# **Exercises in Marine Ecological Genetics**

#### 10. Metabarcoding: microbiome analysis

- Identify ASVs (Amplicon Sequence Variants)
- Visualize microbial community composition
- Assess alpha and beta diversity

Martin Helmkampf

https://github.com/mhelmkampf/meg25



### Microbiome analysis 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

130/E		139	SYRIA	!	40°E
Mediterranean		Haifa .	SIKIA	1 6	IRAO
Measterranean Sea		100	1. 18 B	6,	
	TEL AVIV	YAFO -	AMMĀN		
	18		USALEM		-
A	_Port Said	1 600	- 60 100		
Alexandria	and the same of	Al Arish J	ORDAN \		
John	Ismailia	Al Alian	1		
	1	The state of the s	1		20 3/1/0 at
CAIRO T	Suez	1211			The state of
Black Co.	Train	Eilat Agaba	of his was	59 40	
'Ayn Sukhnah		Mydaba	And the	7 0 2 3	
	Suez	Gulf	a remarked		
	The same	The state of the s			
19 19 DW	Ghárib	Agaba	TO THE STATE OF		
nasi	JII O	The state of the s	Chipping Sec.	SAUDI	1 5
	生美国.	Sharm .		ARABI	1 5000
The state of the s	1 100	ash Shaykh		037-04-0	1120
Asyut 17	Hurghada	X	19 July 19 19 19 19 19 19 19 19 19 19 19 19 19	18 mg	
	J. Proc	Būr Safājah	July 2		9781114
EGYP	T	à		174 M. 1 84	· 10 11 11 11 11 11 11
The same of	10	Al Quşayr		10 / Table	Water Care
of fair			5		· · · · · · · · · · · · · · · · · · ·
and the state of	<b>有</b>	97.00	Sept.	Array - W.	
	The same of	753		The William	
	3 1300	- akin	X/X	2007-200-20	Medina
	(A. (1970)		199	持つまる	INCOME.
	Aswan	, All h	E	<b>人保護</b>	
in the state of	A M	īnā' Baranīs	20	1111	E Service and
Mary Company of	1		>	3.7	
	3	10 8 6 5 6 5		2	The second second
	1	Dans -	-4	1	方。1960年1月1日
State of the party	100			- 4	Trans.
	12	1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2		2 37	1.24
10000		123 7 6 M	484	Y 31	The state of the s
A PARTY OF THE PAR	P. 11.	The state of the		Jeddah •	• Megca
	and the	and the same	STATE OF	(	Libre W
The state of the s	UDAN	149	100	,	( 新新語
	154 00	177	101-30		SEPERATE TO
International	metr	es · ·	10000		The same of the sa
boundary	4000		11 11 11		7
Palestinian territori	3000	10/1/23/10	- 201h		3
0 200	2000		and the state of		7
kilometres	1000	1.30	智能量。		5.
© Australian National University CartoGIS CAP 05-025d	0		11759-16		3.7
			- 15 11	-	



Approach: Amplicon sequencing of **16S** region V3-V4 (~ 460 bp) on **Illumina** MiSeq Microbiome analysis with **QIIME 2** 

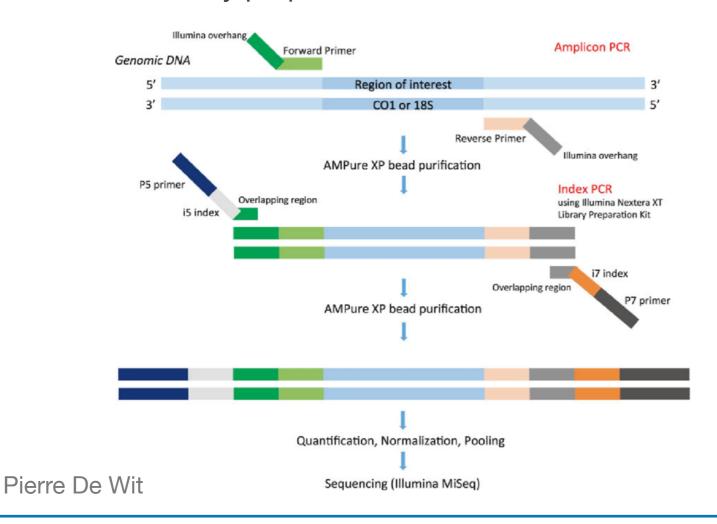


#### **Amplicon sequencing**

Next-gen sequencing of specific genomic regions (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



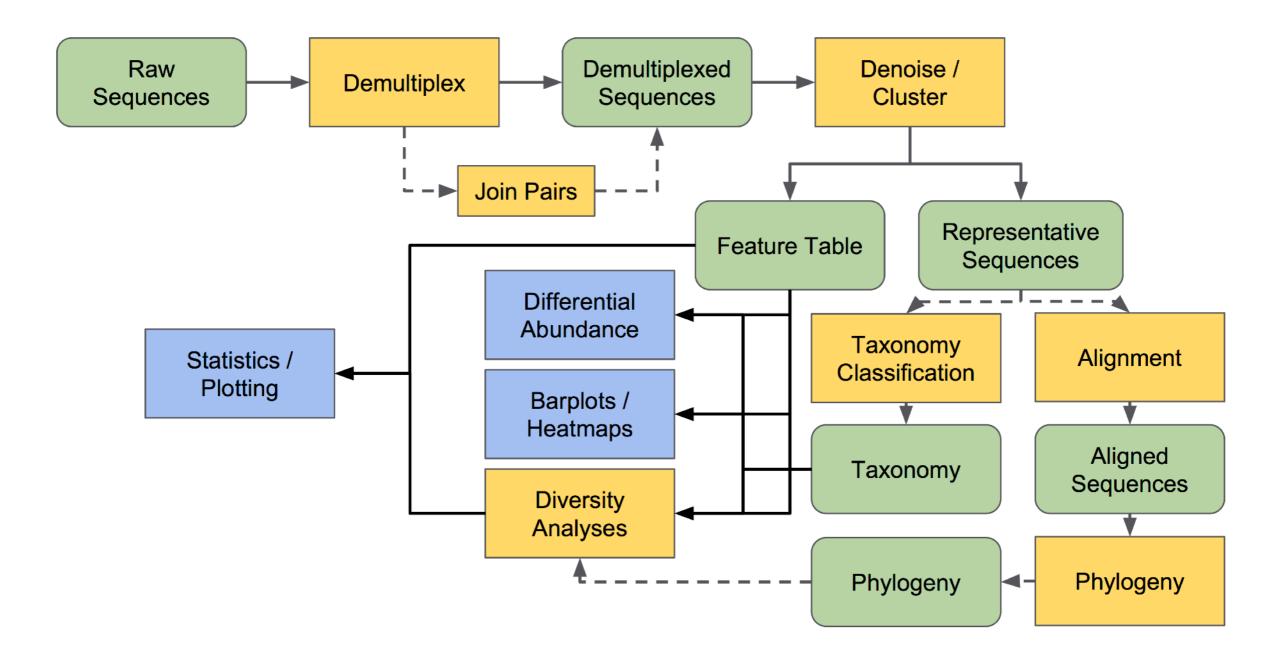
Illumina Inc.



# **General metabarcoding steps**

- Sample collection and DNA extraction
- PCR amplification (e.g. 16S, COI)
- Sequencing (e.g. Illumina, ONT)
- Demultiplexing, quality filtering and trimming (data preprocessing)
- Amplicon Sequence Variant inference (denoising) | 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 <a href="https://view.qiime2.org">https://view.qiime2.org</a>

**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**

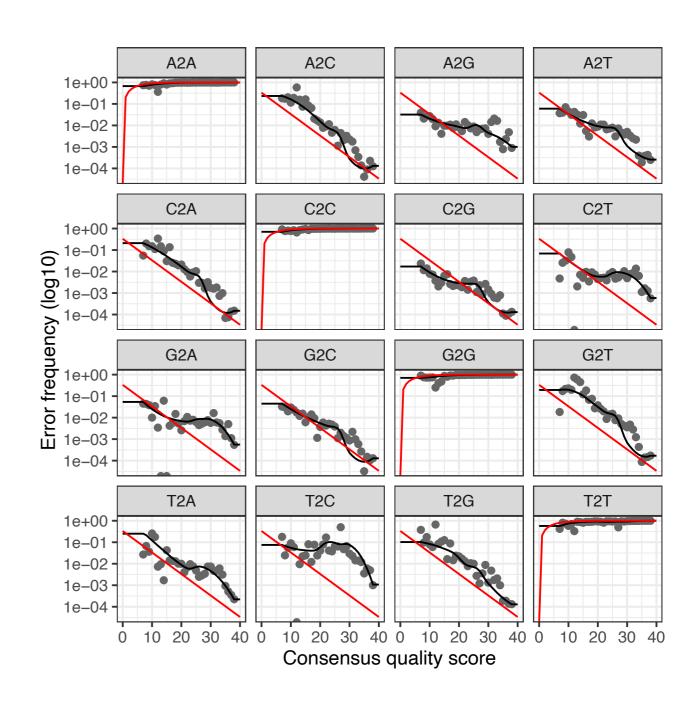
ASV sequences

ASV table (frequencies)



# **Error modeling**

- 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



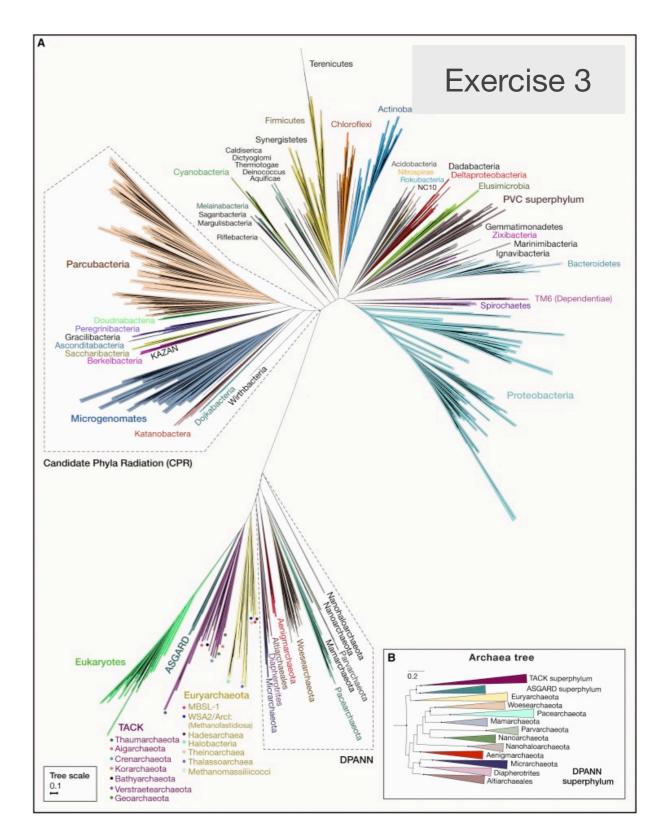


### **Taxonomic assignment**

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

qiime feature-classifier





Castelle & Banfield 2018, Cell



#### Matching sequences to database with BLAST

#### 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		Query Cover			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 —  $H' = -\sum_{i=1}^{n} p_i \ln p_i$  (richness + evenness)

#### **Beta diversity**

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

Bray-Curtis index (relative abundances)

$$igspace BC_{ij} = 1 - rac{2C_{ij}}{S_i + S_j}$$
 0 (same community) on  $-1$  (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



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



Photos: Jake Wilton | Kosmas Hench

OR