Meta-barcoding of microbial communities

Susanne Wilken

Marc Galland

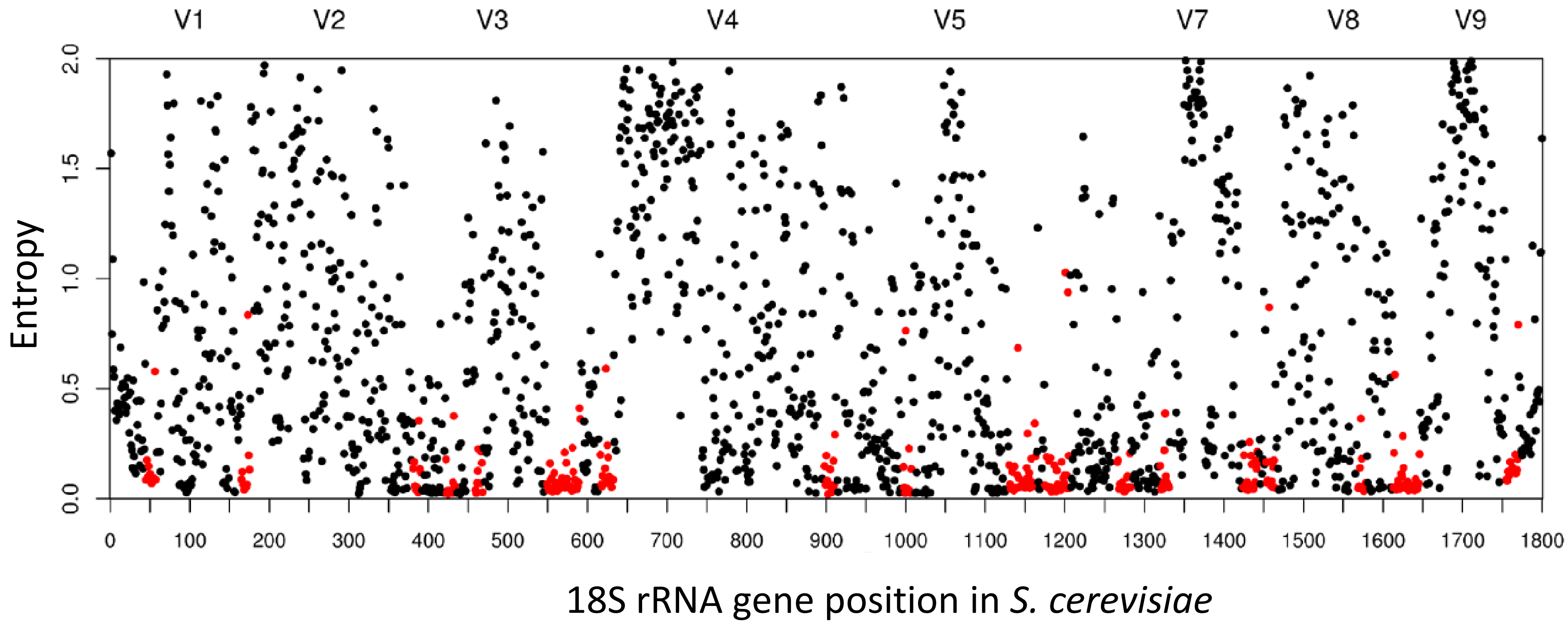
TARA Oceans “studying plankton at planetary scale”

A smaller, faster, cheaper route to fusion energy
Sport, sex, and testosterone go to court p. 858
Making C-N bonds from nitro groups
Science
$10
22 MAY 2015
sciencemag.org
AAAS
A world of plankton
Marine biodiversity surveyed by the schooner Tara
pp. 865 & 873

Which marker-gene to use?

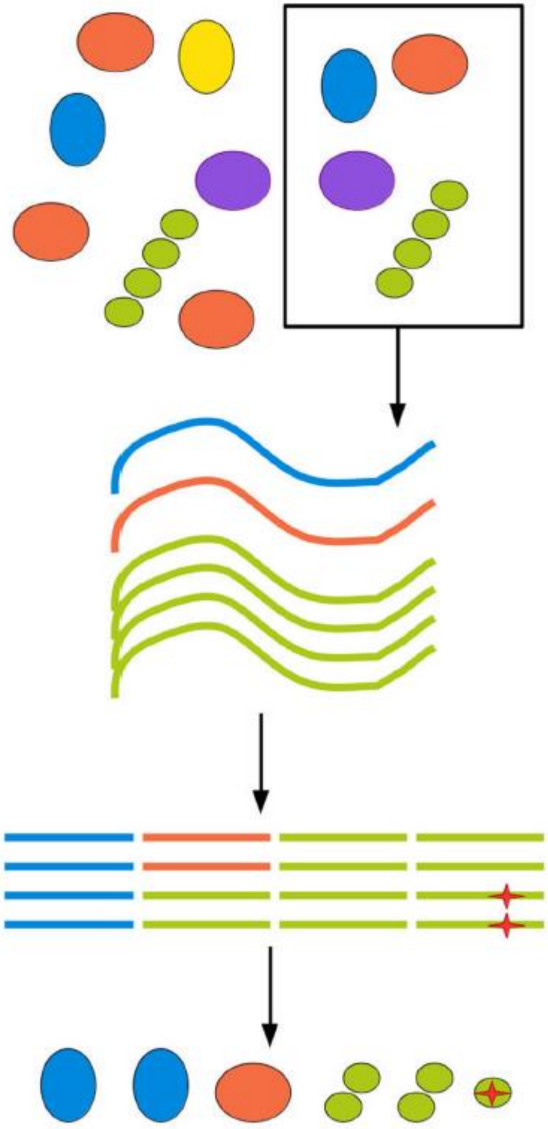
- 16S rRNA gene V4, V1-2 for prokaryotes
- 18S rRNA gene V4, V9 for eukaryotes
- Ribosomal ITS for fungi
- 16S rRNA gene V1-2 for aquatic primary producers (both cyanobacteria and plastids)
- Mitochondrial cytochrome C oxidase for animals
- Rubisco (RbcL) for plants

Choosing primers – Generality vs. phylogenetic resolution



Hadziavdic et al. 2014

Inferring community composition



Sampling bias

Extraction bias

Primer bias

PCR-errors

Chimera formation

Sequencing errors

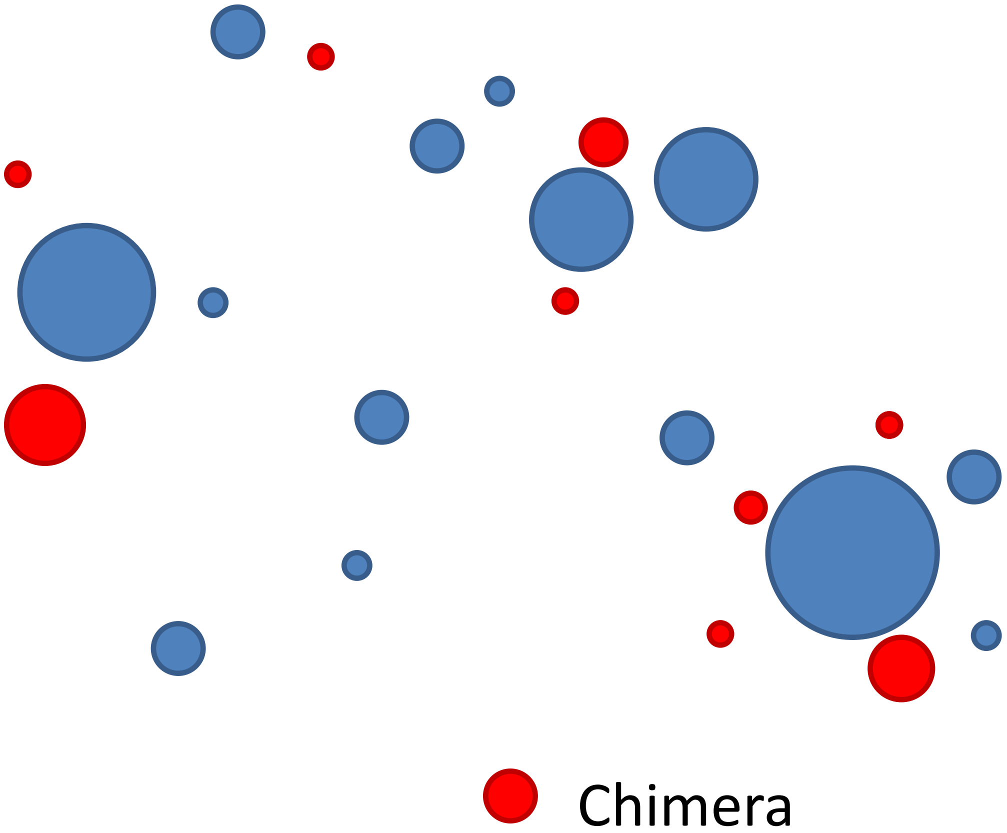
Correction attempted during data analysis

From Hugerth & Andersson 2017

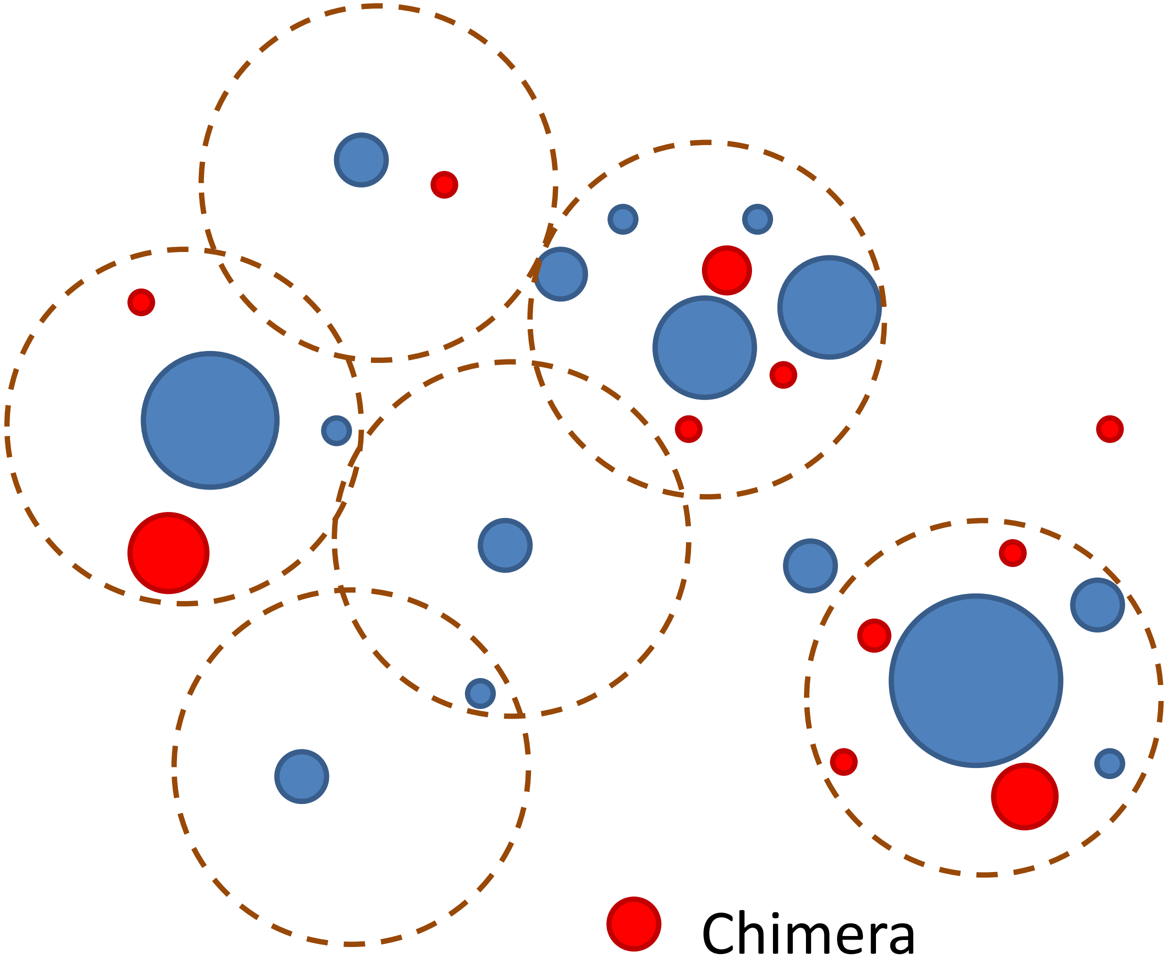
Quality filtering and trimming

- Check presence of primers
- Remove non-biological sequences
- Remove too short reads
- Remove low quality bases (based on individual score or sliding window)
- Merge paired-end reads

Operational taxonomic units (OTUs) vs. amplicon sequence variants (ASVs)



Operational taxonomic units (OTUs)

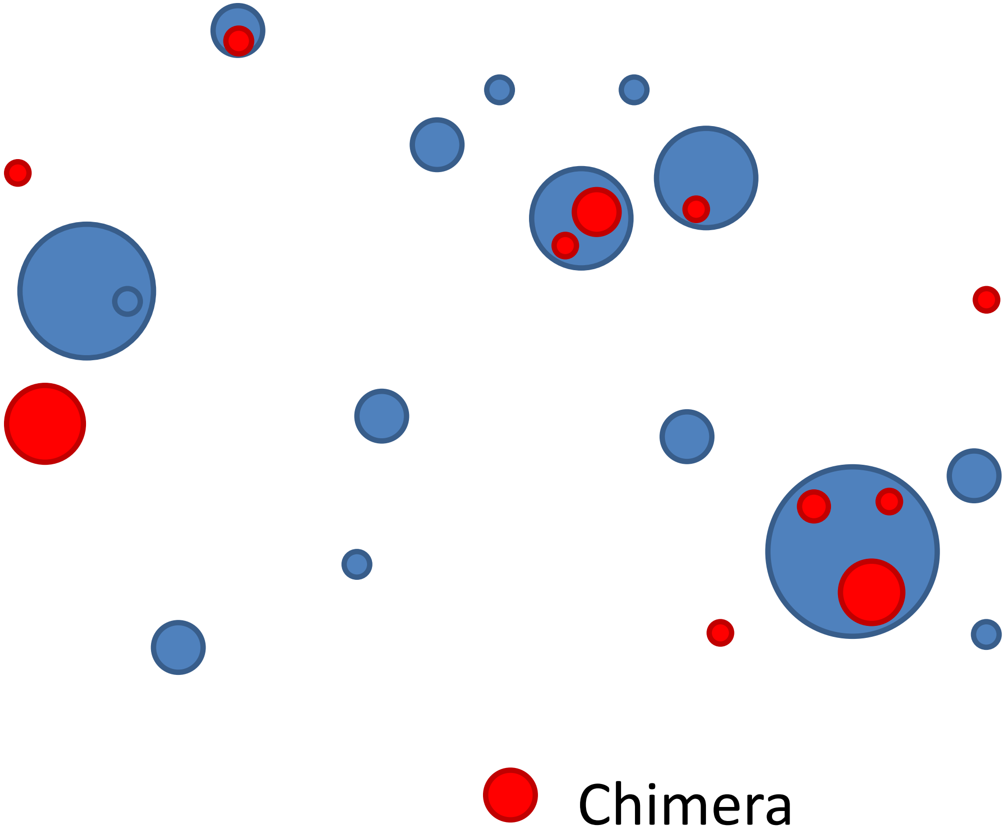


Clustering based on sequence similarity
(erroneous sequences should cluster with original true seqs)

OTUs represented by centroid sequences (dependent on specific datasets)

Approaches include:
- Closed reference
- Open reference
- De novo

Amplicon sequence variants (ASVs)

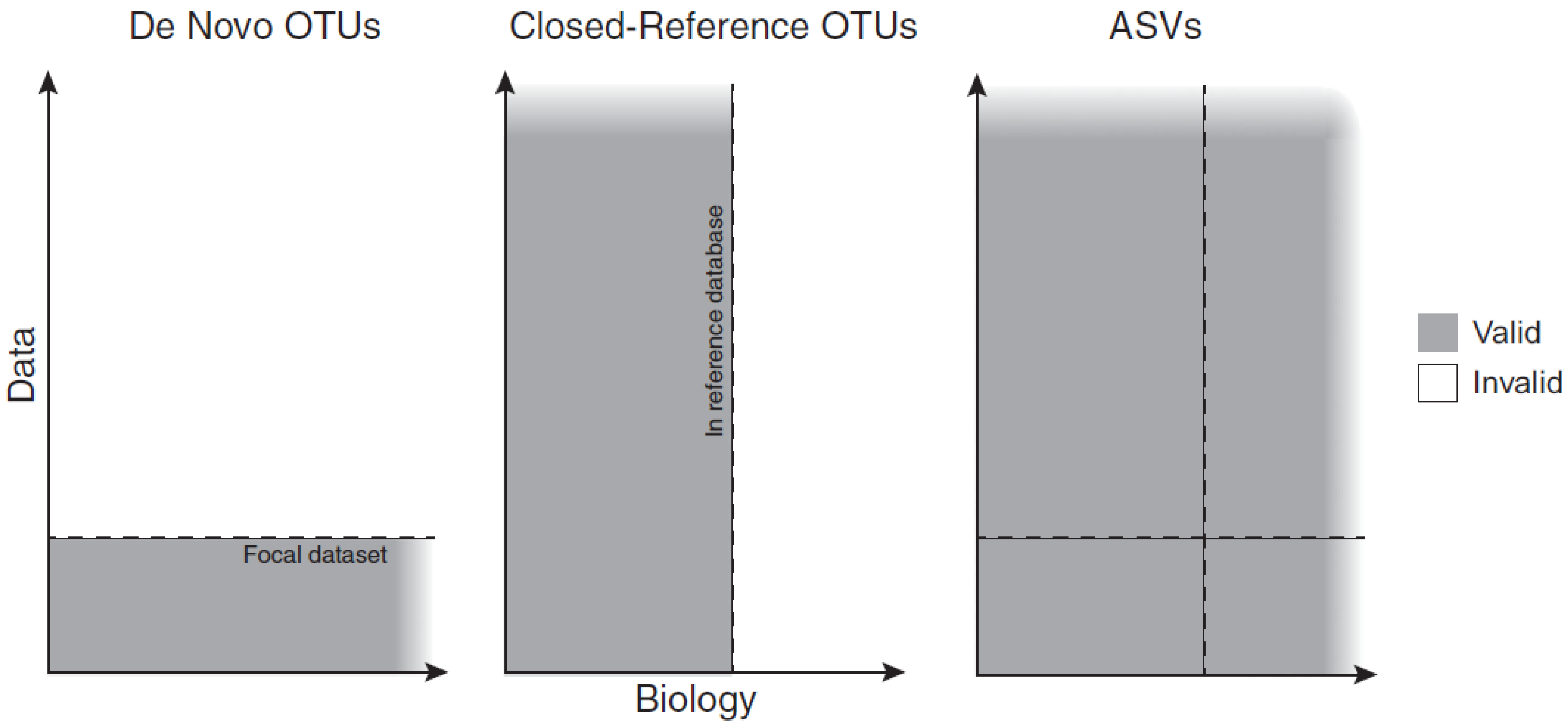


More precise error modeling allows distinction of true and erroneous sequences (still based on similarity and abundance)

ASVs as more meaningful biological units

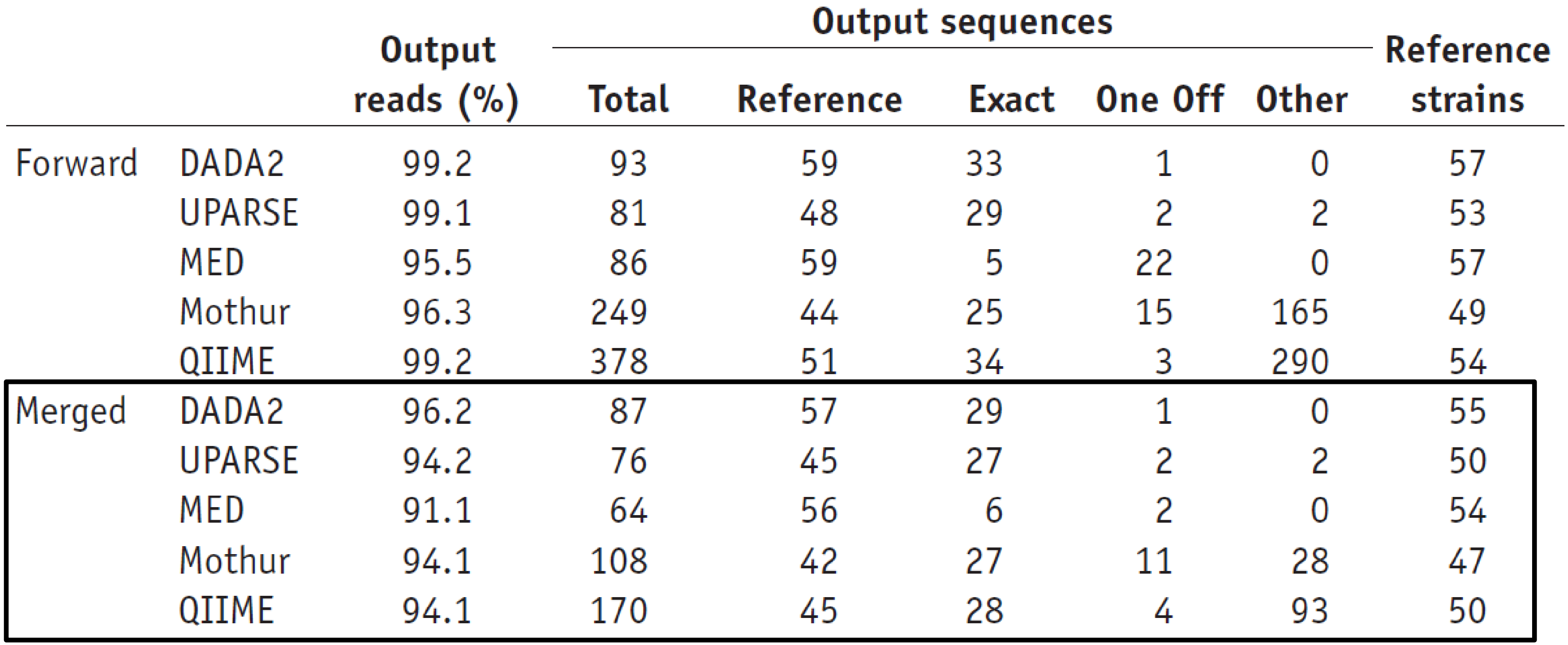
Comparability among studies

Amplicon sequence variants (ASVs)



Callahan et al. 2017

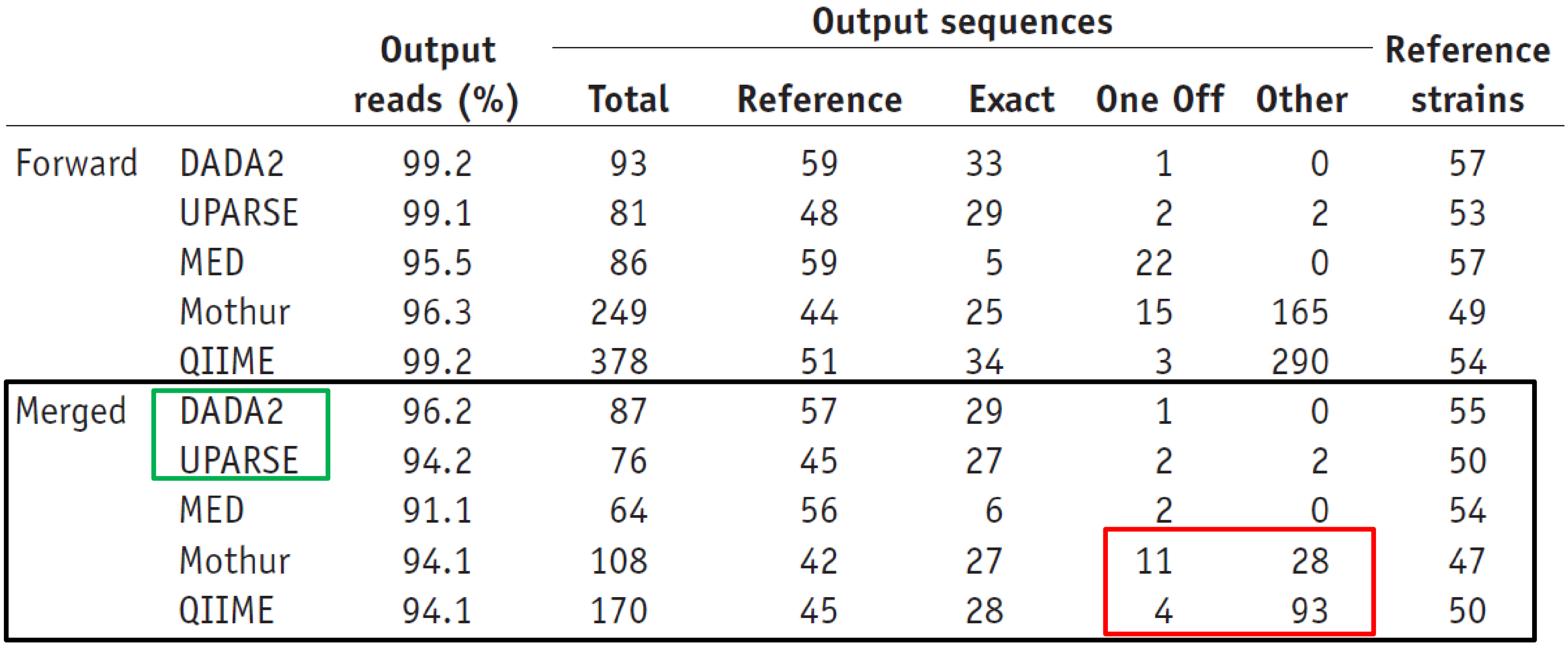
Benchmarking with mock communities



		Output reads (%)	Output sequences					Reference strains
			Total	Reference	Exact	One Off	Other	
Forward	DADA2	99.2	93	59	33	1	0	57
	UPARSE	99.1	81	48	29	2	2	53
	MED	95.5	86	59	5	22	0	57
	Mothur	96.3	249	44	25	15	165	49
	QIIME	99.2	378	51	34	3	290	54
Merged	DADA2	96.2	87	57	29	1	0	55
	UPARSE	94.2	76	45	27	2	2	50
	MED	91.1	64	56	6	2	0	54
	Mothur	94.1	108	42	27	11	28	47
	QIIME	94.1	170	45	28	4	93	50

Callahan et al. 2016

Benchmarking with mock communities



		Output	Output sequences					Reference
		reads (%)	Total	Reference	Exact	One Off	Other	strains
Forward	DADA2	99.2	93	59	33	1	0	57
	UPARSE	99.1	81	48	29	2	2	53
	MED	95.5	86	59	5	22	0	57
	Mothur	96.3	249	44	25	15	165	49
	QIIME	99.2	378	51	34	3	290	54
Merged	DADA2	96.2	87	57	29	1	0	55
	UPARSE	94.2	76	45	27	2	2	50
	MED	91.1	64	56	6	2	0	54
	Mothur	94.1	108	42	27	11	28	47
	QIIME	94.1	170	45	28	4	93	50

Callahan et al. 2016

To learn more:

There’s two excellent reviews of available methods and their potential biases including downstream statistical techniques. These are must-reads for everyone getting started with meta-barcoding:

- Balint et al. 2016: Millions of reads, thousands of taxa: microbial community structure and associations analyzed via marker genes. FEMS Microbiology Reviews 40:686-700.
- Hugerth and Andersson 2017: Analysing microbial community composition through amplicon sequencing: from sampling to hypothesis testing. Frontiers in Microbiology 8: 1561.

More specific to Dada2 and phyloseq:

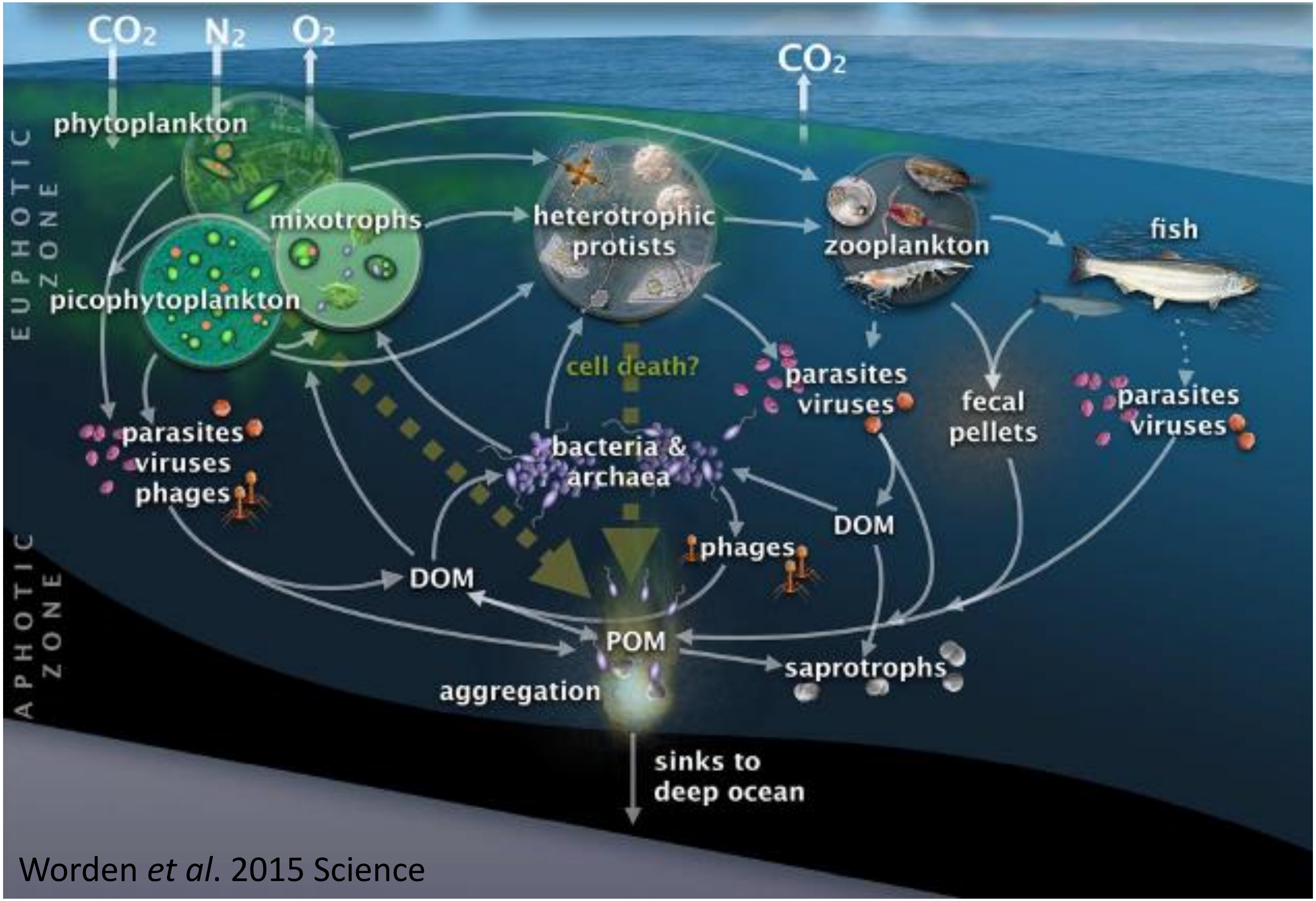
- Callahan et al. 2016: DADA2: high-resolution sample inference from Illumina amplicon data. Nature methods 13: 581-583.
- Callahan et al. 2017: Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME Journal 11: 2639-2643.
- McMurdie and Holmes 2013: phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 8: e61217

Other references:

- Hadziavdic et al. 2014: Characterization of the 18S rRNA gene for designing universal eukaryote specifc primers. PLoS ONE 9: e87624.

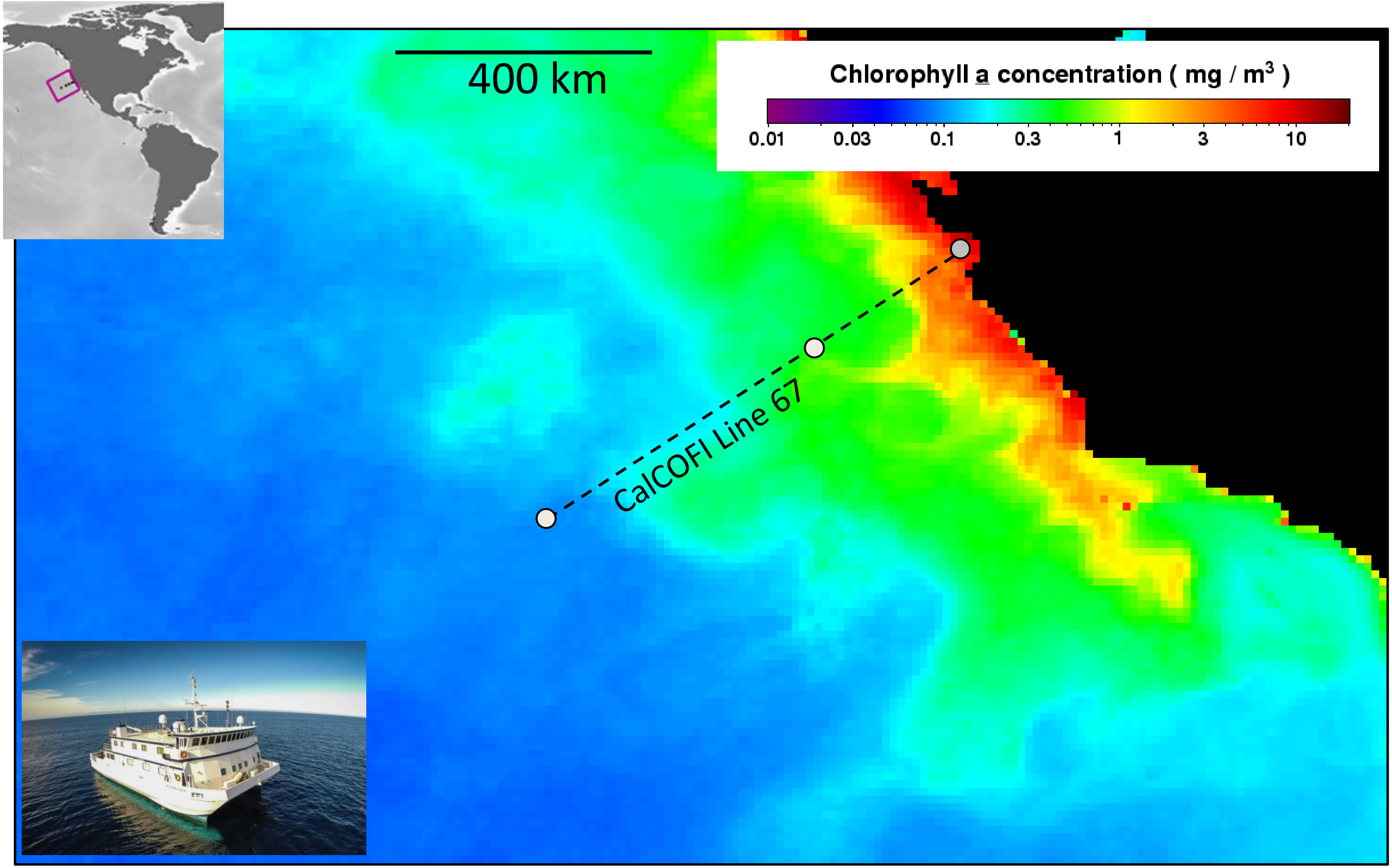
18S rRNA dataset of marine microbial predators

The marine carbon cycle



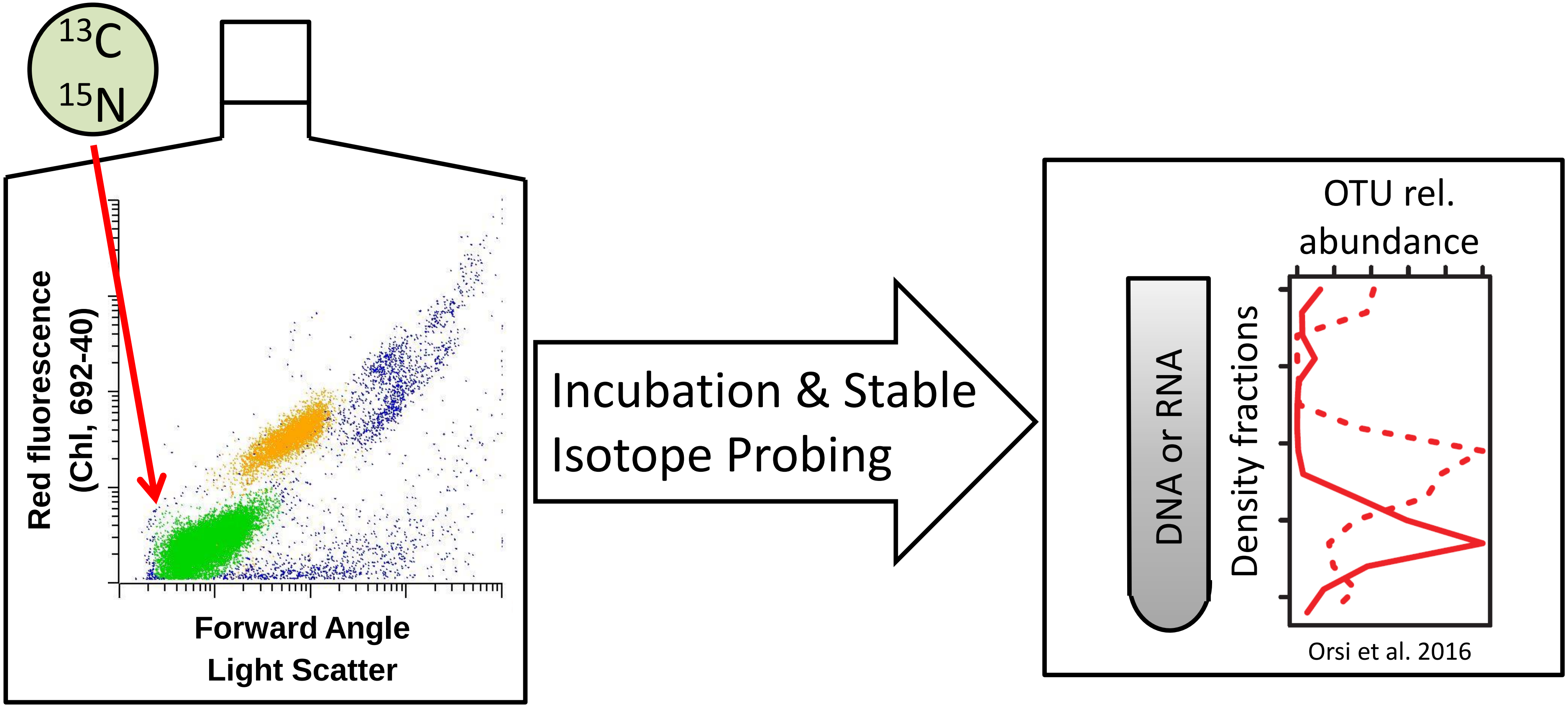
Worden *et al*. 2015 Science

Identifying microbial predators



NASA Goddard Space Flight Center, Ocean Biology Processing Group, Aqua/MODIS

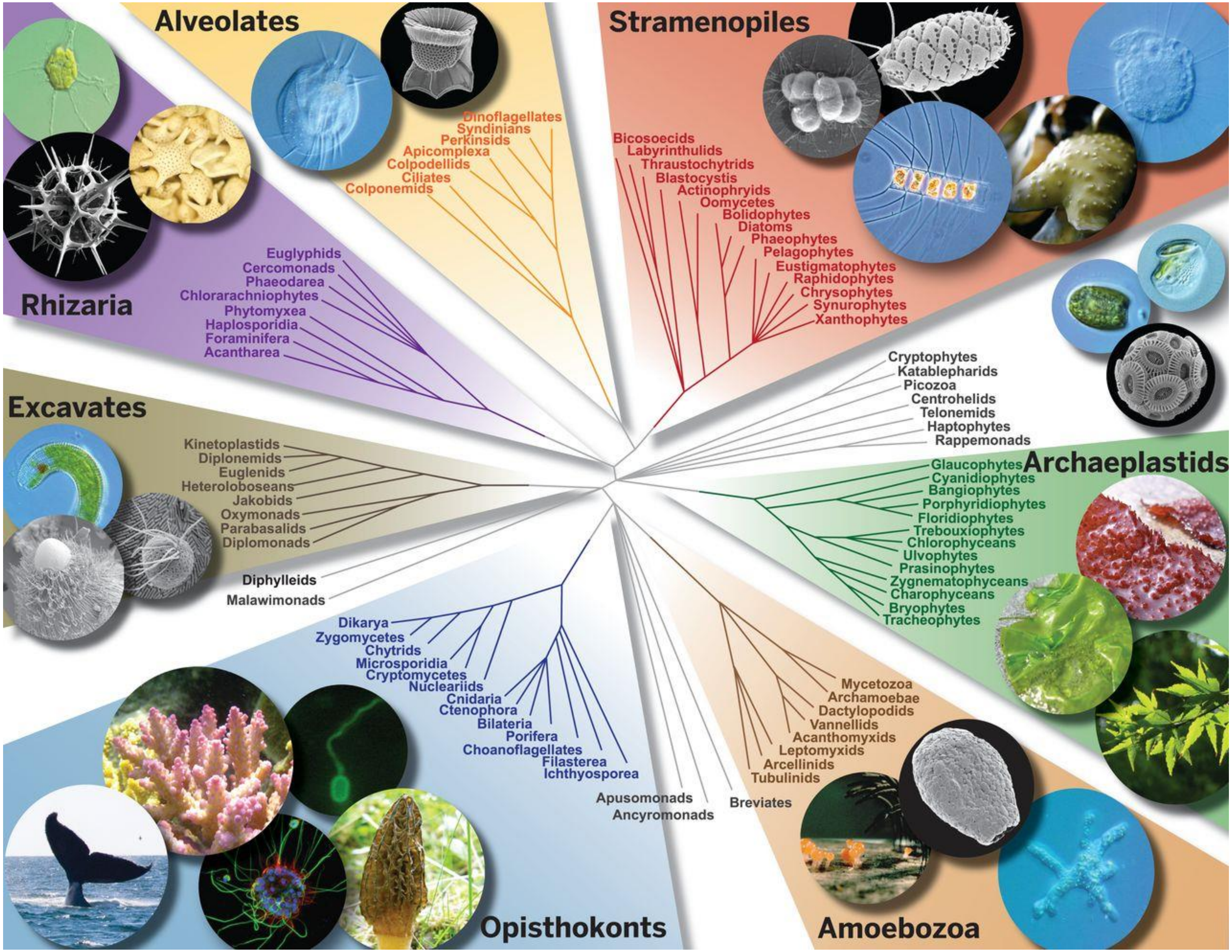
Identifying microbial predators by RNA-SIP



Incubation & Stable Isotope Probing

Orsi et al. 2016

The eukaryotic tree of life



Worden
et al. 2015