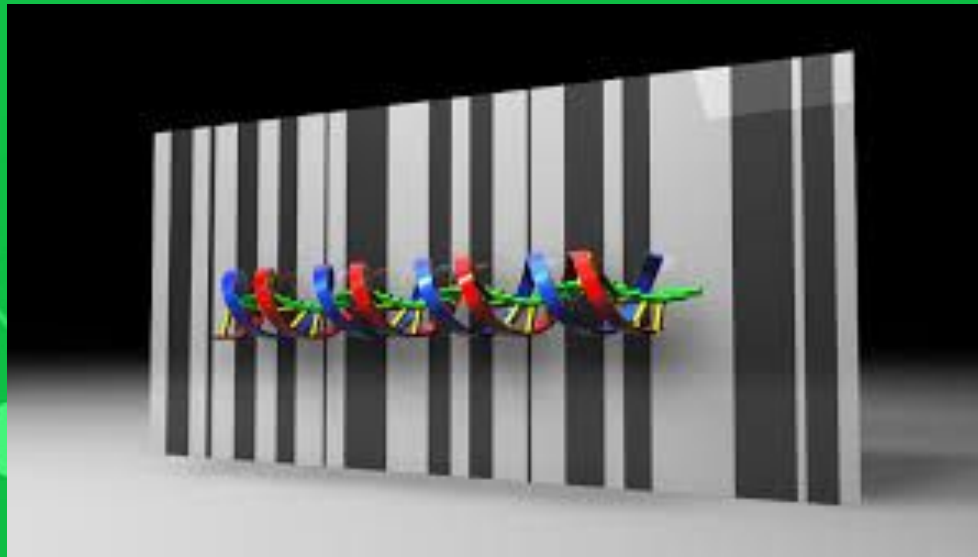
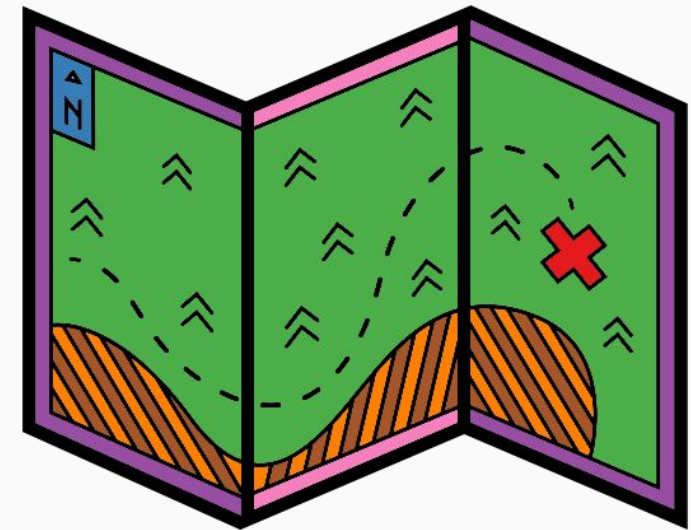


Metabarcoding for diet analysis and environmental DNA – day 2



Outline

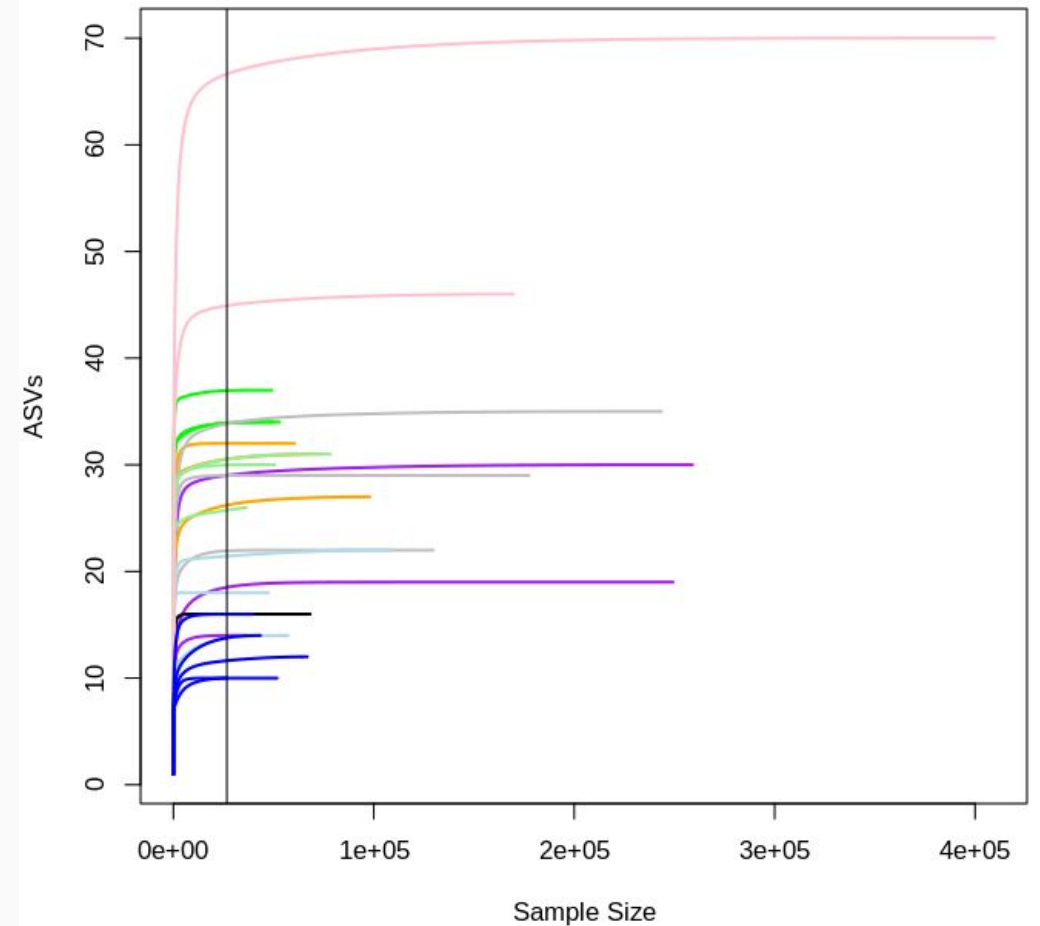
- Normalising
- Assign taxonomy with dada2
- Exporting data from dada2
- Classifying ASVs with GenBank and BLAST
- Using MEGAN to assign taxa
- Creating your own reference database



Data Normalisation



- Have we sequenced enough to detect all taxa?
- Should we randomly subsample to the same depth?
 - May not be appropriate where species are differentially abundant
 - Could generate artificial variation between samples
- Other normalization methods:
 - Using relative abundance
 - Variance stabilising transformation

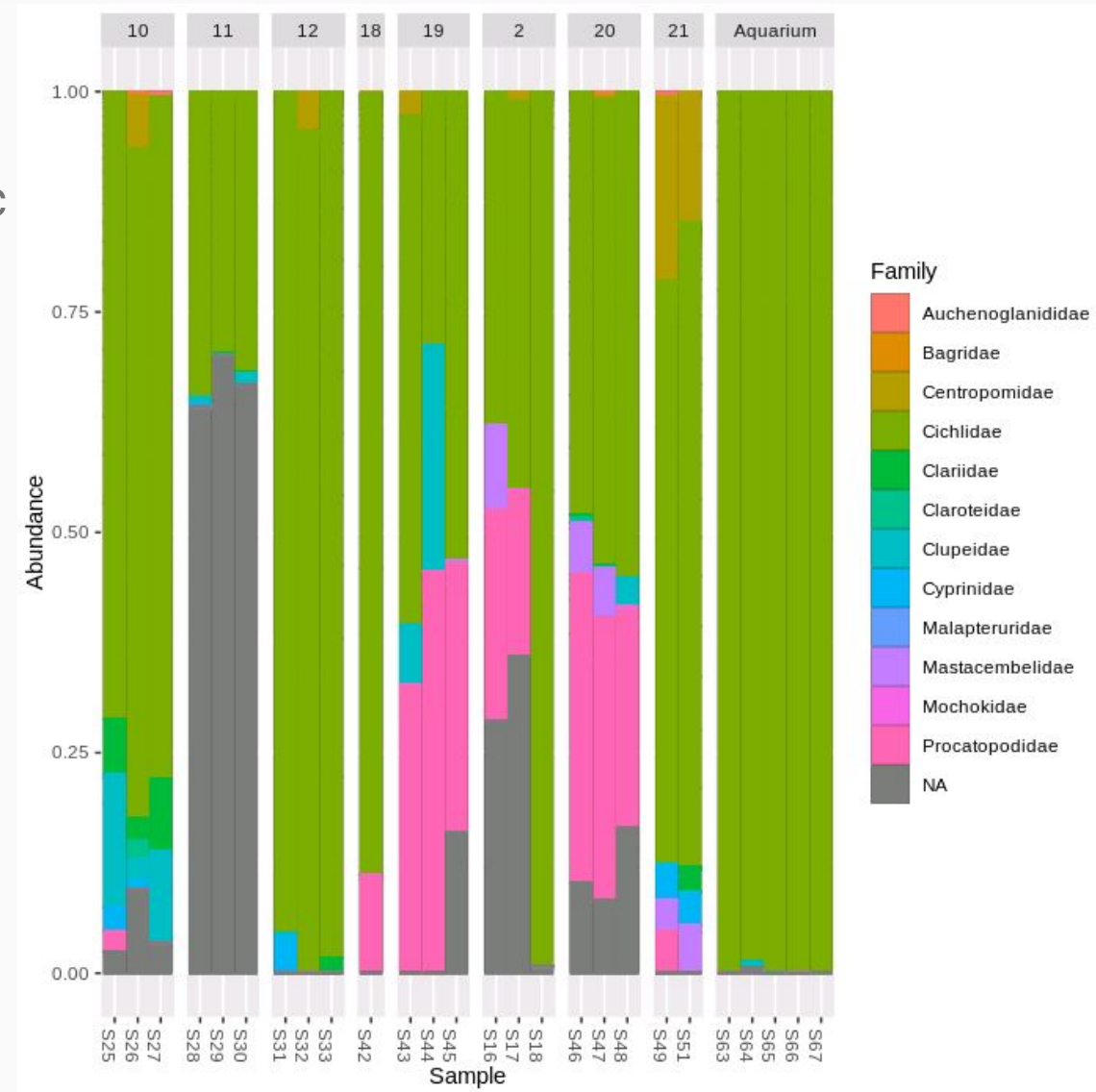


Assigning taxonomy with dada2



Two methods:

- **assignTaxonomy**
 - assigns across multiple taxonomic ranks
 - finds the most similar sequence in the reference file
 - bootstrap approach to assess confidence of assignment at each taxonomic level
-
- **assignSpecies**
 - requires *Genus species* binomials in the reference database
 - ASVs are assigned if they are an exact match
 - most applicable to 16S bacterial amplicon data



Exporting data from dada2



```
>ASV_1
CACCGCGTTATACGAGAGGCTCAAGTTGATAGACATCGGCGTAAAGAGTGGTTAGGAAGTTTTTAACTAAAGCCGAACGCCCTCAGAA
>ASV_2
CACCGCGTTATACGAGAGGCCCAAGTTGATAGACGCCGGCGTAAAGAGTGGTTAGGAAGTTTTTAAAATAAGCCGAATGCCCTCAGA
>ASV_3
CACCGCGTTATACGAGCGGCTCAAGTTGACAGACGTCGGCGTAAAGAGTGGTTAGGAAATTCTTAACTAAAGCCGAACGCCCTCAGAG
>ASV_4
CACCGCGTTATACGAAAGGCTCAAGTTGATTGTACACGGCGTAAAGTGTGGTTAAGGAACTACCTAACTAAAGCTGAACACTCTCAAA
>ASV_5
CACCGCGTTATACGAGAGGCTCAAGTTGATAGACATCGGCGTAAAGAGTGGTTAGGAAGTTTTTAACTAAAGCCGAACGCCCTCAGAA
>ASV_6
CACCGCGTTATACGAGCGGCTCAAGTTGACAGACGTCGGCGTAAAGAGTGGTTAGGAAATTCTTAACTAAAGCCGAACGCCCTCAGAG
```

fasta file

counts

	A	B	C	D	E	F
1		S16	S17	S18	S25	S26
2	ASV_1	27453	0	102667	12198	9096
3	ASV_2	73373	13078	15	1192	3700
4	ASV_3	69	0	0	0	0
5	ASV_4	58385	6027	0	1173	0
6	ASV_5	0	0	134357	1111	0
7	ASV_6	0	0	0	496	868

taxonomy

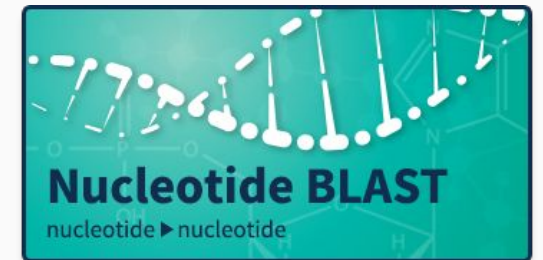
	A	B	C	D	E	F	G	H
1		kingdom	Phylum	Class	Order	Family	Genus	Species
2	ASV_1	Eukaryota	Chordata	Actinopteri	Cichliiformes	Cichlidae	Limnochromis	NA
3	ASV_2	NA	NA	NA	NA	NA	NA	NA
4	ASV_3	Eukaryota	Chordata	Actinopteri	Cichliiformes	Cichlidae	Neolamprologus	NA
5	ASV_4	Eukaryota	Chordata	Actinopteri	Cyprinodontiformes	Procatopodidae	Lamprichthys	Lamprichthys_tanganicanus
6	ASV_5	Eukaryota	Chordata	Actinopteri	Cichliiformes	Cichlidae	NA	NA
7	ASV_6	Eukaryota	Chordata	Actinopteri	Cichliiformes	Cichlidae	Neolamprologus	NA



Classifying ASVs with GenBank and BLAST



- GenBank – database of publicly available DNA sequences
- BLAST – Basic Local Alignment Search Tool
- Can obtain multiple hits for each ASV/query sequence
- Very useful but proceed with caution! GenBank is not error-free.



Using MEGAN to assign taxa



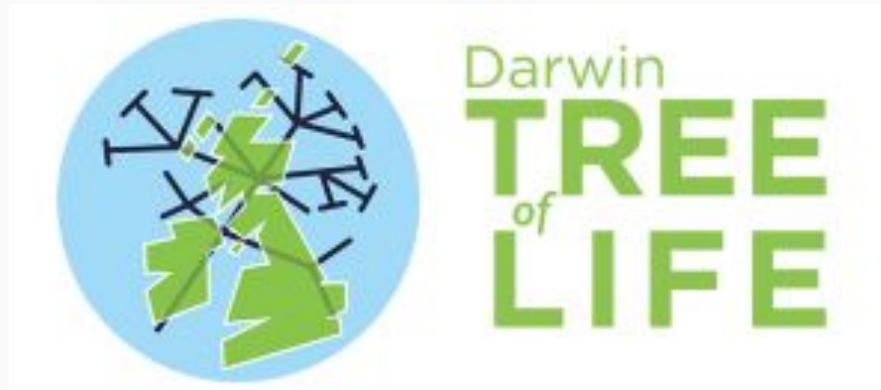
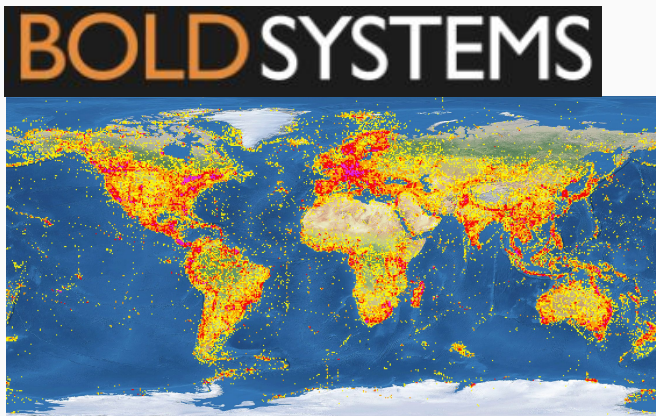
- MEtaGenome ANalyzer
- Import BLAST results
- Lowest Common Ancestor algorithm
 - Multiple BLAST hits to different taxa
 - ASV assigned at lowest node that incorporates all 'good' BLAST hits.



Creating your own reference database



- Sample individual species and generate barcode sequences
- Download specific set of sequence from GenBank
- BOLD database – curated DNA barcode, > 300,000 species
- Darwin Tree of Life – whole genome sequencing for every species in the UK





Thank you!

Questions?

