Exercises in Marine Ecological Genetics

10. Metabarcoding: microbiome analysis

- Infer ASVs (Amplicon Sequence Variants)
- Compare microbial community composition
- Assess alpha and beta diversity

Martin Helmkampf

https://github.com/mhelmkampf/meg25

Marker gene-based microbiome analysis

- Uses conserved marker gene for bacteria / archaea: **16S**
- Enables profiling of **whole microbial communities** (including non-cultivable / unknown taxa)
- **Applications**
  - Characterization of environments (e.g. sediment, ocean water)
  - Host-associated microbiomes (e.g. human gut, animal-microbe symbioses)
  - Bioindicators for ecosystem health, food safety, and public health monitoring
- More **cost-effective** and less complex than shotgun metagenomics

Image: Nadezhda Buravleva / vecteezy.com

Example project

Sabrin Abdelghany

Do sea cucumbers of the species *Actinopyga crassa* in the Red Sea

- feed selectively?
- change the microbial community of the sediment?

Sample	*N*
Seawater	3
Seagrass	4
Sediment	4
Foregut	5
Midgut	5
Hindgut	5

Approach: Amplicon sequencing of **16S** region V3-V4 (~ 460 bp) on **Illumina MiSeq**

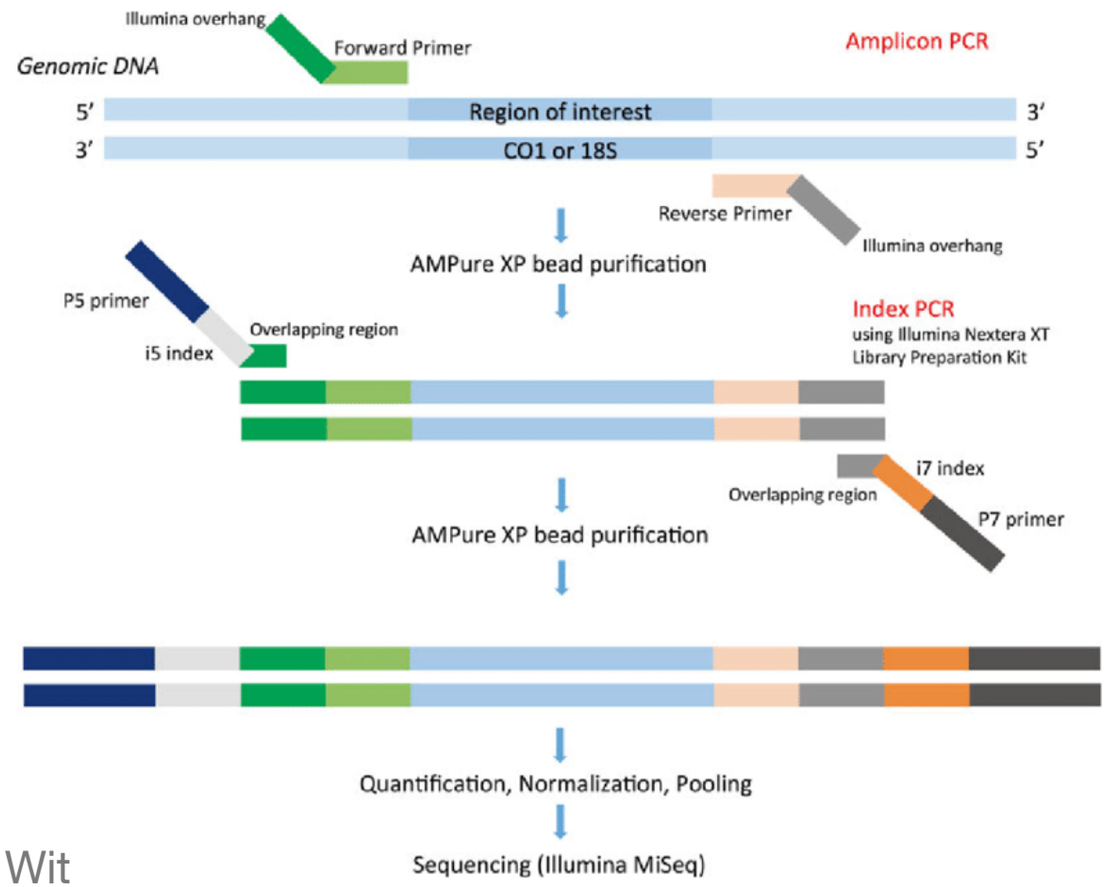
Microbiome analysis with **QIIME 2**

Amplicon sequencing

Next-gen sequencing of specific genomic region (e.g. 16S)

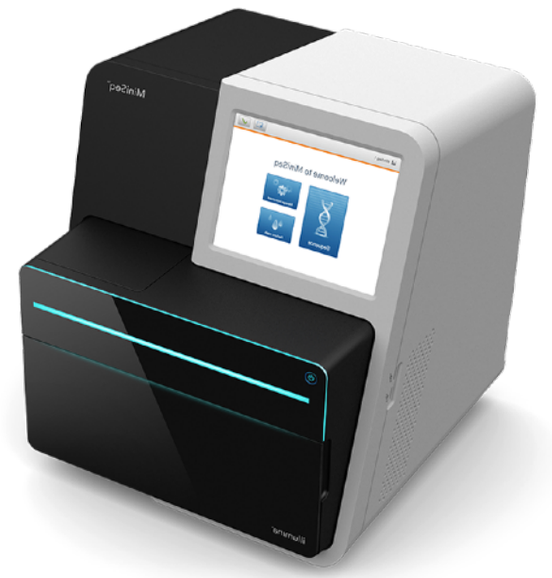
Many samples in parallel by **multiplexing** (using unique molecular index for each sample)

Dual-PCR library preparation



Illumina MiSeq bench top sequencer

7–25 million reads, up to 2 × 300 bp



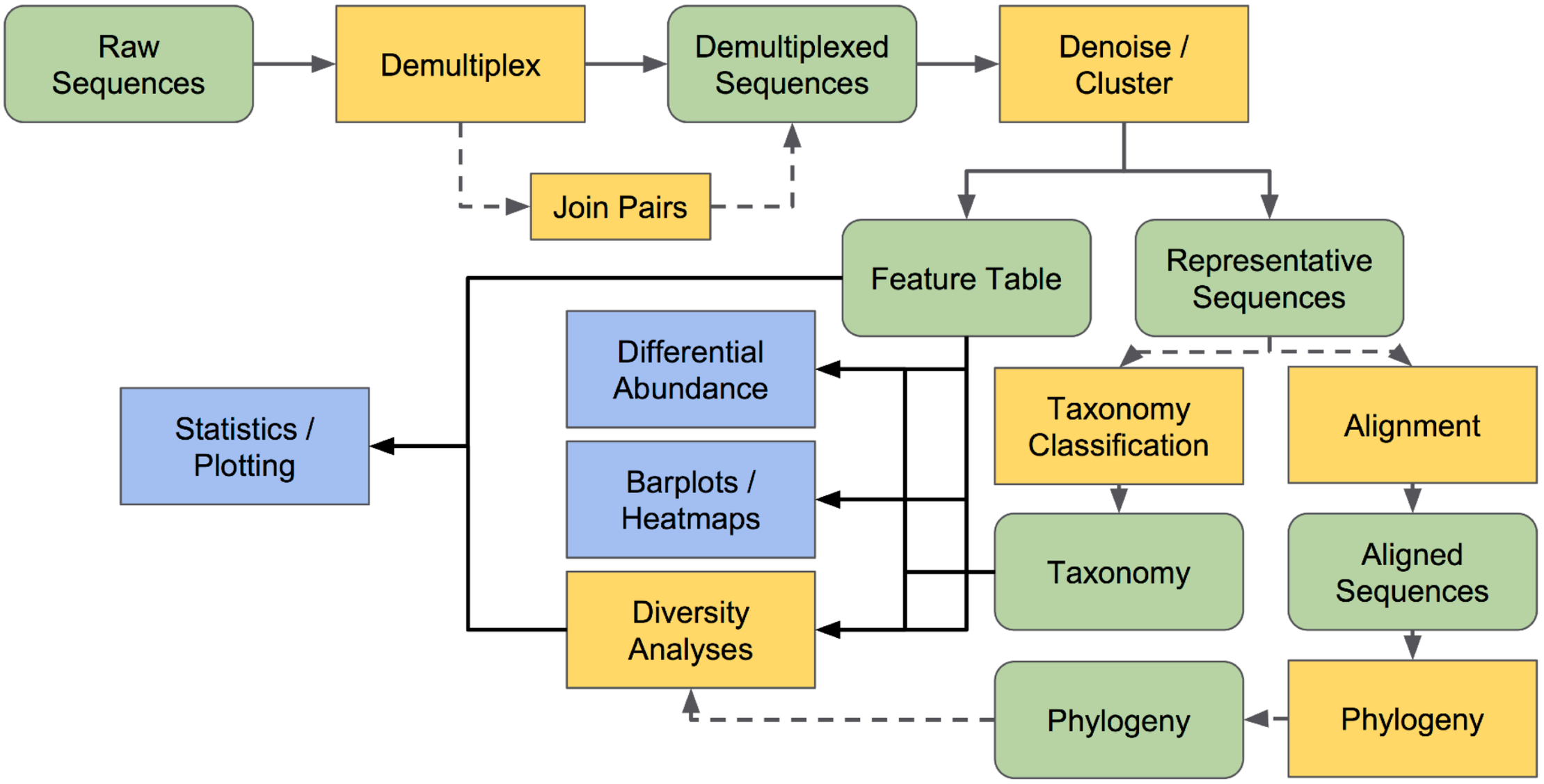
Pierre De Wit

Illumina Inc.

General metabarcoding steps

- Sample collection and DNA extraction
- PCR amplification (e.g. 16S, COI)
- Sequencing (e.g. Illumina, ONT)
- **Demultiplexing, trimming and quality filtering** (data preprocessing)
- **Amplicon Sequence Variant inference** (denoising) | alternative: OTU clustering
- **Taxonomic assignment**
- **Downstream analyses** (community composition, diversity metrics)

QIIME 2 microbiome analysis workflow



qiime2.org

Data preprocessing

Exercise 1

- Import of raw reads (e.g. Fastq files) and metadata — `qiime tools import`
- Demultiplexing, primer removal — `qiime demux-paired`
- Joining forward and reverse reads — `qiime vsearch join-pairs`
- Quality check — `qiime demux summarize`

output.qza (data)

output.qzv (visualization)

View at https://view.qiime2.org

Denoising

Exercise 2

- Quality filter

  Filter out low-quality sequences, trim reads

`qiime deblur denoise-16S`

- Dereplicate

  Collapse identical reads into unique sequences with counts

- Deblur

  Statistically infer error-free sequences, remove chimeras and singletons

Amplicon Sequence Variants

ASV sequences

ASV table (frequencies)

Error modeling

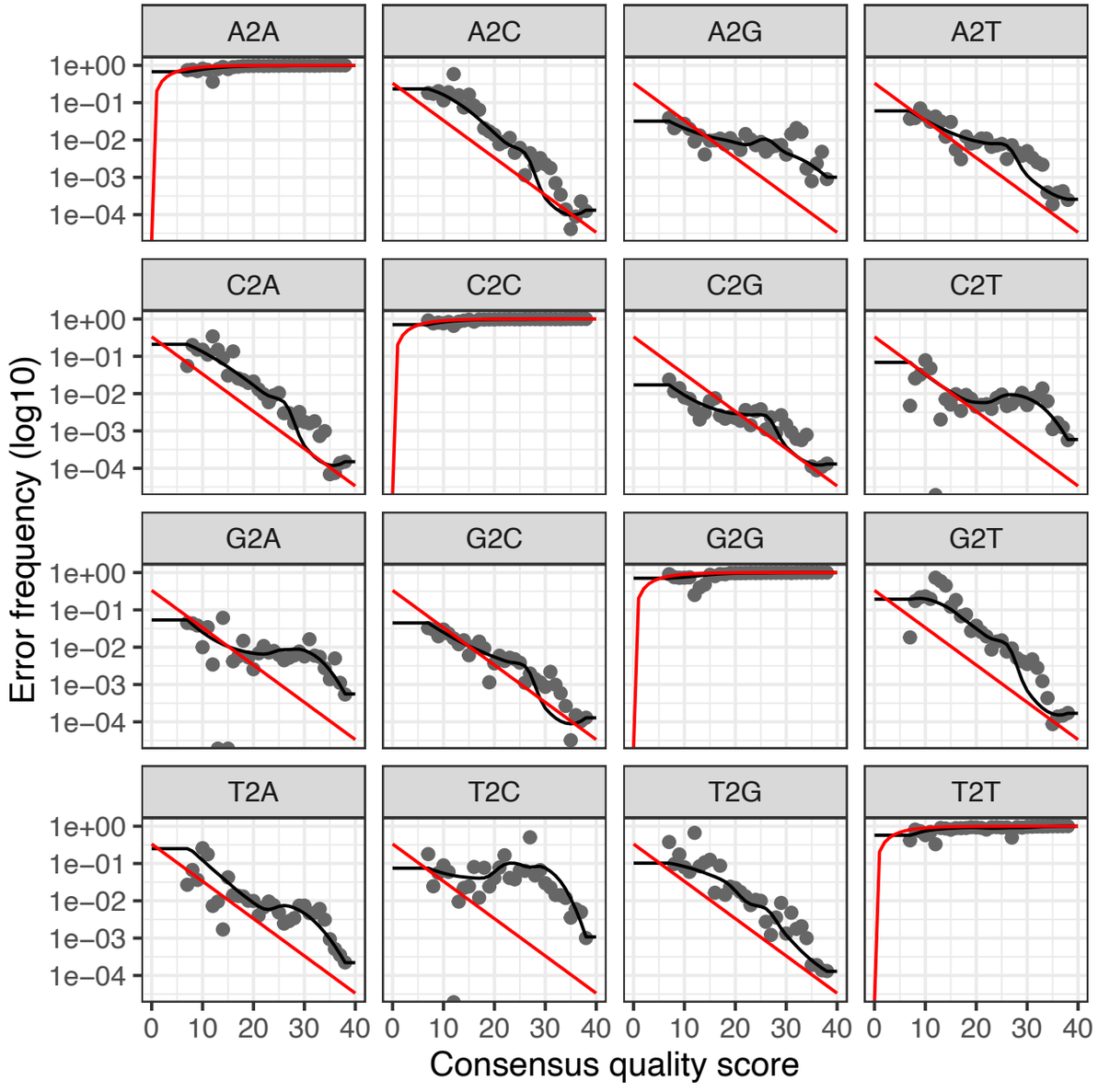
Exercise 2

- Error rates are **estimated directly** from data

- Model accounts for each base and its quality score

- Sequences are **statistically tested** against the error model

- If probability of being an error is low, sequence is kept as true variant

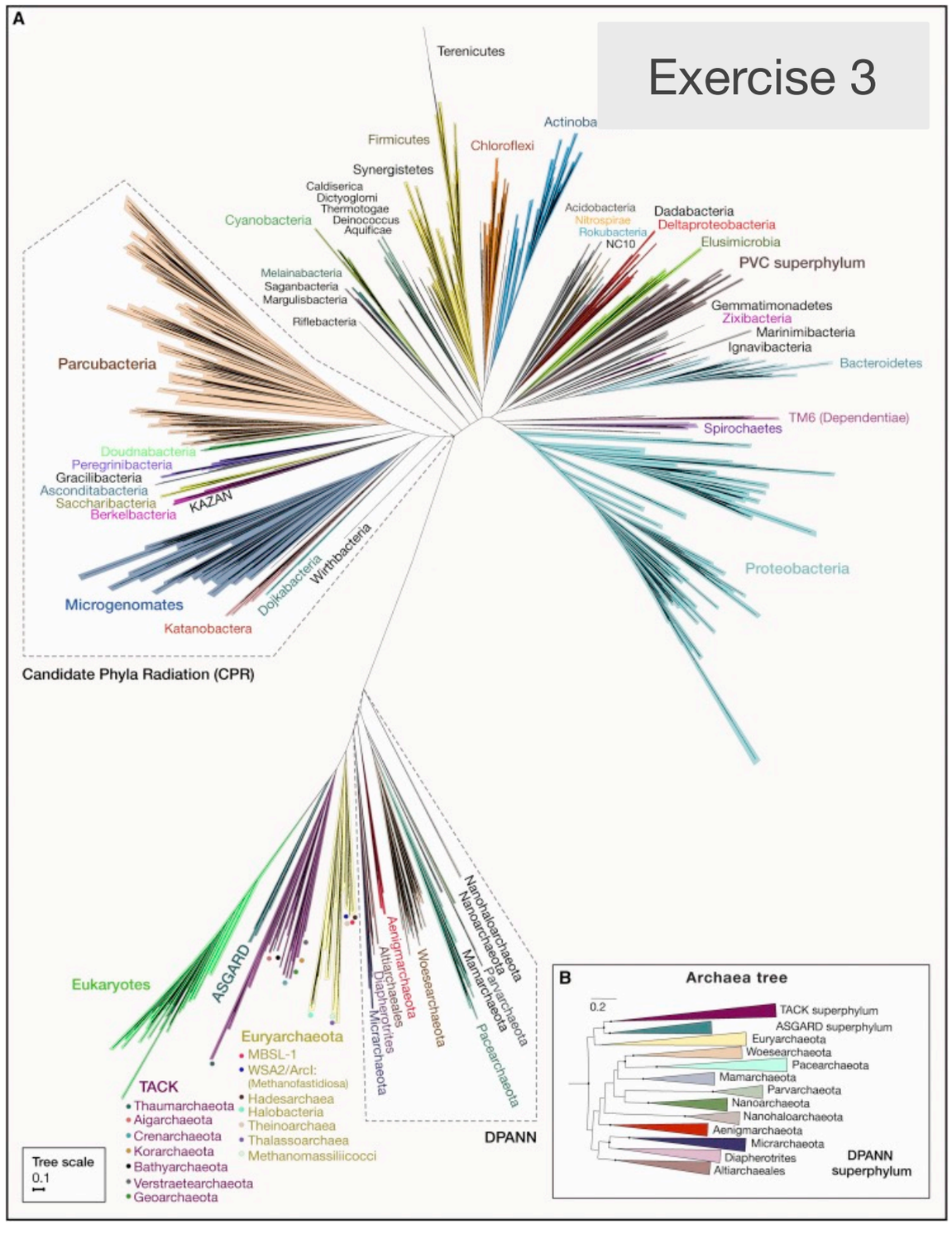


Taxonomic assignment

- Similarity / alignment-based (e.g BLAST)
- Phylogenetic placement
- Probabilistic classifiers (e.g. QIIME2)

```
qiime feature-classifier
```

silva
high quality ribosomal RNA databases



Castelle & Banfield 2018, Cell

Matching sequences to database with BLAST

Exercise 3

Algorithm overview

- Split query into very short segments (*k*-mers or words)
- Find exact matches between words and sequences in database (seeds)
- Extend matches to **local alignments** (HSP; stop once too many mismatches occur)
- Evaluate statistical significance of each HSP (**e-value**)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Oncorhynchus keta mitochondrial COX1 gene for cytochrome c oxidase subunit 1, partial cds, isolate: OK_M08F	Oncorhynchus keta	1029	1029	100%	0.0	99.64%	772	LC094471.1
☑	Oncorhynchus keta mitochondrial COX1 gene for cytochrome c oxidase subunit 1, partial cds, isolate: OK_M01F	Oncorhynchus keta	1029	1029	100%	0.0	99.64%	772	LC094464.1
☑	Oncorhynchus keta isolate 10_Narva cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial	Oncorhynchus keta	1029	1029	100%	0.0	99.64%	655	KR778851.1

...

Diversity metrics

Exercise 4

How diverse is the community?

How similar / dissimilar are the communities?

`qiime diversity`

Alpha diversity

Within-community diversity, e.g. species richness, Shannon's index —— $H' = -\sum_{i=1}^{R} p_i \ln p_i$

(richness + evenness)

Beta diversity

Dissimilarity between communities, e.g. Jaccard distance (presence / absence only),

Bray-Curtis index (relative abundances)

$BC_{ij} = 1 - \frac{2C_{ij}}{S_i + S_j}$

0 (same community)

– 1 (no similarity)

Metabarcoding

Summary

```
qiime deblur denoise-16S    # filter, dereplicate and denoise reads (infer ASVs)

qiime feature-classifier                          # classify ASVs taxonomically

qiime diversity                                   # calculate diversity metrics
```

- Metabarcoding uses marker genes (e.g. 16S, COI) to identify and compare organisms in **complex communities**

- Reads can be “denoised” into high resolution **Amplicon Sequence Variants** or more simply clustered into OTUs

- Downstream analyses like community composition, diversity metrics, and comparisons across samples or environments enable powerful **ecological insights**

Options for next week

- Lecture: Introduction to phylogenetics
- Exercises: wedgefish phylogeny based on COI barcodes

OR

- Lecture: none
- Exercises: Do hamlet species differ in their diet? A gut content metabarcoding analysis using R (~ eDNA)

Photos: Jake Wilton | Kosmas Hench

Course evaluation — please participate!

https://elearning.uni-oldenburg.de/plugins.php/unizensusplugin/show?cid=3660d16e8eb3daf479389cf8233c12fb