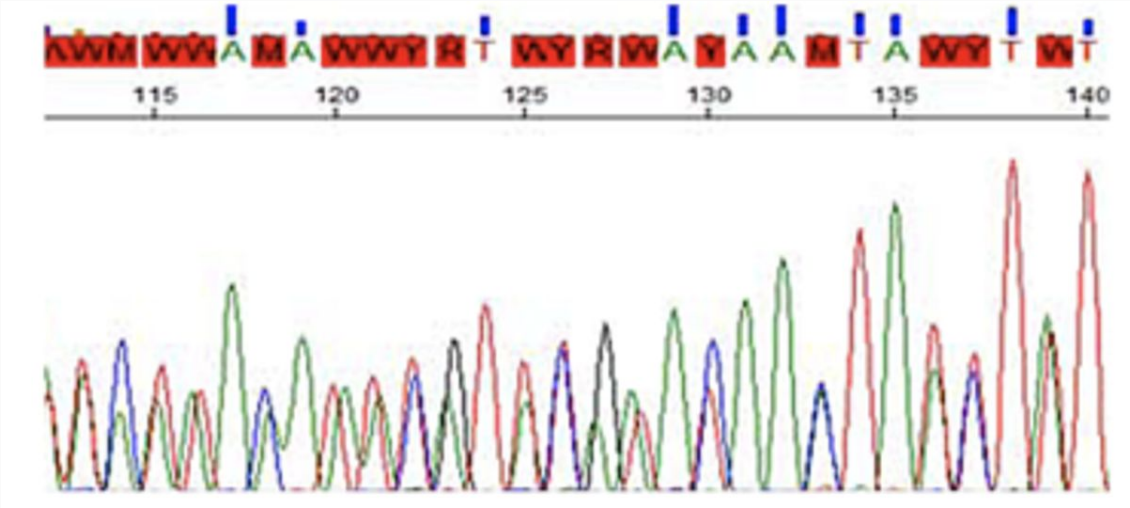
Single unknown sequences

- Use barcoding primers to amplify and sequence a short gene region for target group of taxa (e.g. mammals, plants)
- Sanger sequence & BLAST



Many unknown sequences

- Use generic barcoding primers to amplify short region of gene for target group of taxa
- Amplify mixed DNA, clone amplicons, sanger sequence & BLAST
    - OR
- Amplify mixed DNA, tag & pool samples, sequence (high-throughput sequencing), BLAST/classify for species ID

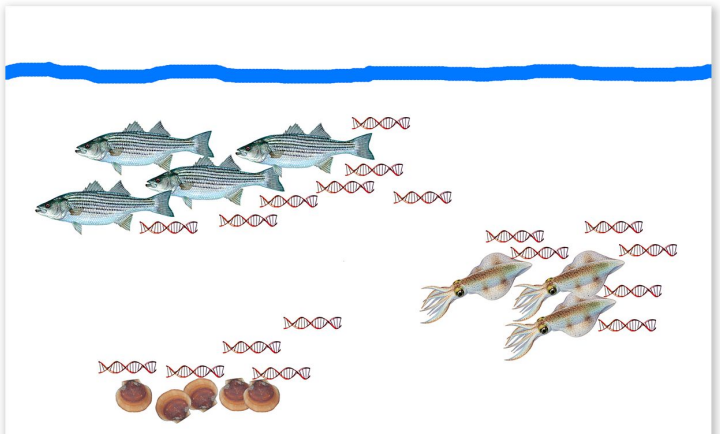


Environmental or complex DNA samples

- Trace DNA from soil, water, air
- DNA comes from skin, mucous, saliva, gametes, blood, pollen, fruit, urine, faeces
- Potentially degraded
- DNA from animal blood meals, faeces, stomach contents
- DNA from traded ‘complex’ products, e.g. herbal supplements, teas, fish paste

UK species identified as mosquito feeding hosts

Surimi



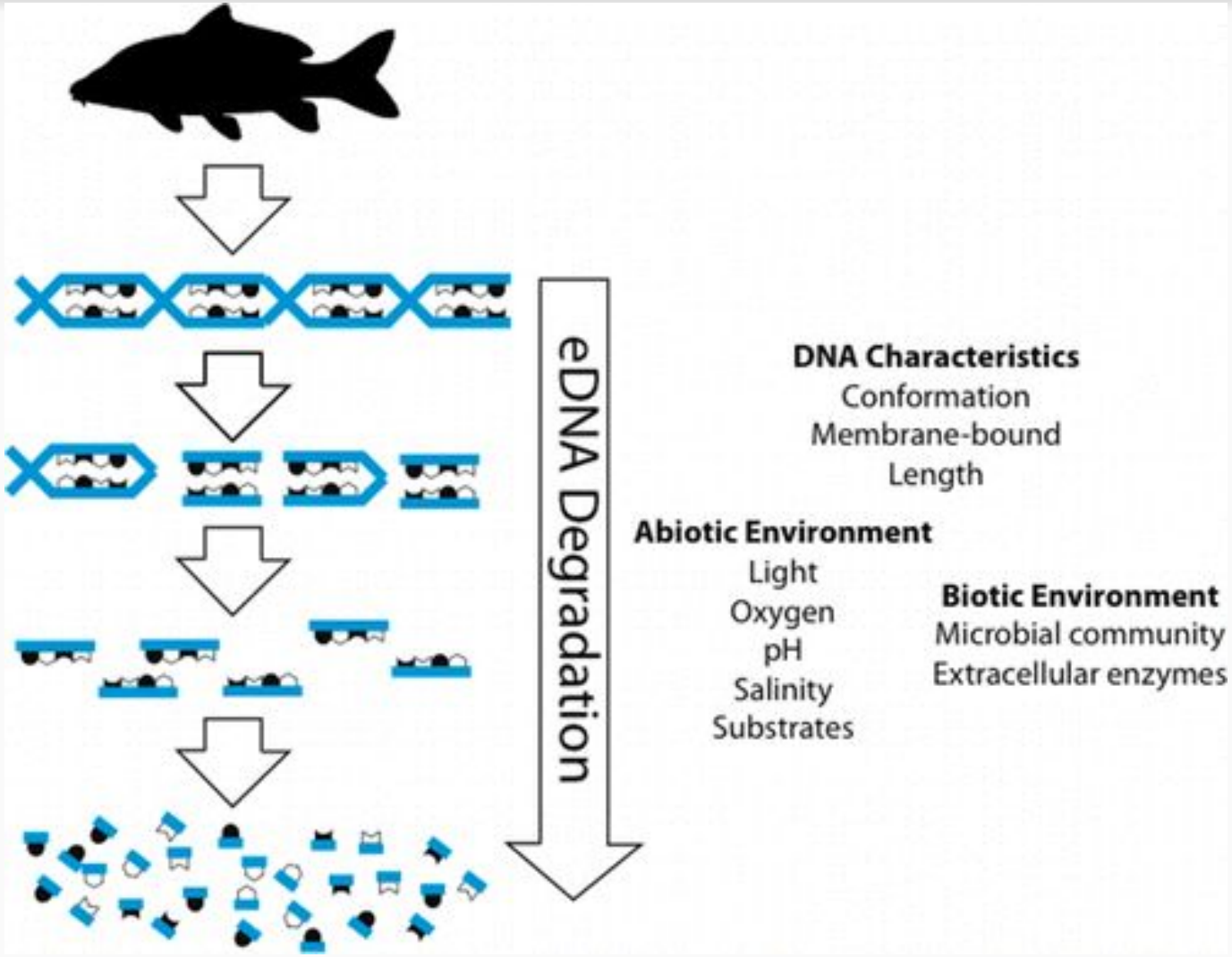
Water samples will contain mixed species DNA

Why study macrobial eDNA?

- Invasive species – early detection/detect low density populations
- Rare species – detect low density populations
- Analysis of diet (is this strictly speaking eDNA?)
- Difficult to identify or survey species
- Historical biodiversity data from frozen sediments
- Often quicker & cheaper!

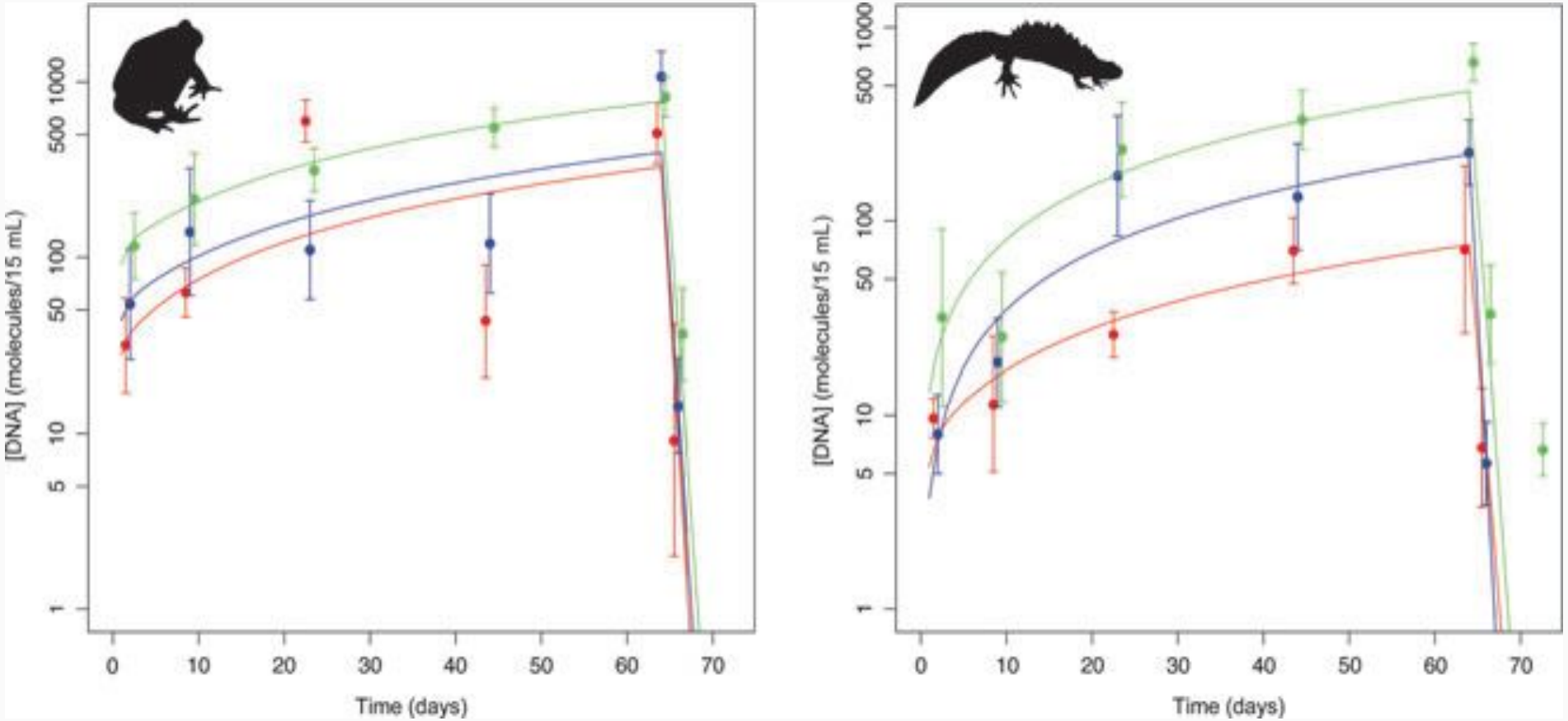
Technical challenges

- DNA can be in short fragments & low abundance – amplify short fragments, target taxa of interest
- How long does DNA persist in the environment/complex samples?

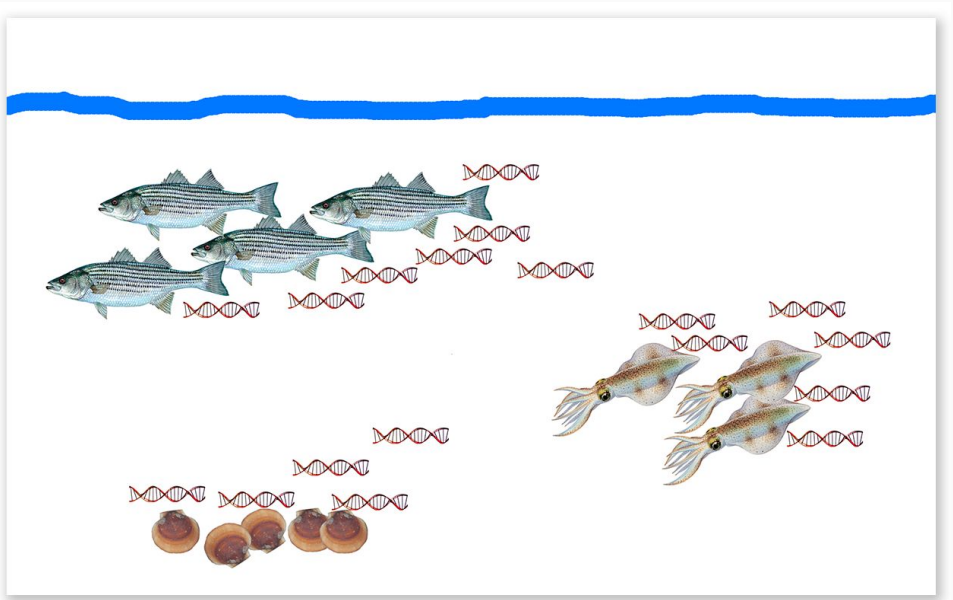


DNA degradation

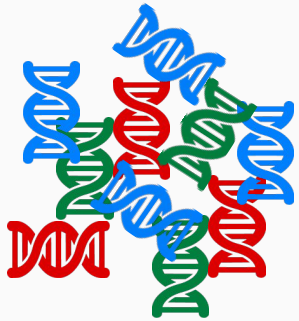
- eDNA from water indicates recent presence of organism (Thomsen *et al.* (2012) Mol Ecol.)



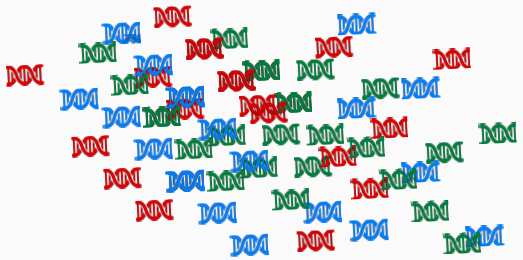
DNA metabarcoding methods, e.g. Cyt-b barcode from sea water



Mixed DNA from water sample

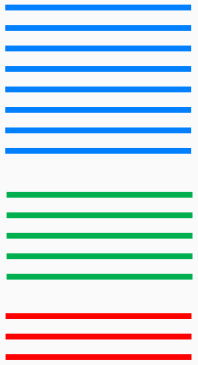


Amplify Cyt b marker



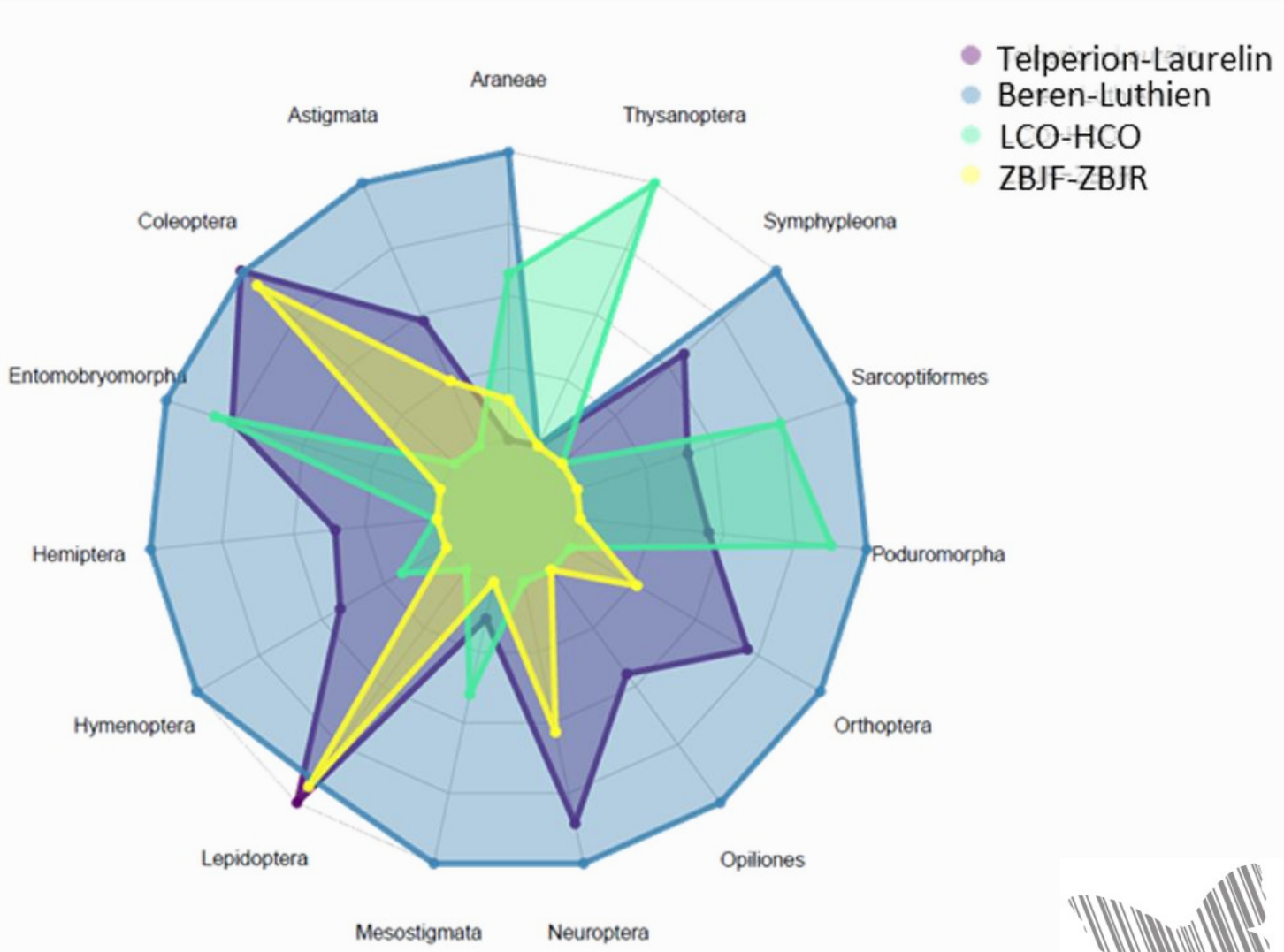
High-throughput sequencing

Bioinformatics and taxon ID



Primer choice

- Which primer pair(s) to use?
- Maximise taxa of interest & minimise those not (e.g. predator DNA)
- Comparison software available – e.g. PrimerMiner (https://github.com/VascoElbrecht/PrimerMiner)
- A good discussion of the considerations: https://biocoenosis.org/2021/03/29/primer-time/



Massively parallel sequencing

- Much higher degree of parallelism than Sanger sequencing – sequence millions of DNA fragments simultaneously

- Much lower costs

- Different platforms differ in terms:
  - read lengths
  - bp output
  - costs of run
  - costs of library preparation
  - error rates

www.illumina.com

www.pacb.com

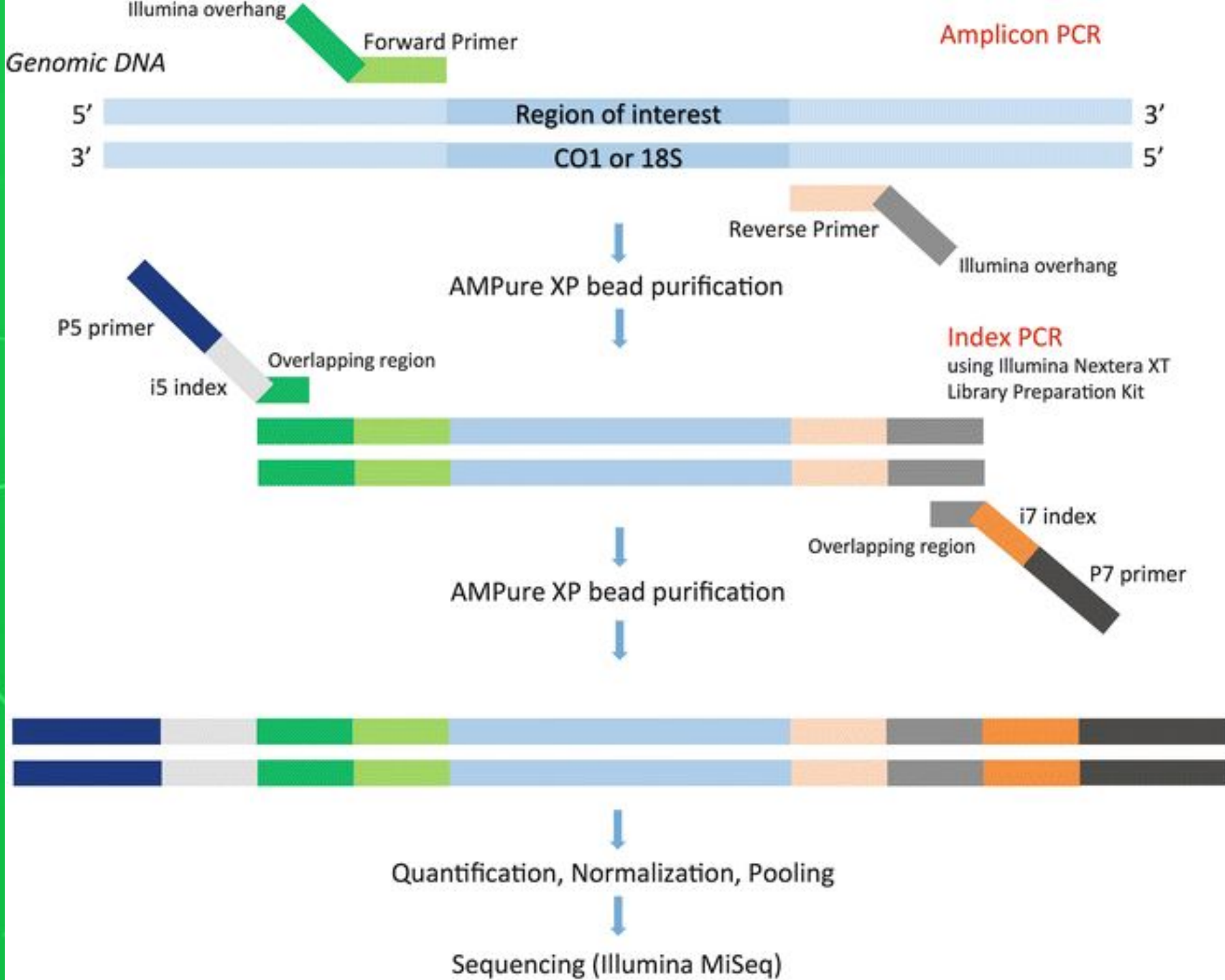
‘Homemade’ library preparation for metabarcoding

- Need to pool and track large numbers of samples, sometimes for multiple genes/amplicons.

- Two main approaches:
  - 1. PCR amplicons using MID-tagged primers, prepare library with blunt-end ligation of Illumina adapters.
  -
  - 2. PCR amplicons using overhang primers, prepare library using PCR to attach Illumina adapters.

Preparing DNA for metabarcoding using Illumina sequencing

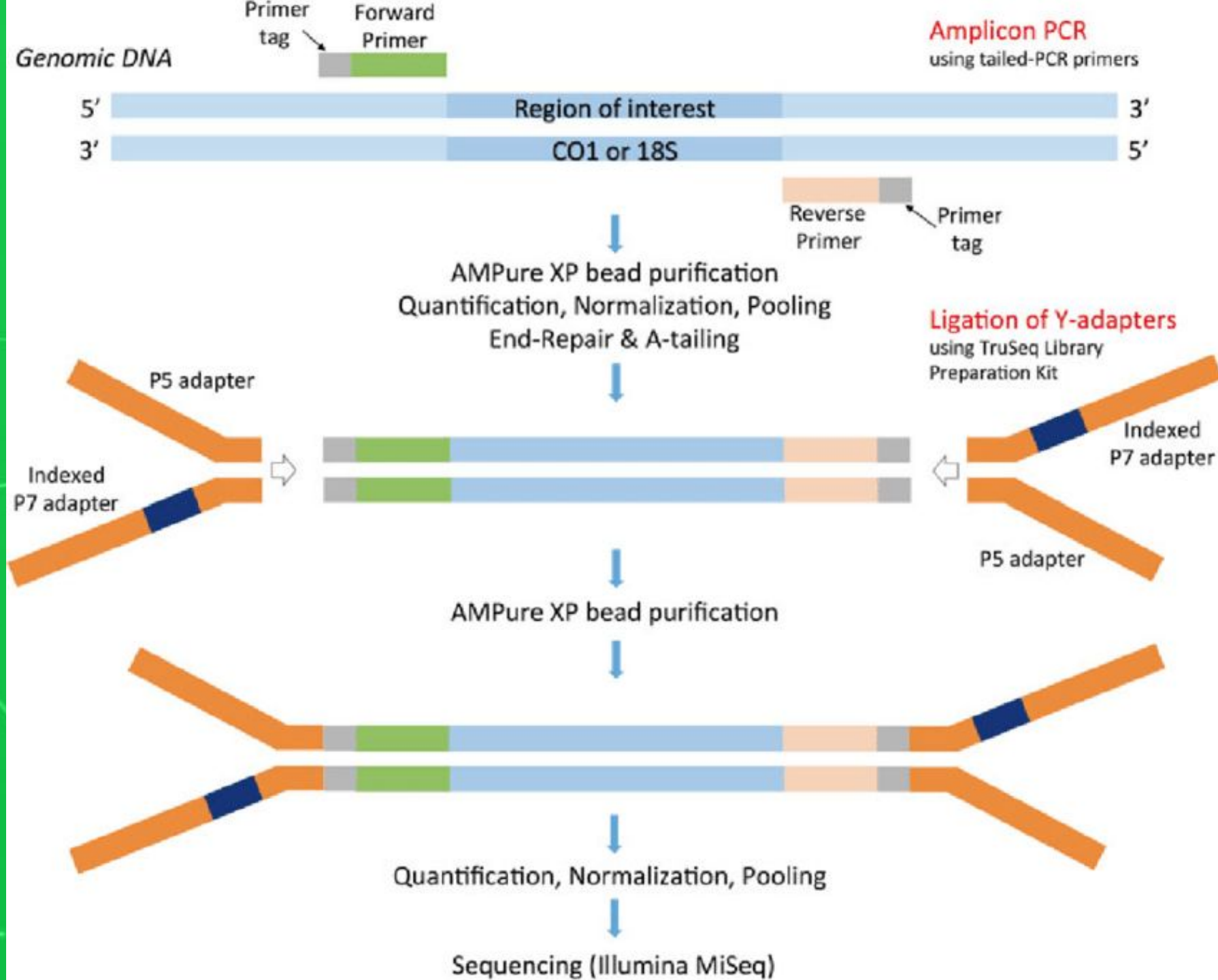
Dual PCR



Bourlat et al. (2016) Preparation of Amplicon Libraries for Metabarcoding of Marine Eukaryotes Using Illumina MiSeq: The Dual-PCR Method

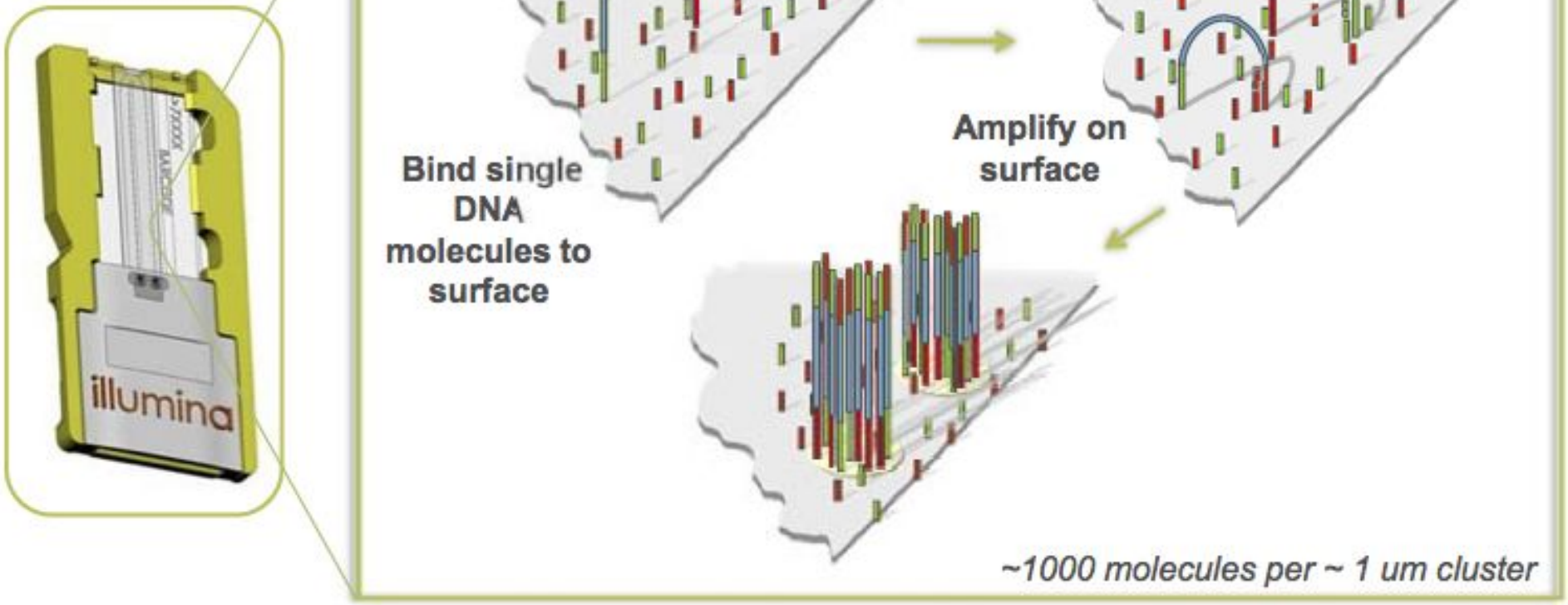
Preparing DNA for metabarcoding using Illumina sequencing

Adapter ligation



Leray et al. (2016) Preparation of Amplicon Libraries for Metabarcoding of Marine Eukaryotes Using Illumina MiSeq: The Adapter Ligation Method

Cluster Generation

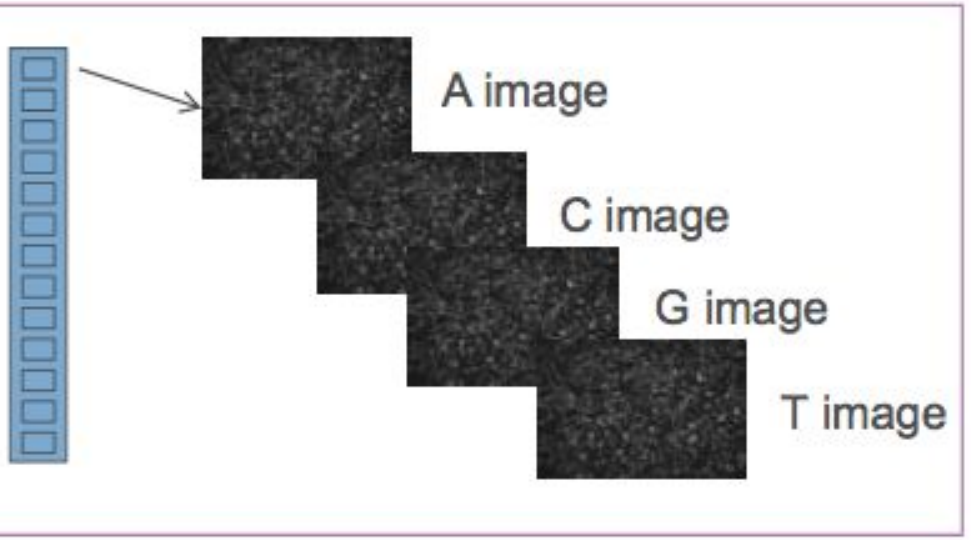


Sequencing

Clusters are images using LED and filter combinations specific for each fluorescently-labeled nucleotide

After imaging is complete for one section (tile), the flow cell is moved to the next tile and the process is repeated

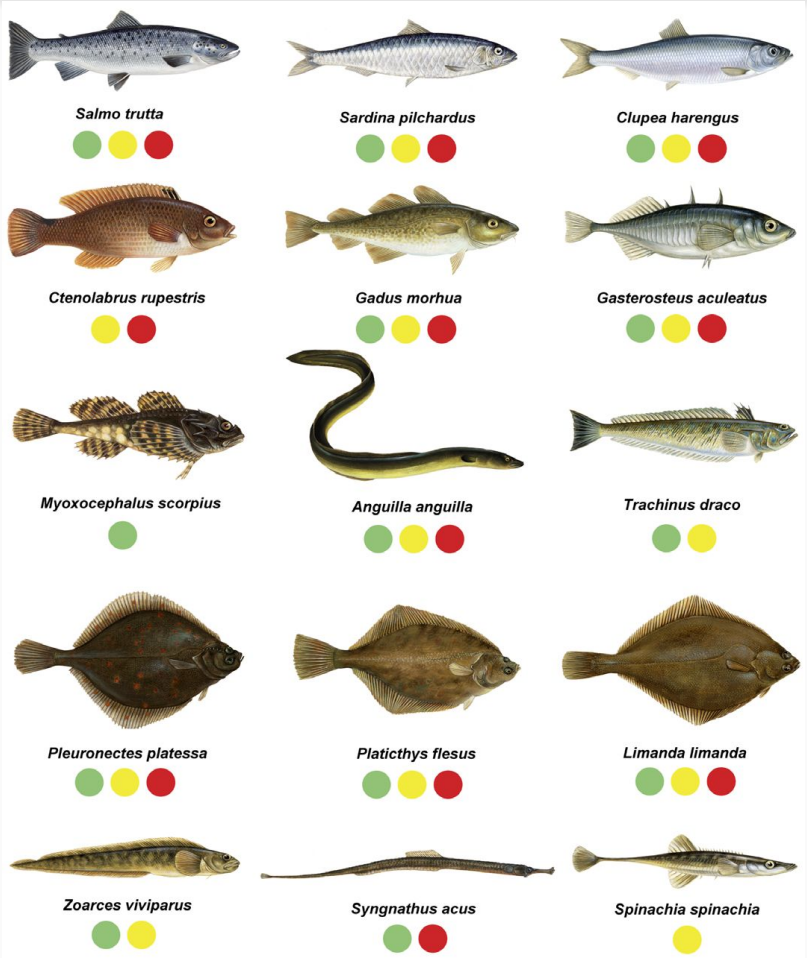
Imaging for the 1st cycle takes ~3 min., including focusing routines



A video of Illumina sequencing workflow: https://youtu.be/fCd6B5HRaZ8

How does eDNA compare to traditional surveys?

- Seawater success
  - Detection of fish species
  - Including species never/rarely detected
  - Thomsen *et al*. (2012) PLoS One

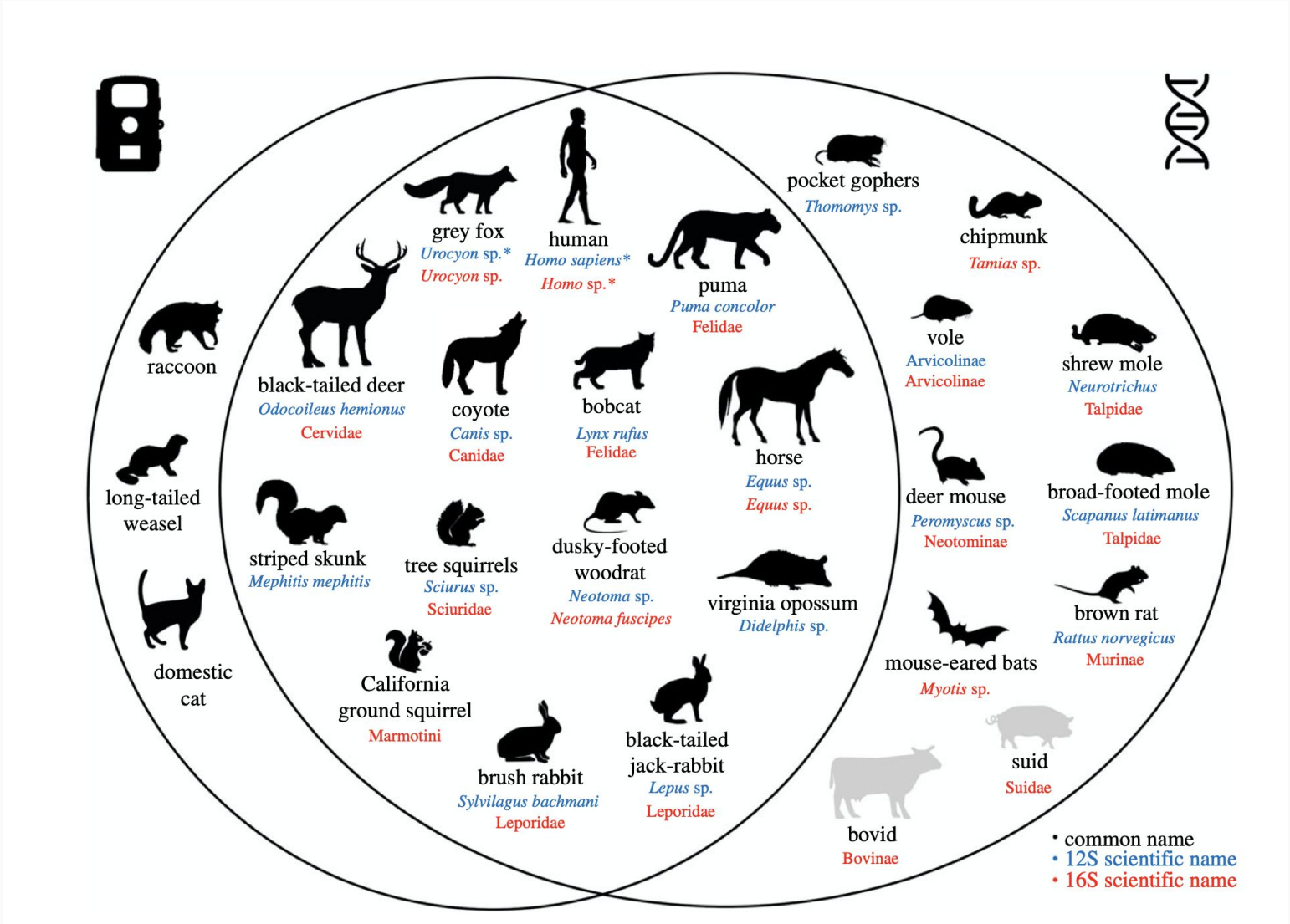


How does eDNA compare to traditional methods?

Terrestrial success

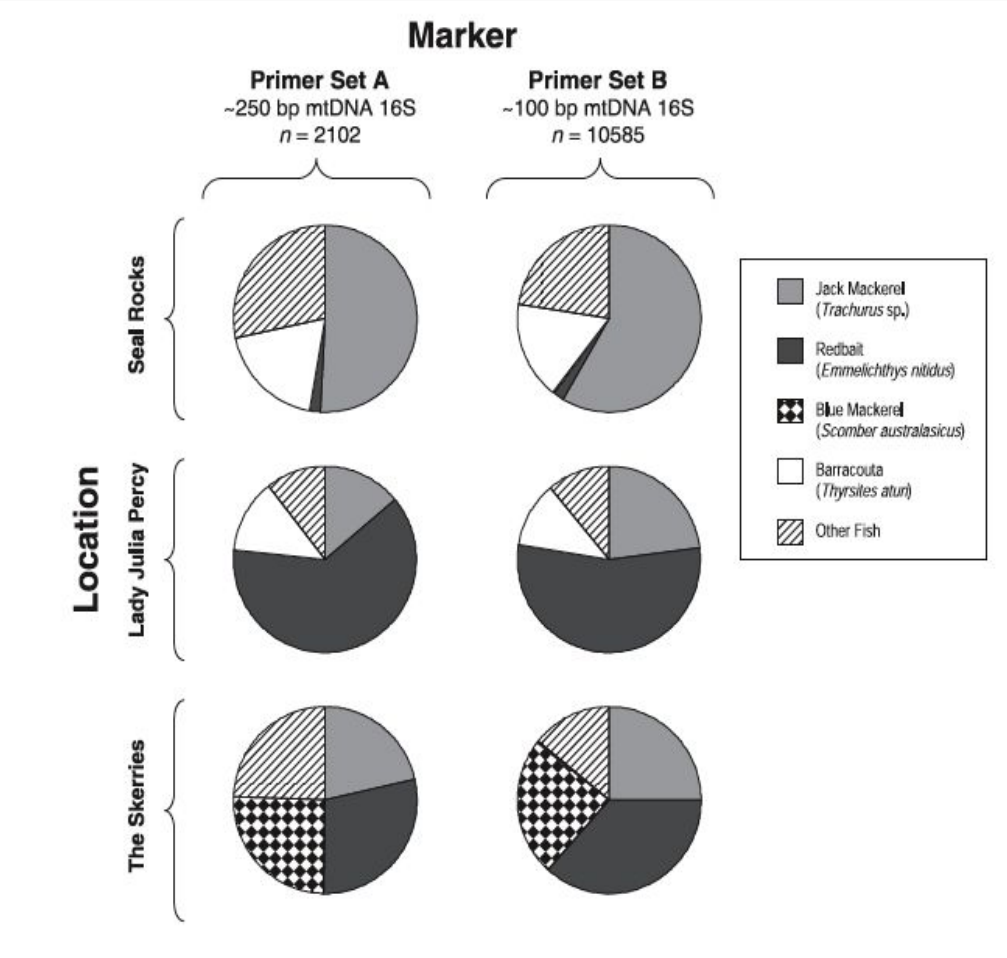
- Detection of mammal species from soil
- Long term camera trapping data
- Including species never/rarely detected

Leempoel et al. (2020) Proc. R. Soc. B

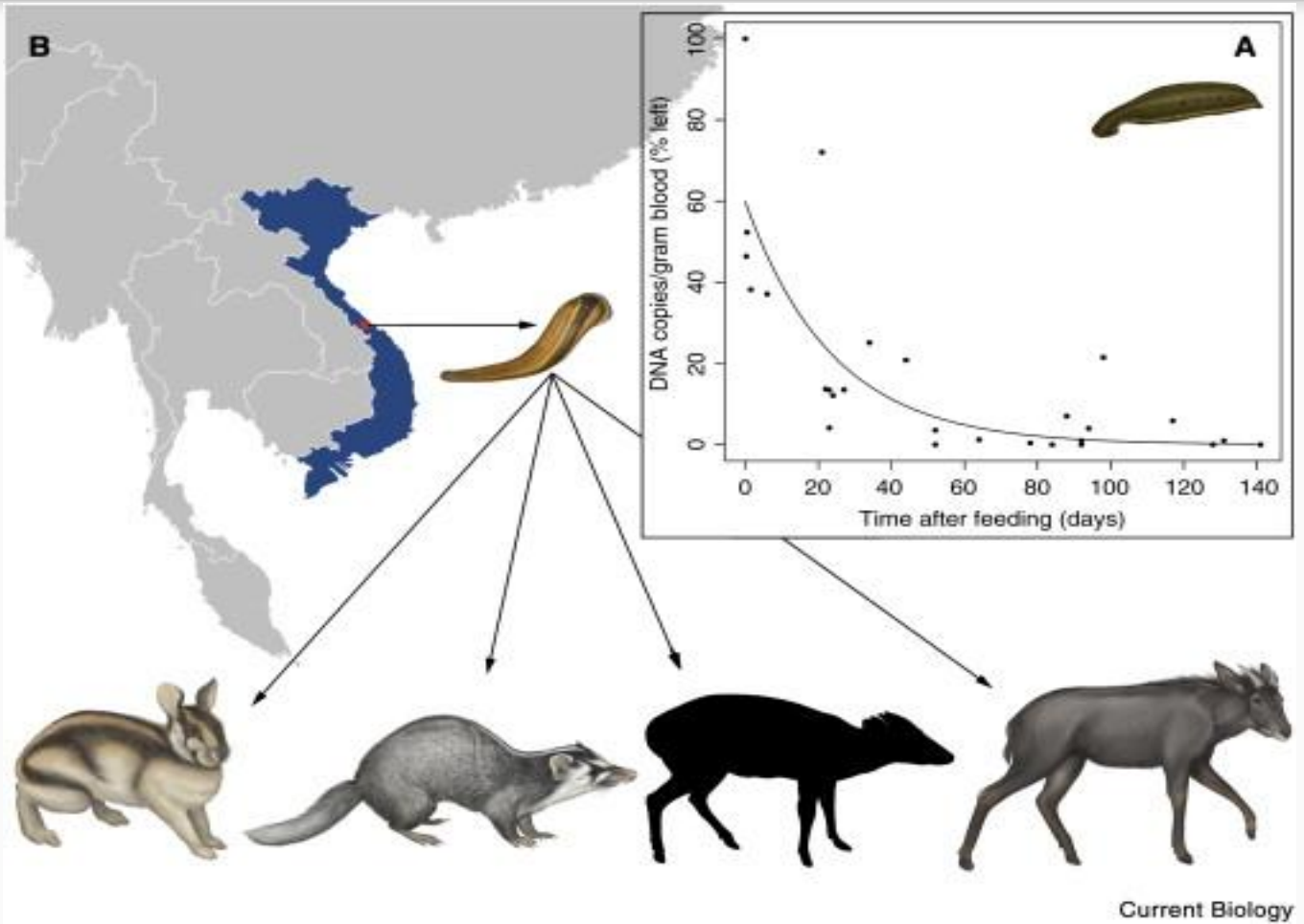


Diet Analysis

- Deagle et al. (2009) pyrosequenced 270 fecal samples
- 58 different species, higher diversity than traditional analyses
- Geographic variation in diet composition



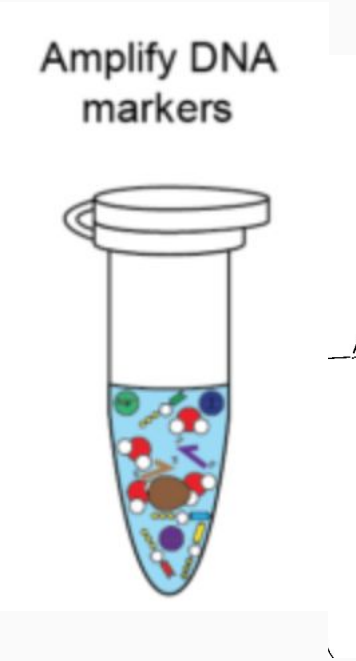
Detecting rare mammals from leech blood meals in Vietnam

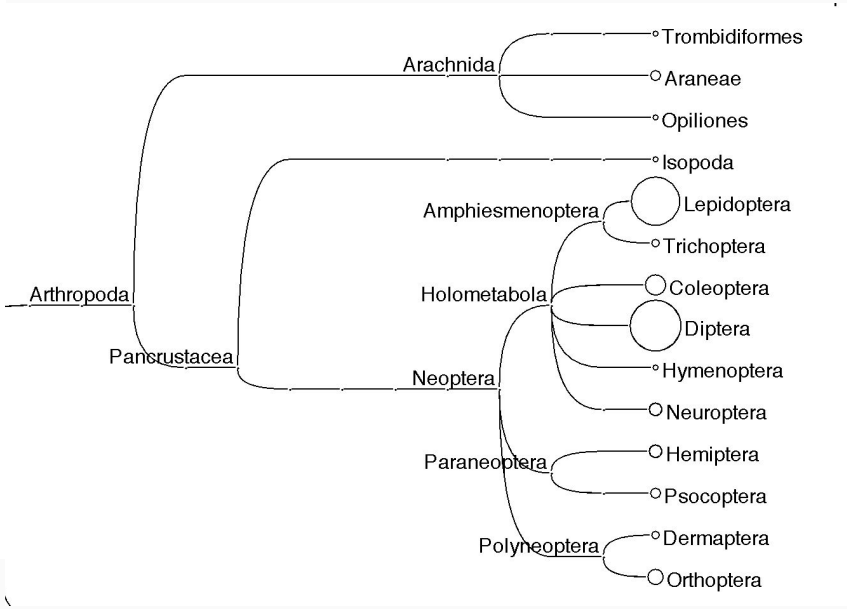


Schnell *et al.* (2003)

Metabarcoding on bulk samples ...

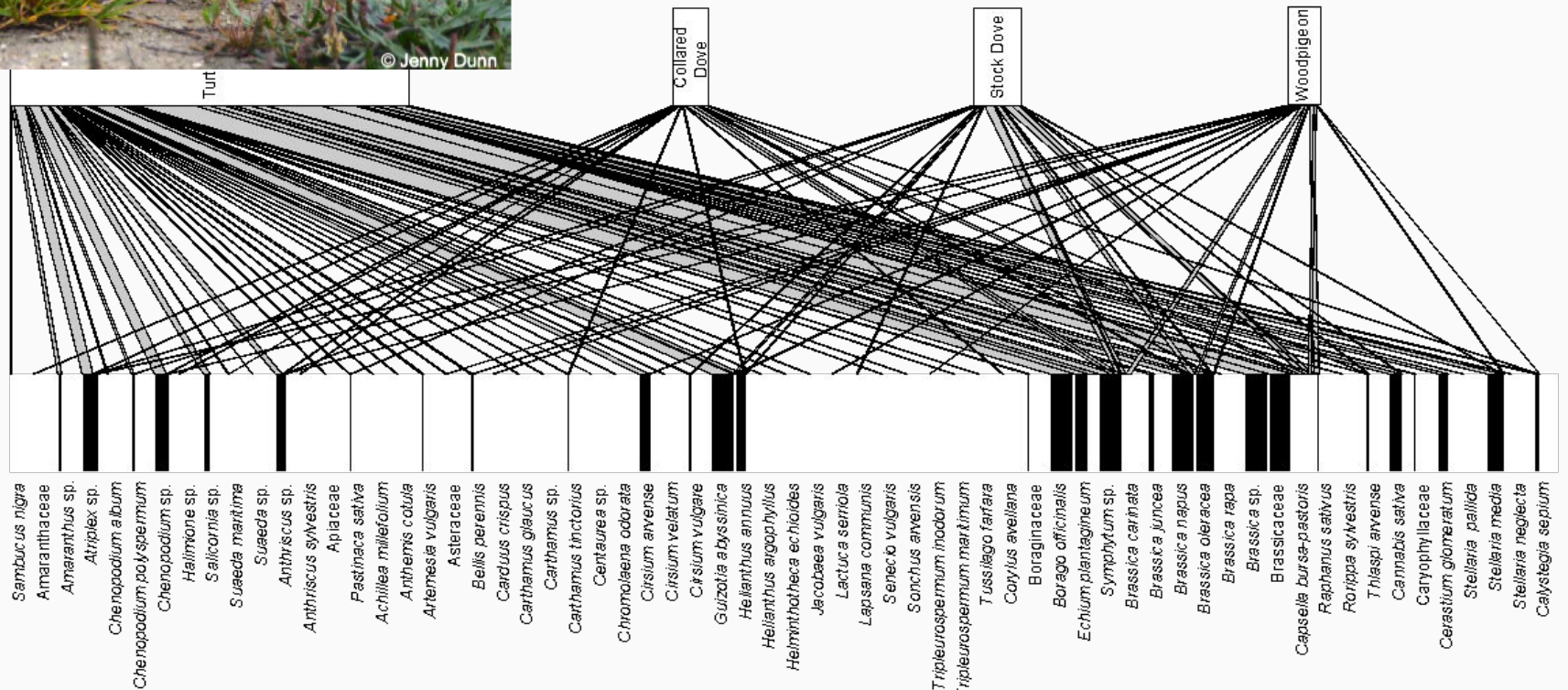
What are the consequences for biodiversity of different farming practices?





Declining UK Turtle Doves have high dietary overlap with other pigeons, especially Stock Doves, and are using anthropogenic food sources on farms

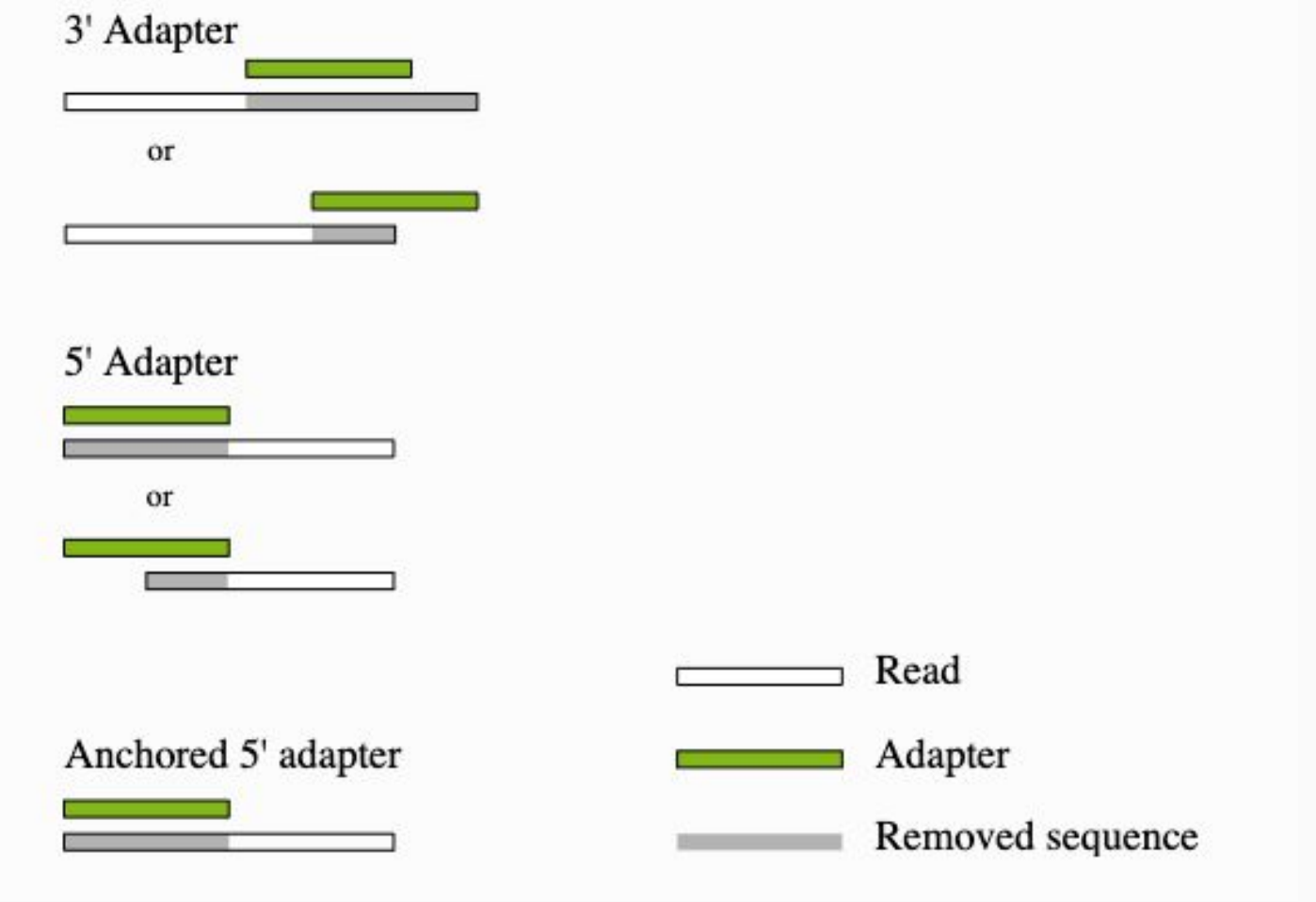
Dunn et al (2018)
https://onlinelibrary.wiley.com/doi/full/10.1111/mec.14766



- Divisive Amplicon Denoising Algorithm
- Models and corrects Illumina-sequenced amplicon errors
- Pipeline
  - Filtering
  - Dereplication
  - Merging paired-reads
  - Chimera Identification
  - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4927377/

Cutadapt

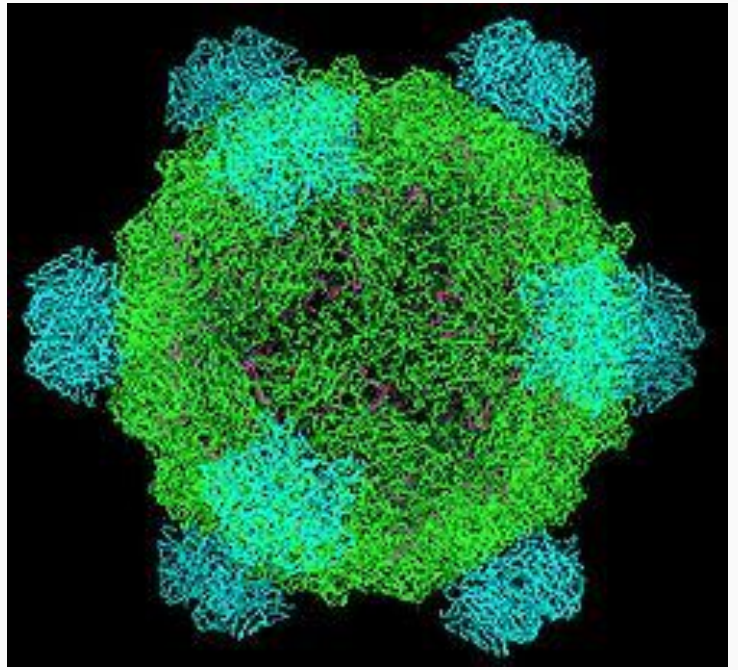
- DADA2 requires that primers are removed
- A flexible tool for removing known sequences (primers/adapters)
- Online docs: ‘recipes’ for different library prep/sequencing methods



https://cutadapt.readthedocs.io/en/stable/index.html

Filtering & Dereplication

- Remove sequences containing ‘N’ bases
- Remove poor quality sequences
- Remove sequences similar to PhiX
- Optional minimum length filtering



https://en.wikipedia.org/wiki/Phi_X_174

Merging reads

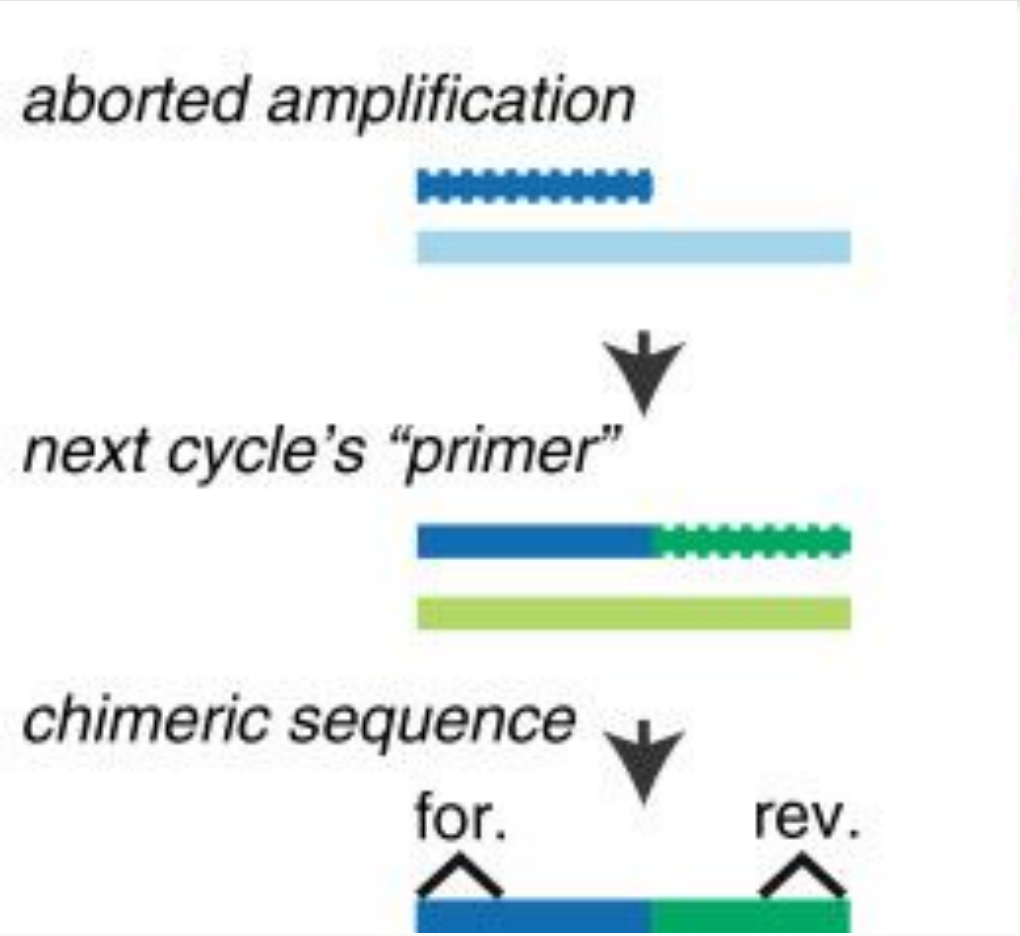
- Align read pairs R1/R2 to each other
- Improves quality of reads
- Longer reads

R1 (30bp)	ACCGTACGTATGCGTAGCTGACGTAGCATG----------
R2 (30bp)	----------TGCGTAGCTGACGTAGCATGCGCGATTCGA
Overlap (20bp)	----------TGCGTAGCTGACGTAGCATG----------
Stitched read (40bp)	ACCGTACGTATGCGTAGCTGACGTAGCATGCGCGATTCGA

Chimera removal

DNA from two or more parent molecules

- PCR artifact
- Erroneous “novel” sequence



OTUs vs Amplicon Sequence Variants (ASV)

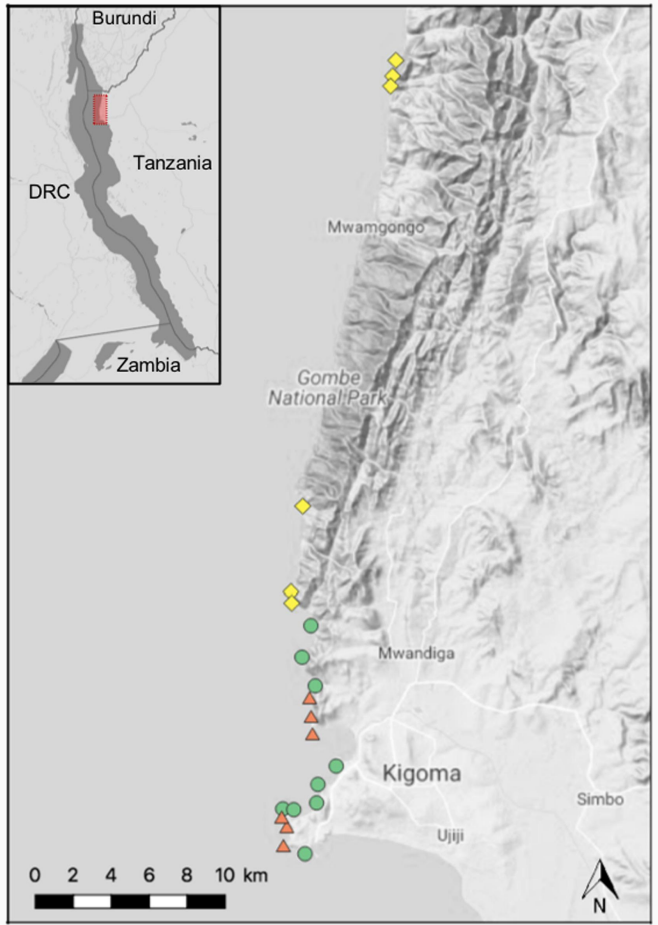
- Previous methods
  - OTUs (Operational taxonomic units)
  - Cluster sequences by identity e.g. 97% similarity for species
  - Can cause over clustering
  - 97% is chosen due to errors within Illumina data
- DADA2
  - Denoises and cleans reads so they represent real sequences
  - Much finer resolution on sequences
  - Can differentiate sequences that have only 1bp difference

Next steps with ASVs

- Assign taxonomy
  - NCBI Blast search or create a custom reference database
- Rarefaction
  - Enough sequencing depth?
- Alpha diversity – how many species are there and how evenly abundant are they within a sample?
- Beta diversity – how different is the species composition in different samples/groups?

- We will work in the R environment within the VNC
- Command line interface, also displays plots
- Many packages available for bioinformatics
- VNC:
  - Packages are installed
  - Code is provided in the workbook for you to type in
  - Further information on R (including further tutorials) provided in the workbook

Testing the performance of environmental DNA metabarcoding for surveying highly diverse tropical fish communities: A case study from Lake Tanganyika



- Used 4 primer pairs to assess effectiveness of metabarcoding to survey tropical fish
- Example data set to work on this week
  - subset of 31 samples, including replicates from several sites and negative controls
  - two primer pairs targeting 12S
  - Illumina 2 x 150bp MiSeq data
- Doble et al. (2019) https://onlinelibrary.wiley.com/doi/full/10.1002/edn3.43

Reminders and Tips

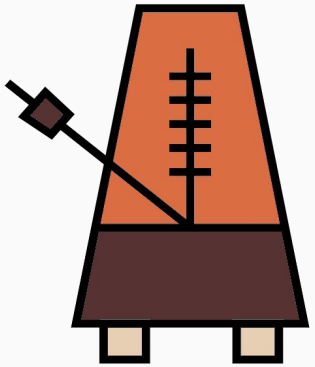
Work at your own pace

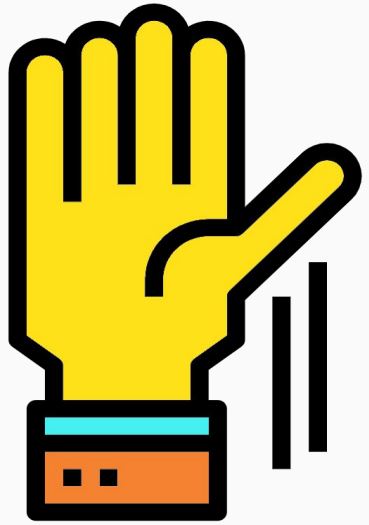
Typos

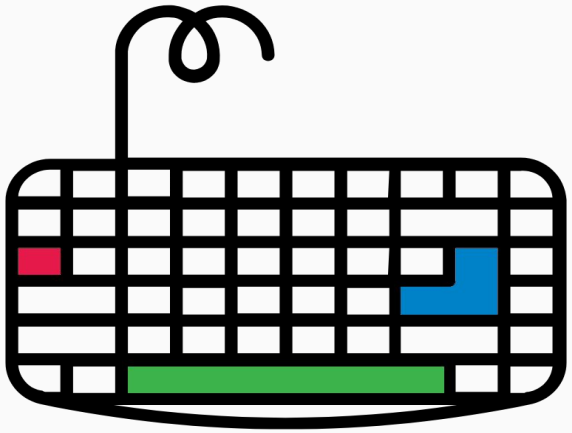
Ask questions

Breaks are important

Tab, space, and enter

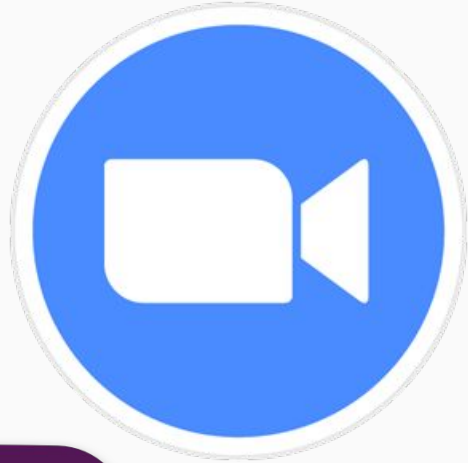






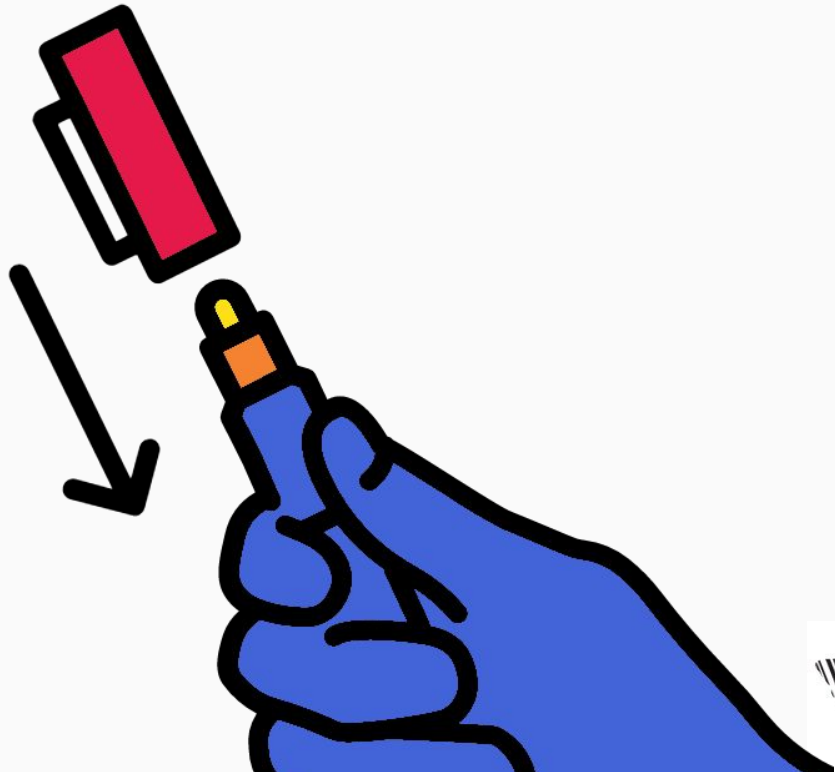
Online class info

Zoom - Ask via microphone if no question currently being asked/answered
Zoom breakout - Quiet room 4
Slack - Ask questions via the channel or ask to go into a zoom breakout room with one of us
WebVNC - We can connect to your webVNC to see and help with issues.



Recap

- What is metabarcoding?
- Environmental DNA
- Diet analysis
- Lab methods & sequencing data
- Analysis pipeline
- Using R
- Assigning taxonomy



Thank you!

Questions?