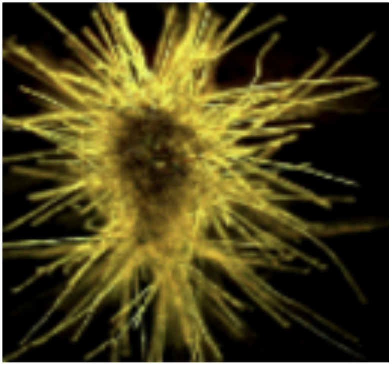
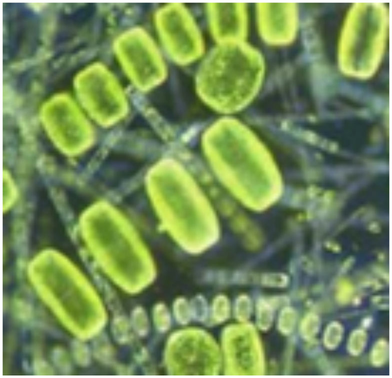
Vignettes



- From genome to biome: Tracking the metabolism and microbiome of a keystone N_2 fixer

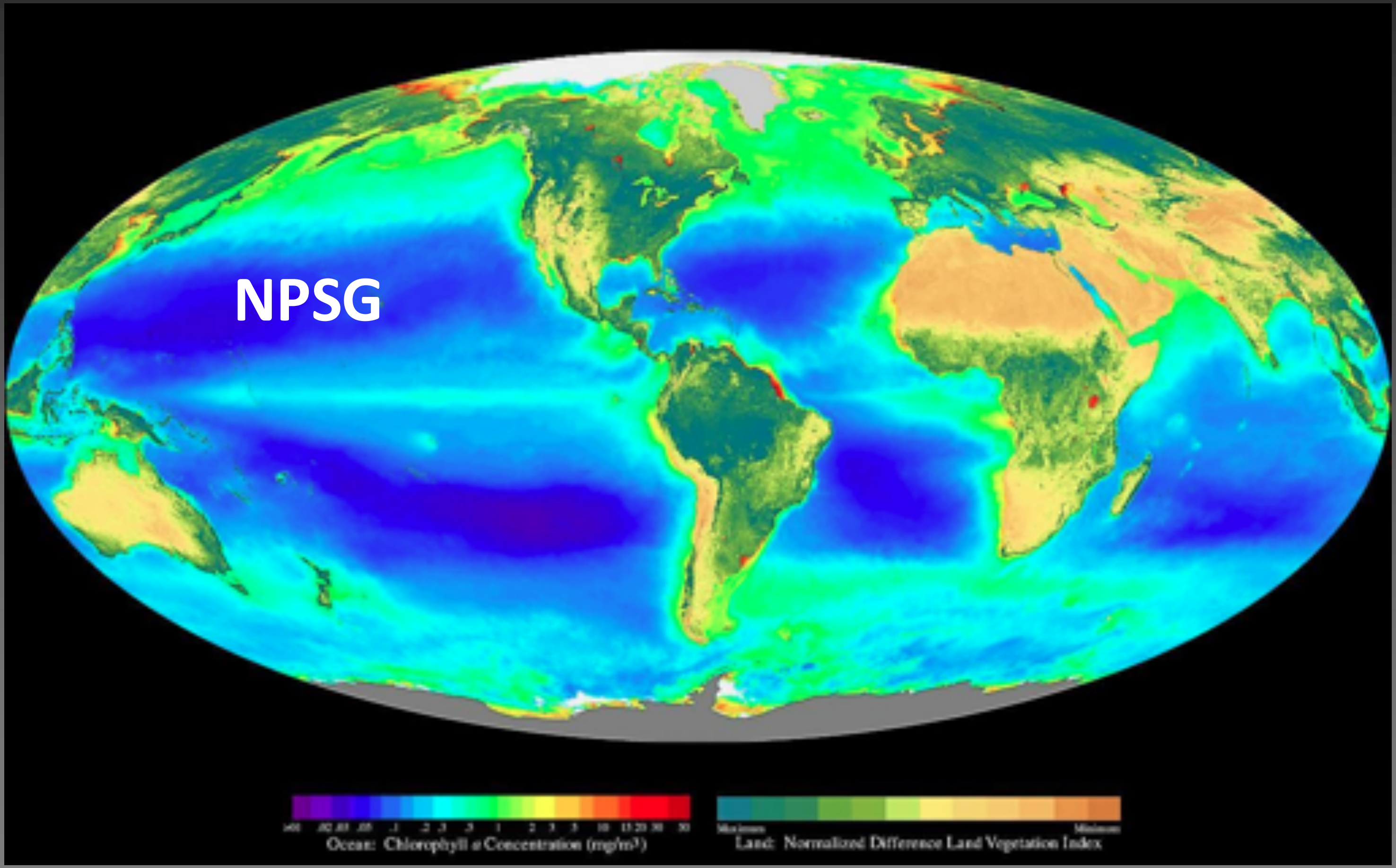
Genome - enabled



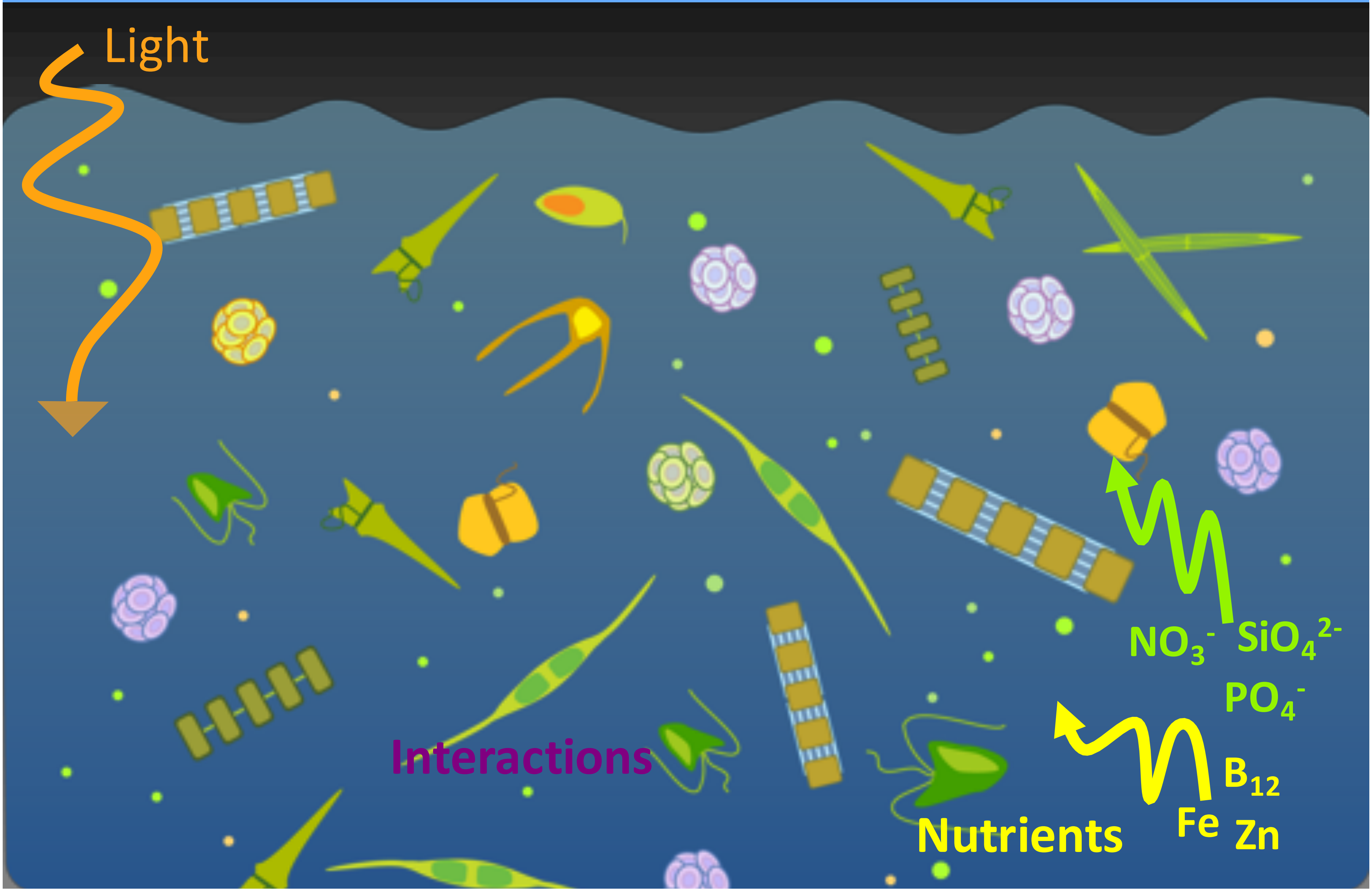
- Co-existing in a sea of competition: Leveraging transcriptome data to track the physiological ecology of phytoplankton from key groups

Transcriptome - enabled

North Pacific is one of the largest biomes on the planet

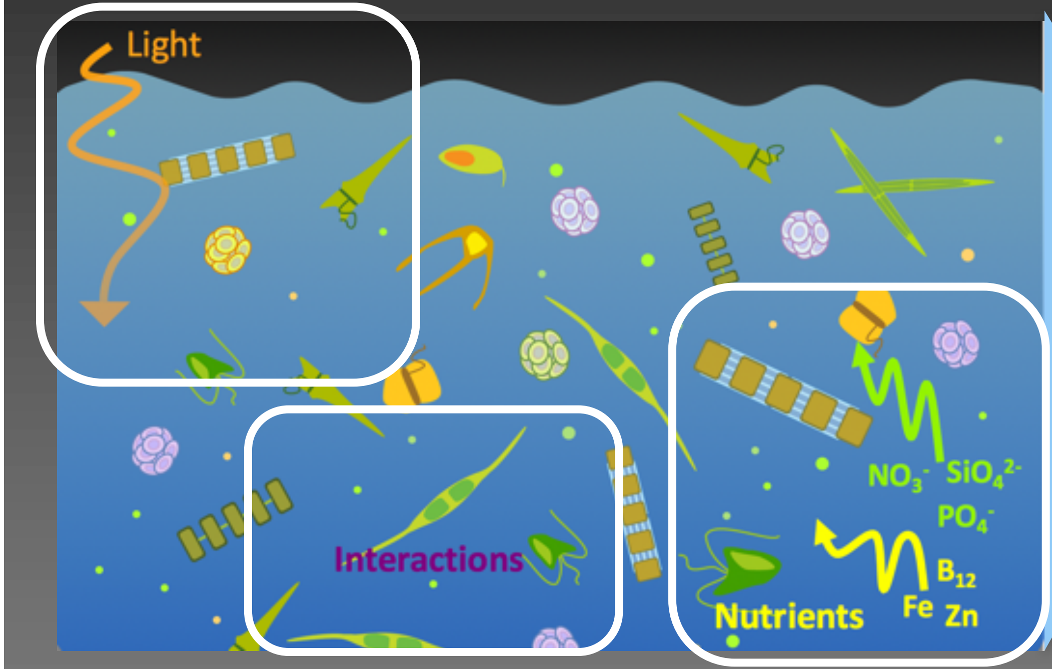


Complex community dynamics driven in part by resources



Tracking phytoplankton physiological ecology

Ecosystem function and biogeochemistry



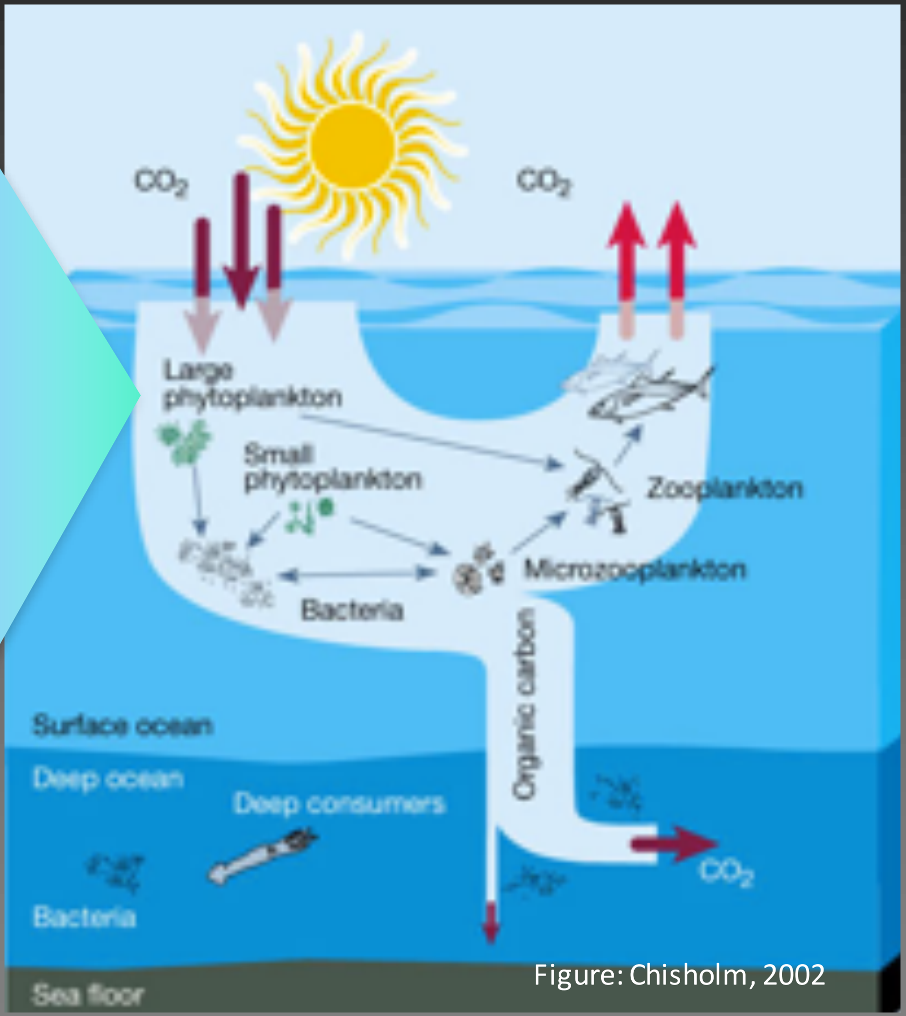


Figure: Chisholm, 2002

How do these factors drive phytoplankton distributions and activities?

Core questions

Nutrient responses

- Proof of concept: How do different groups respond to nutrient pulses?

Light responses

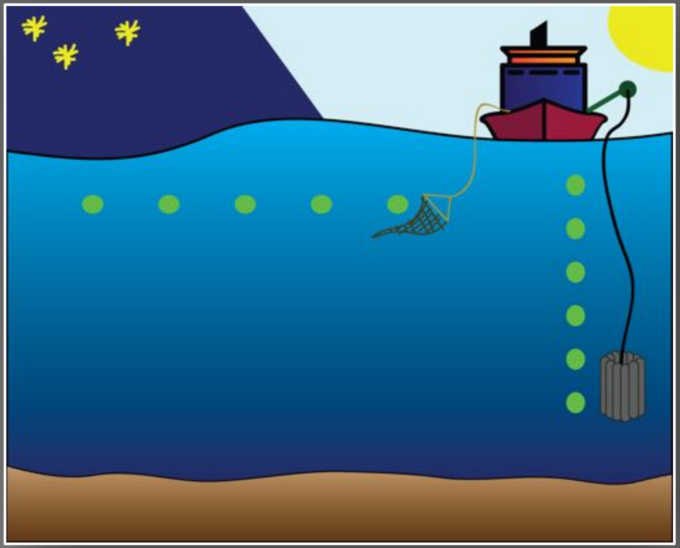
- How does light drive metabolic cascades across phyla?

Interactions

- Are there patterns of coordinated expression that underpin host-symbiont physiology?
- How does light drive interactions between the microbiome and host?

HOE:DYLAN

MAHALO



Challenges and opportunities in microbial oceanography

- Long standing challenges:
    - Populations are dilute
    - Few species-specific assays
    - Few genome or transcriptome sequences
    - Meta 'omics approaches for marine eukaryotes are still in their infancy
    - Diverse communities

- Can we get enough RNA?
- How deep do we sequence?
- Can we ID reads with sufficient resolution?

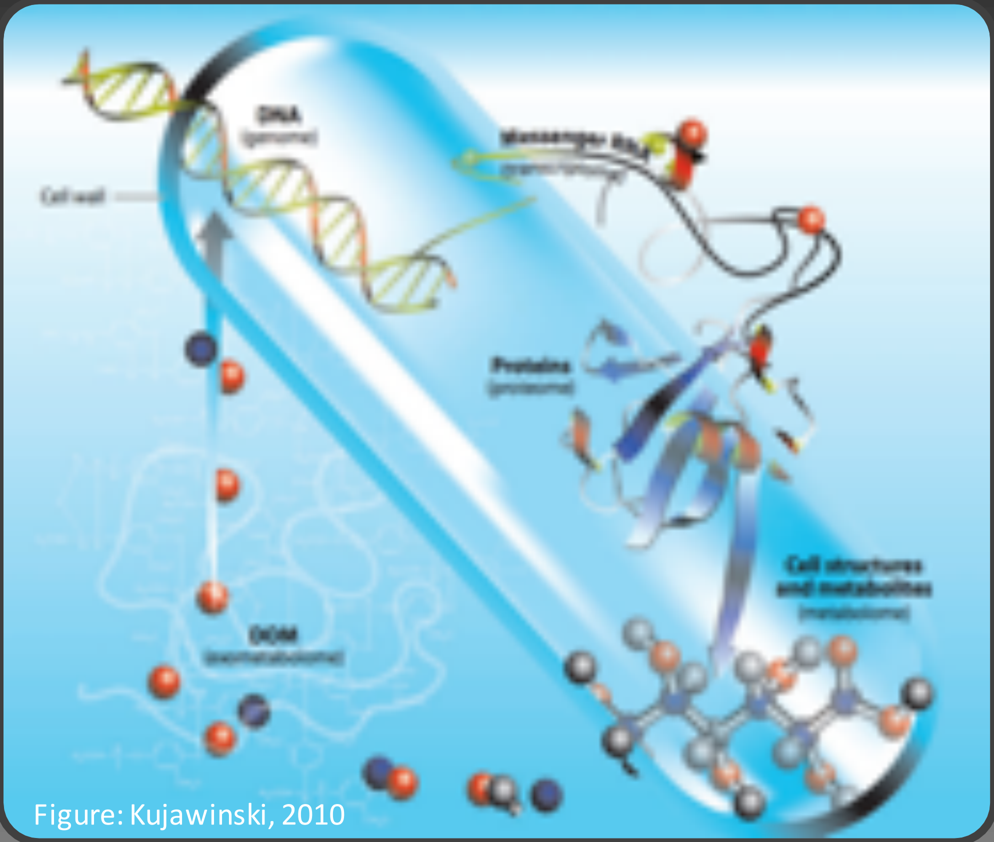


Figure: Kujawinski, 2010

Core questions

Nutrient responses

- Proof of concept: How do different groups respond to nutrient pulses?

Light responses

- How does light drive metabolic cascades across phyla?

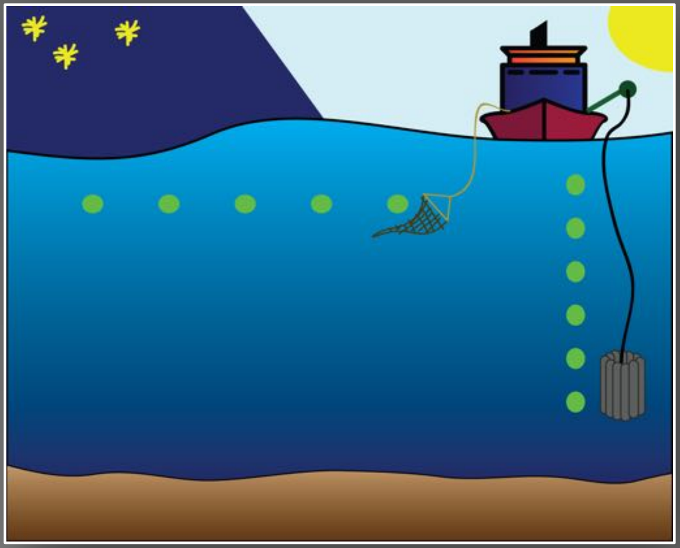
Interactions

- Are there patterns of coordinated expression that underpin host-symbiont physiology?

- How does light drive interactions between the microbiome and host?

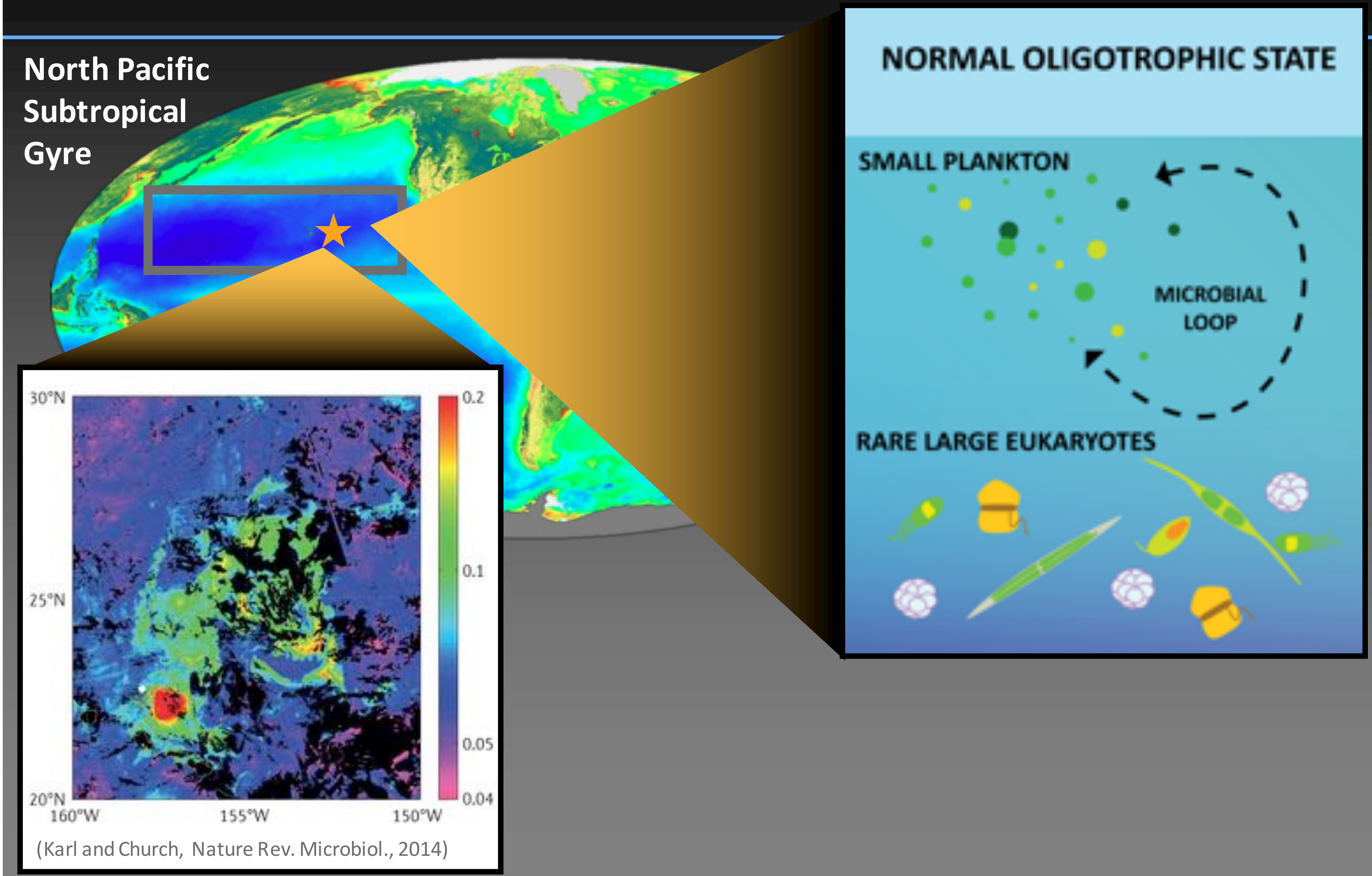
HOE:DYLAN

MAHALO



The significance of the NPSG

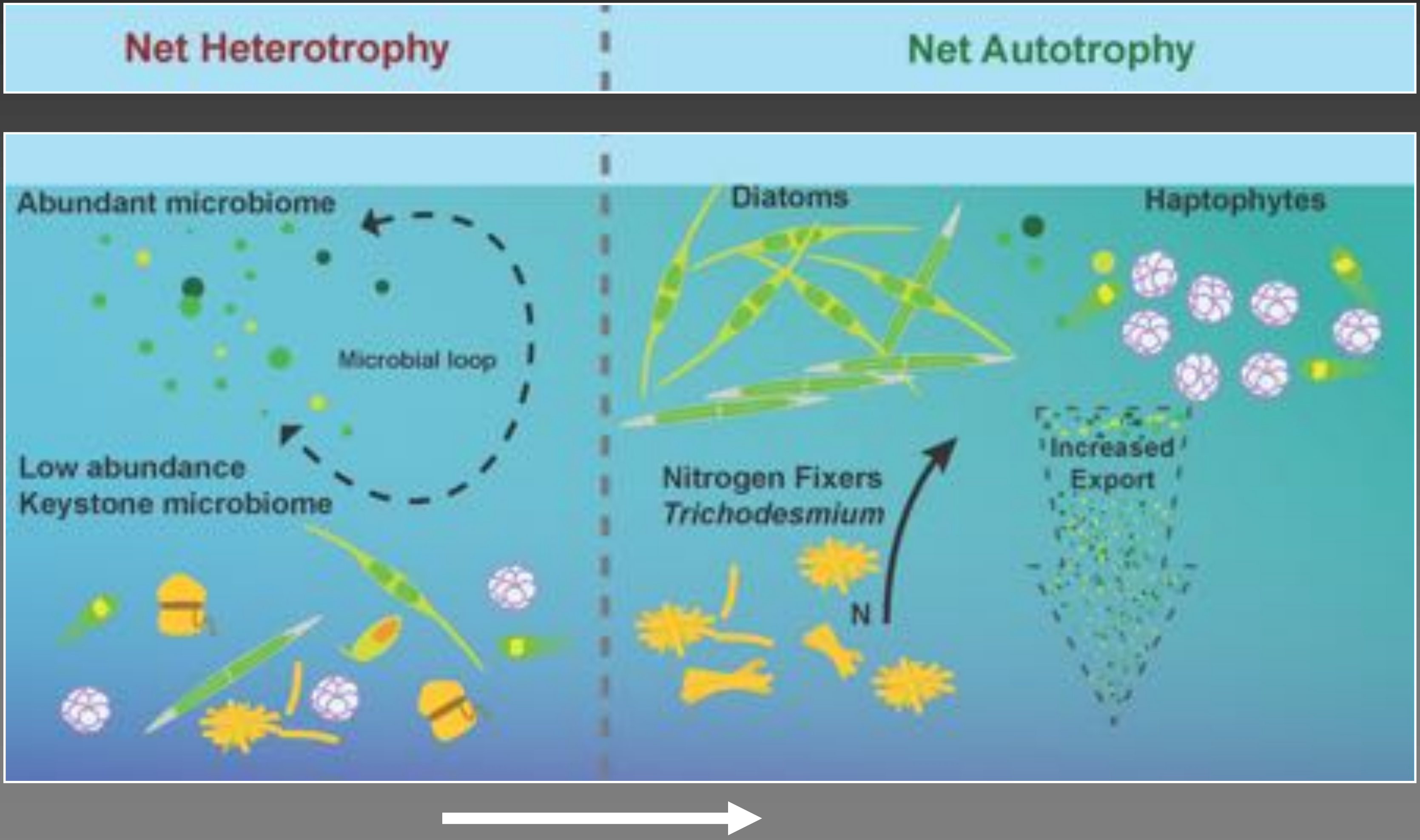
North Pacific Subtropical Gyre



(Karl and Church, Nature Rev. Microbiol., 2014)

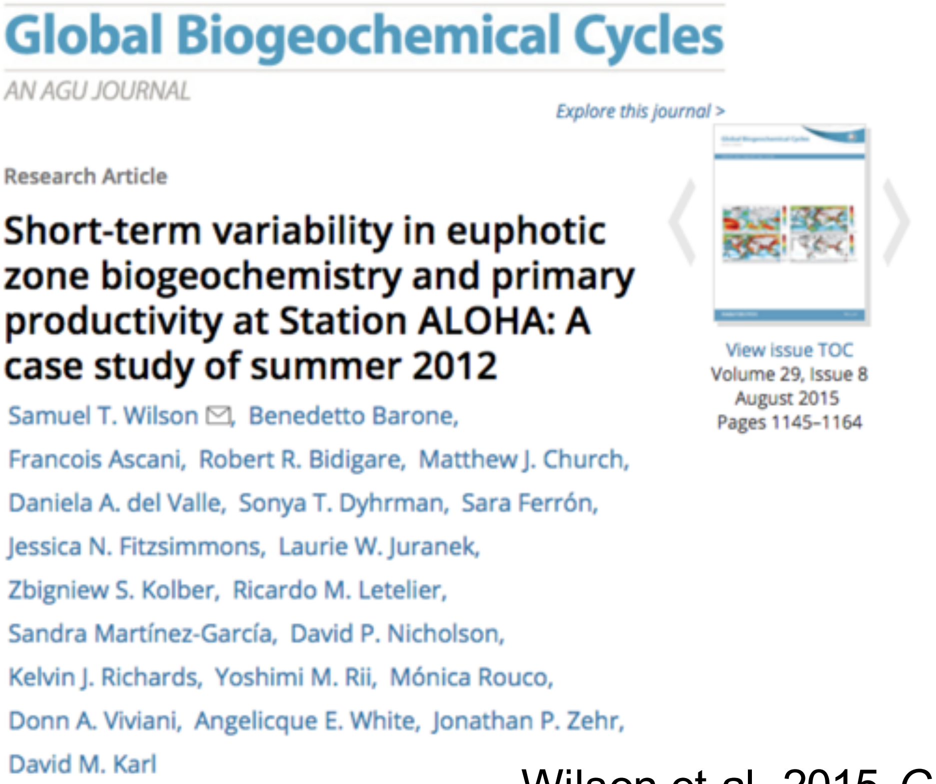
Image credits: SeaWIFS Global Chlorophyll

Blooms of keystone species can shift ecosystem state



Nutrient input,or other forcings

Hawaii Ocean Experiment: Dynamics of Light and Nutrients



Global Biogeochemical Cycles

AN AGU JOURNAL

Explore this journal >

Research Article

Short-term variability in euphotic zone biogeochemistry and primary productivity at Station ALOHA: A case study of summer 2012

Samuel T. Wilson, Benedetto Barone, Francois Ascani, Robert R. Bidigare, Matthew J. Church, Daniela A. del Valle, Sonya T. Dyhrman, Sara Ferrón, Jessica N. Fitzsimmons, Laurie W. Juranek, Zbigniew S. Kolber, Ricardo M. Letelier, Sandra Martínez-García, David P. Nicholson, Kelvin J. Richards, Yoshimi M. Rii, Mónica Rouco, Donn A. Viviani, Angelicque E. White, Jonathan P. Zehr, David M. Karl

View issue TOC
Volume 29, Issue 8
August 2015
Pages 1145–1164

Wilson et al. 2015 *GBC*

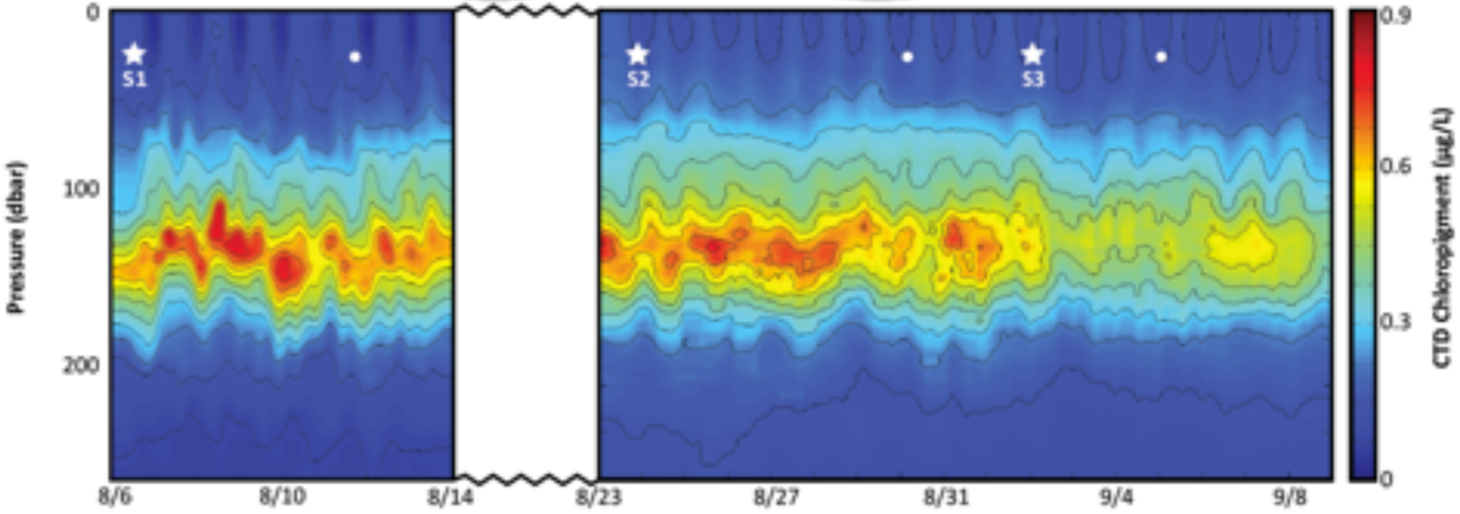
center for microbial oceanography: research and education

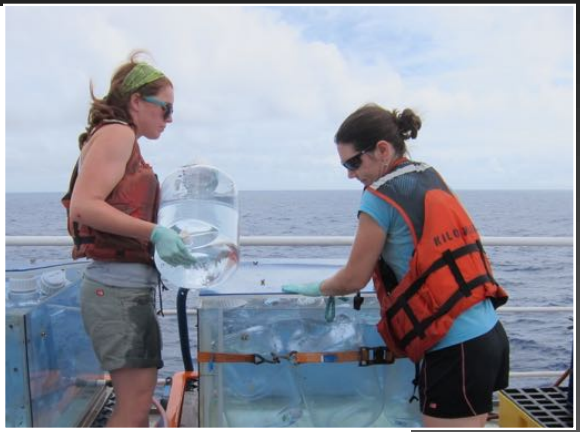
c·more *linking genomes to biomes*

(HOE-DYLAN)

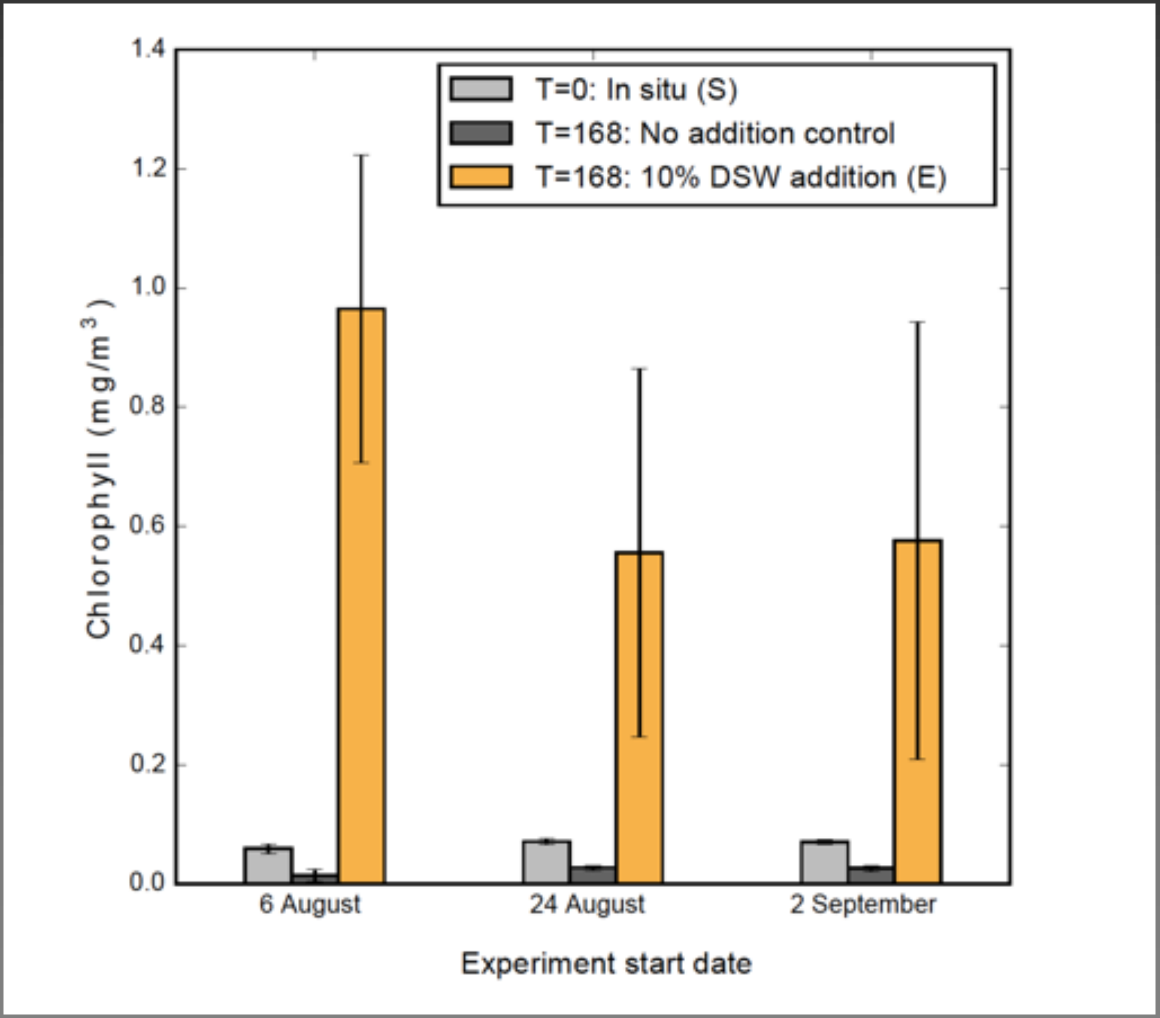
Sampling a bloom

Sampling at Station ALOHA coupled with bloom simulations



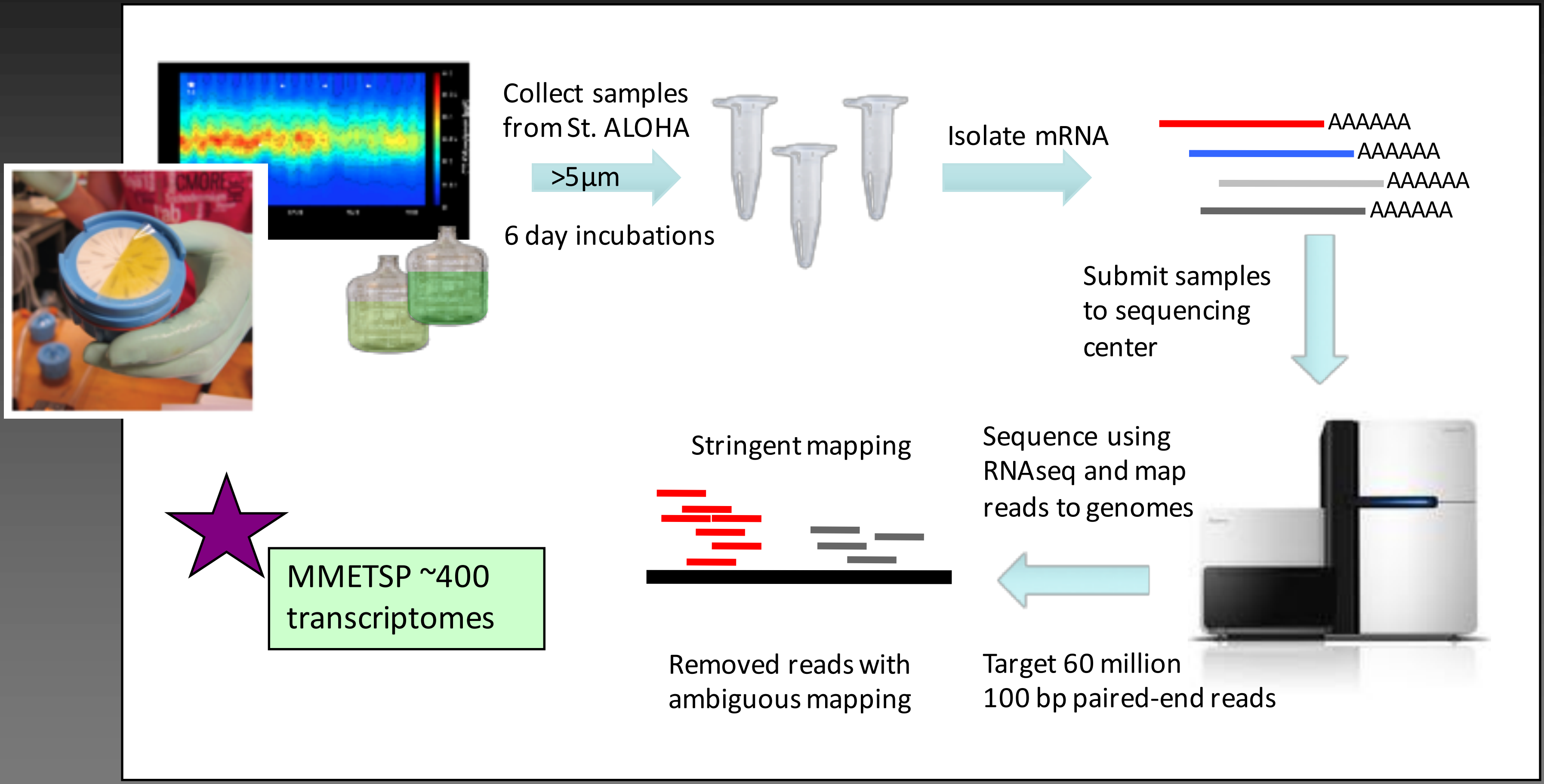


Deep water addition led to simulated “bloom”




CMORE
lab

Sampling and pipeline



Collect samples from St. ALOHA

>5µm

6 day incubations

Isolate mRNA

Submit samples to sequencing center

Sequence using RNAseq and map reads to genomes

Target 60 million 100 bp paired-end reads

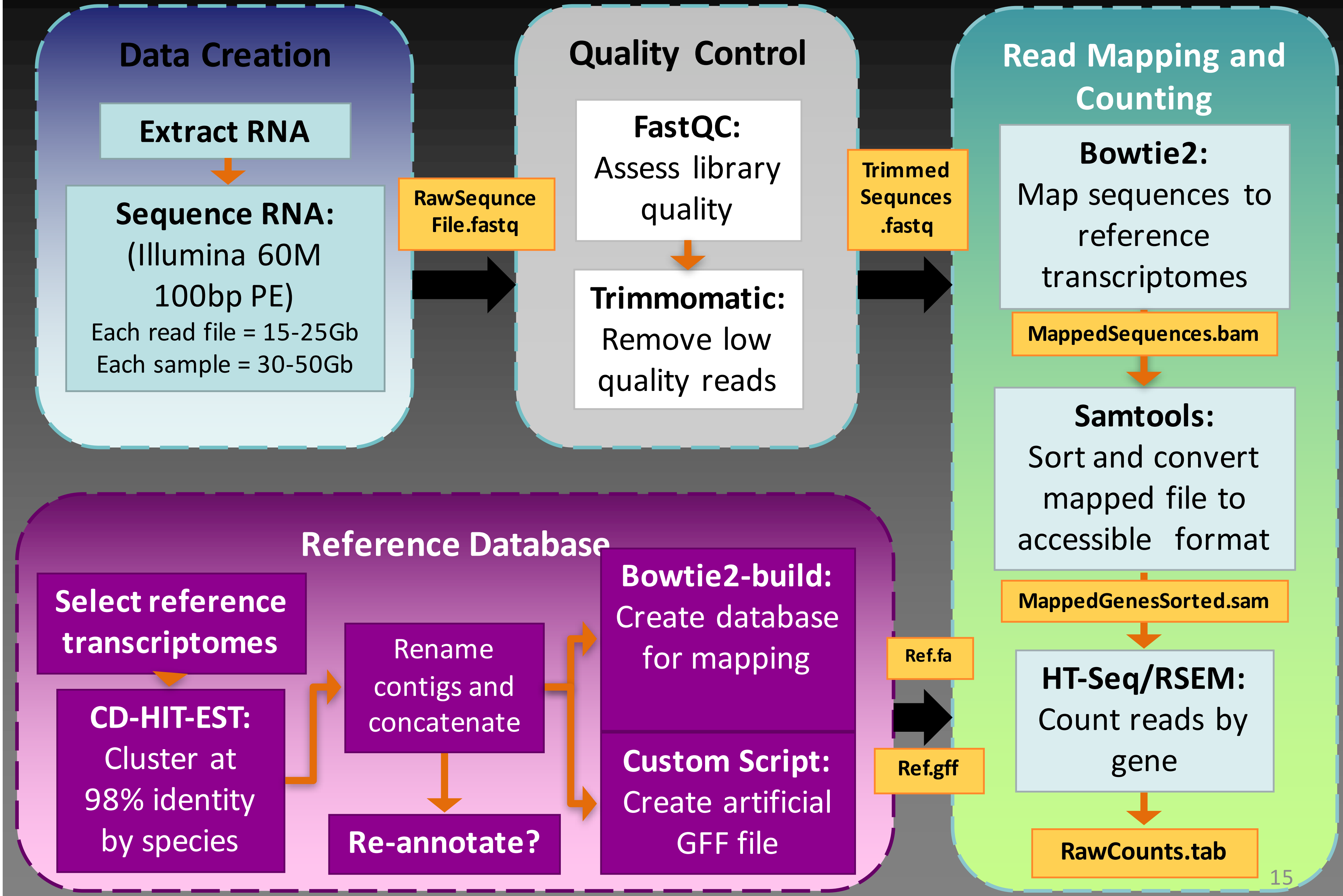
Stringent mapping

Removed reads with ambiguous mapping

MMETSP ~400 transcriptomes

MMETSP = Marine Microbial Eukaryotic Transcriptome Project

Metatranscriptome Data Pipeline - the gory details ....



Data Creation

Extract RNA

Sequence RNA:
(Illumina 60M 100bp PE)
Each read file = 15-25Gb
Each sample = 30-50Gb

RawSequnce File.fastq

Quality Control

FastQC:
Assess library quality

Trimmomatic:
Remove low quality reads

Trimmed Sequnces .fastq

Read Mapping and Counting

Bowtie2:
Map sequences to reference transcriptomes

MappedSequences.bam

Samtools:
Sort and convert mapped file to accessible format

MappedGenesSorted.sam

HT-Seq/RSEM:
Count reads by gene

RawCounts.tab

Reference Database

Select reference transcriptomes

CD-HIT-EST:
Cluster at 98% identity by species

Rename contigs and concatenate

Re-annotate?

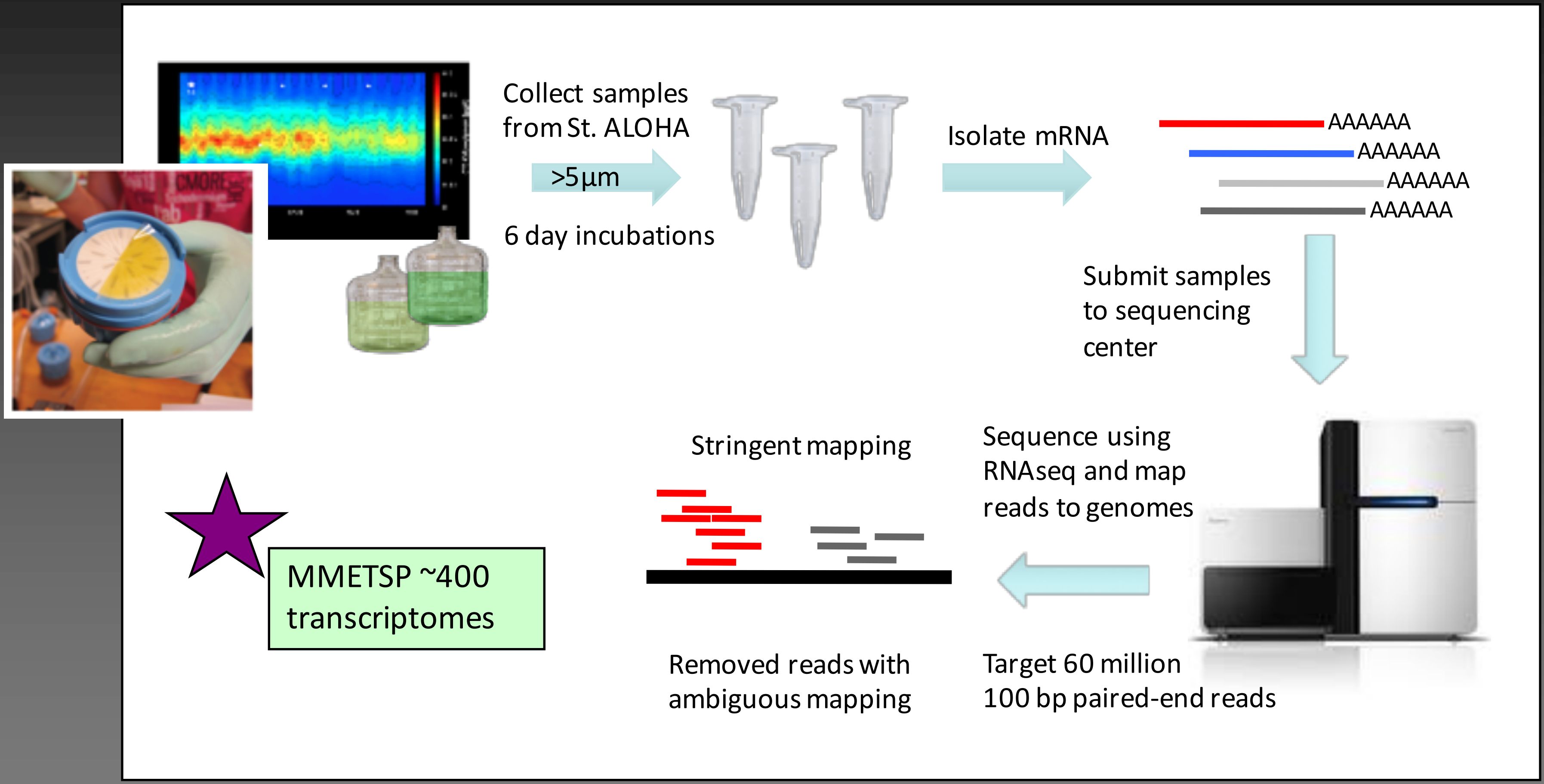
Bowtie2-build:
Create database for mapping

Custom Script:
Create artificial GFF file

Ref.fa

Ref.gff

Sampling and pipeline



Collect samples from St. ALOHA

>5µm

6 day incubations

Isolate mRNA

Submit samples to sequencing center

Sequence using RNAseq and map reads to genomes

Target 60 million 100 bp paired-end reads

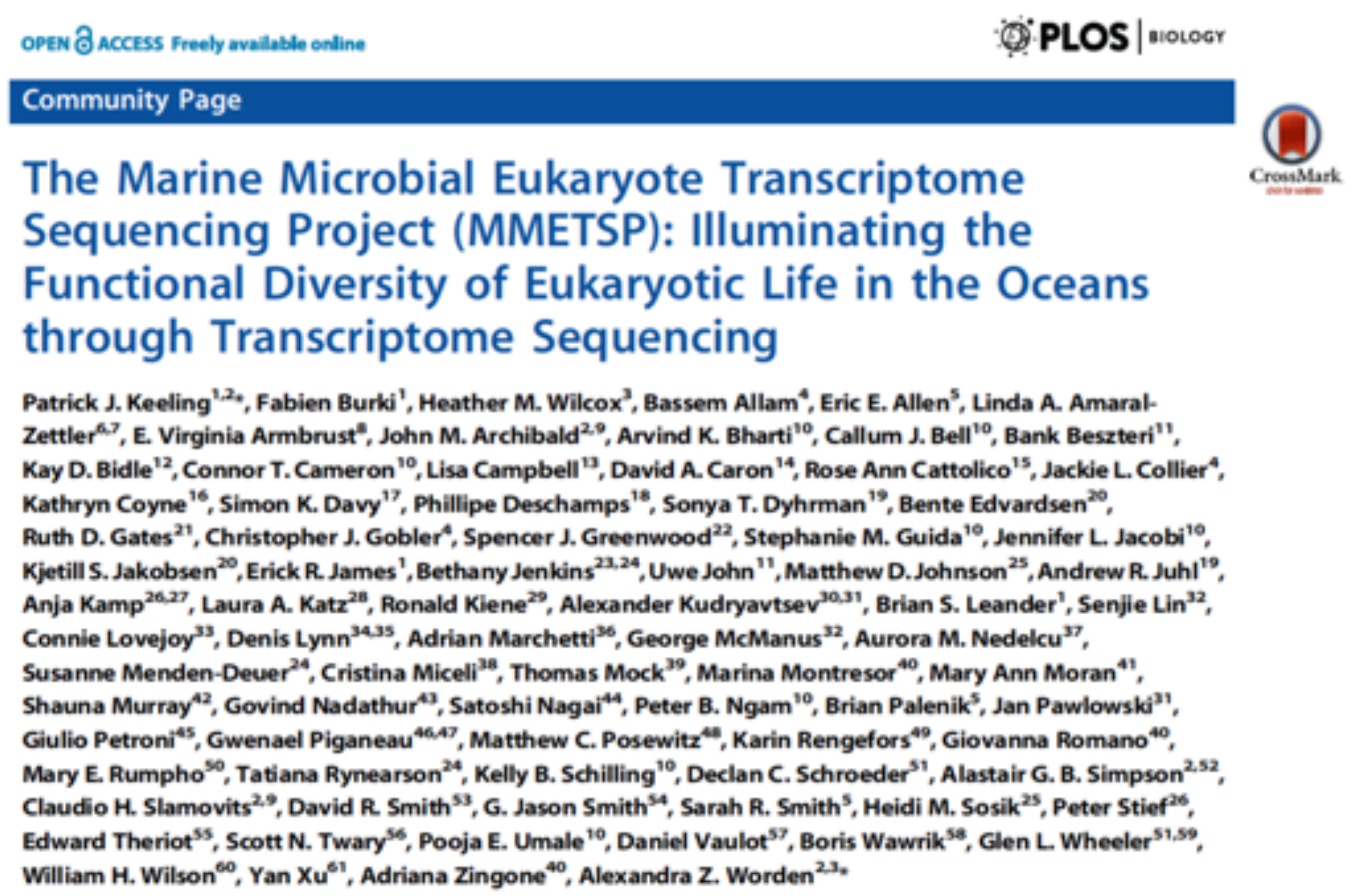
Stringent mapping

Removed reads with ambiguous mapping

MMETSP ~400 transcriptomes

MMETSP = Marine Microbial Eukaryotic Transcriptome Project

Marine Microbial Eukaryote Transcriptome Sequencing Project



OPEN ACCESS Freely available online

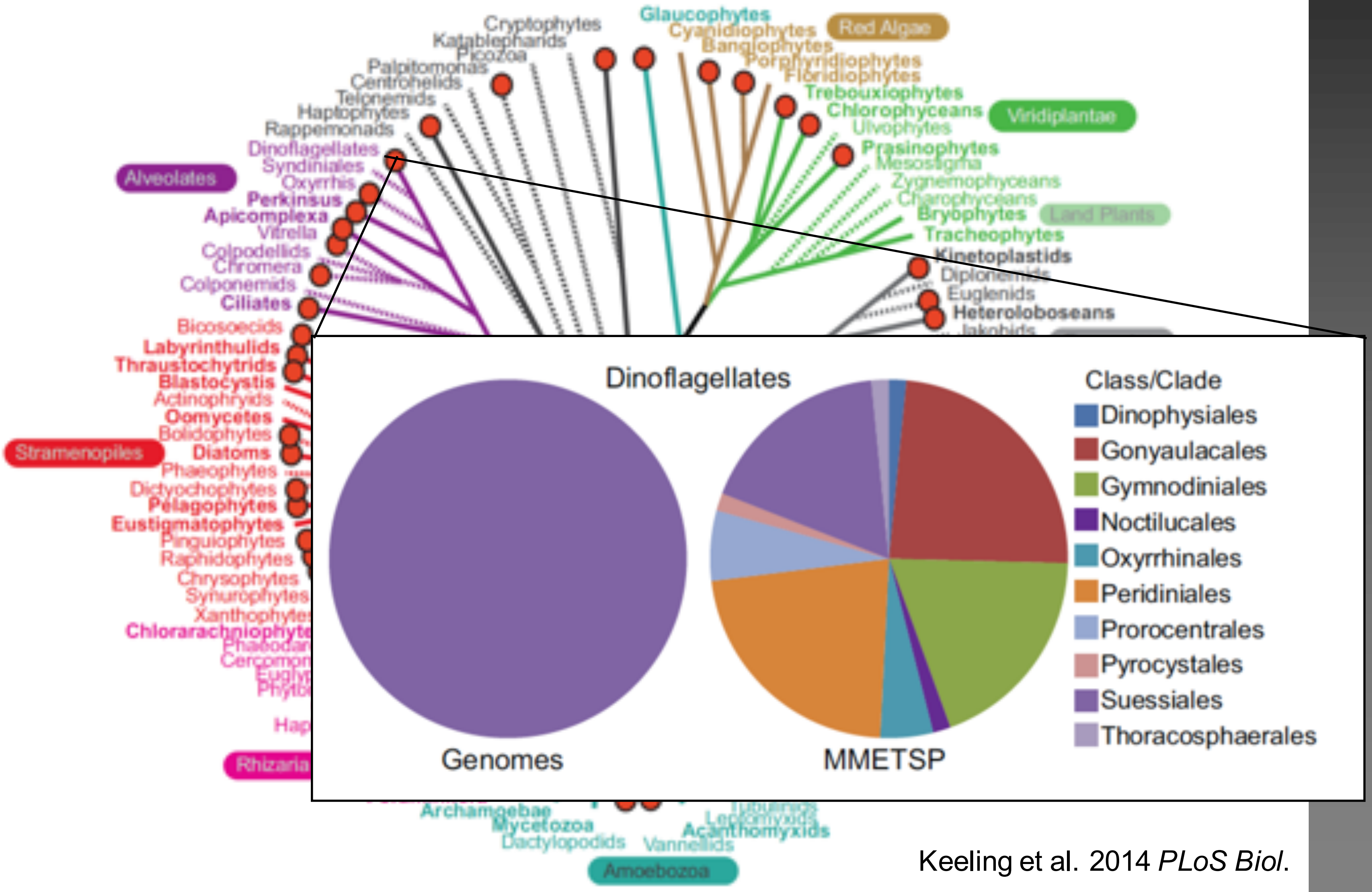
PLOS | BIOLOGY

Community Page

The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing

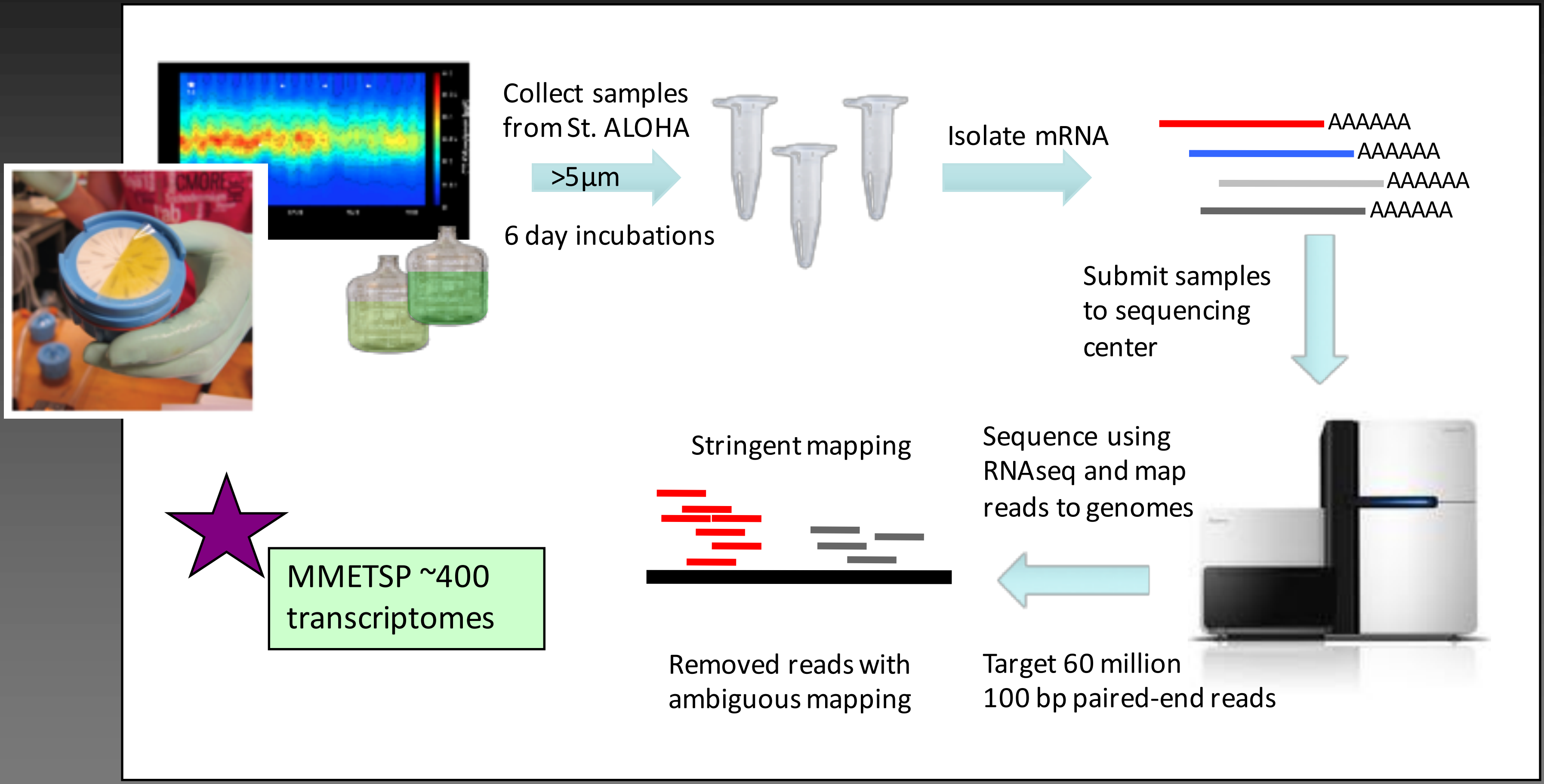
Patrick J. Keeling[1,2,*], Fabien Burki[1], Heather M. Wilcox[3], Bassem Allam[4], Eric E. Allen[5], Linda A. Amaral-Zettler[6,7], E. Virginia Armbrust[8], John M. Archibald[2,9], Arvind K. Bharti[10], Callum J. Bell[10], Bank Beszteri[11], Kay D. Bidle[12], Connor T. Cameron[10], Lisa Campbell[13], David A. Caron[14], Rose Ann Cattolico[15], Jackie L. Collier[4], Kathryn Coyne[16], Simon K. Davy[17], Phillipe Deschamps[18], Sonya T. Dyhrman[19], Bente Edvardsen[20], Ruth D. Gates[21], Christopher J. Gobler[4], Spencer J. Greenwood[22], Stephanie M. Guida[10], Jennifer L. Jacobi[10], Kjetill S. Jakobsen[20], Erick R. James[1], Bethany Jenkins[23,24], Uwe John[11], Matthew D. Johnson[25], Andrew R. Juhl[19], Anja Kamp[26,27], Laura A. Katz[28], Ronald Kiene[29], Alexander Kudryavtsev[30,31], Brian S. Leander[1], Senjie Lin[32], Connie Lovejoy[33], Denis Lynn[34,35], Adrian Marchetti[36], George McManus[32], Aurora M. Nedelcu[37], Susanne Menden-Deuer[24], Cristina Miceli[38], Thomas Mock[39], Marina Montresor[40], Mary Ann Moran[41], Shauna Murray[42], Govind Nadathur[43], Satoshi Nagai[44], Peter B. Ngam[10], Brian Palenik[5], Jan Pawlowski[31], Giulio Petroni[45], Gwenael Piganeau[46,47], Matthew C. Posewitz[48], Karin Rengefors[49], Giovanna Romano[40], Mary E. Rumpho[50], Tatiana Rynearson[24], Kelly B. Schilling[10], Declan C. Schroeder[51], Alastair G. B. Simpson[2,52], Claudio H. Slamovits[2,9], David R. Smith[53], G. Jason Smith[54], Sarah R. Smith[5], Heidi M. Sosik[25], Peter Stief[26], Edward Theriot[55], Scott N. Twary[56], Pooja E. Umale[10], Daniel Vaulot[57], Boris Wawrik[58], Glen L. Wheeler[51,59], William H. Wilson[60], Yan Xu[61], Adriana Zingone[40], Alexandra Z. Worden[2,3,*]

Greatly expanded reference sequences in the tree of life



Keeling et al. 2014 *PLoS Biol*.

Sampling and pipeline



Collect samples from St. ALOHA

>5µm

6 day incubations

Isolate mRNA

Submit samples to sequencing center

Sequence using RNAseq and map reads to genomes

Target 60 million 100 bp paired-end reads

Stringent mapping

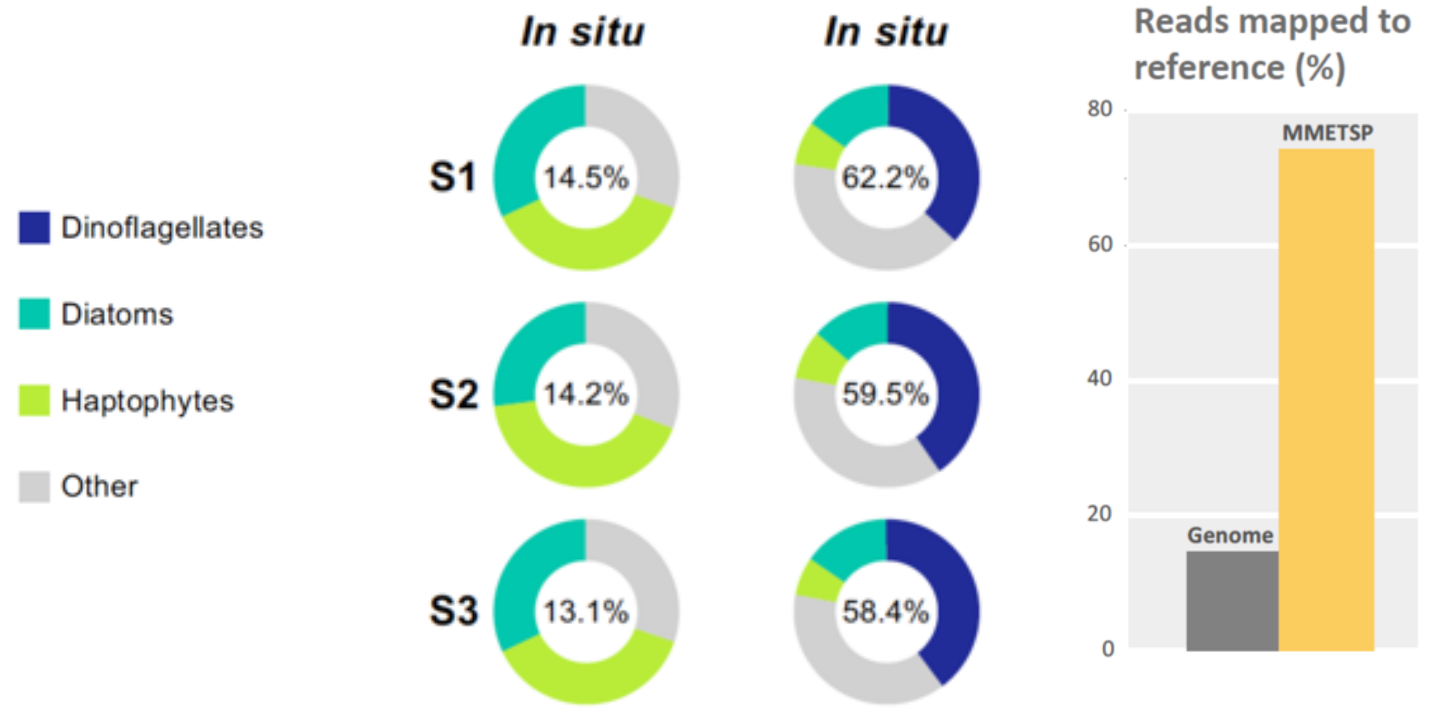
Removed reads with ambiguous mapping

MMETSP ~400 transcriptomes

MMETSP = Marine Microbial Eukaryotic Transcriptome Project

MMETSP improves read identification

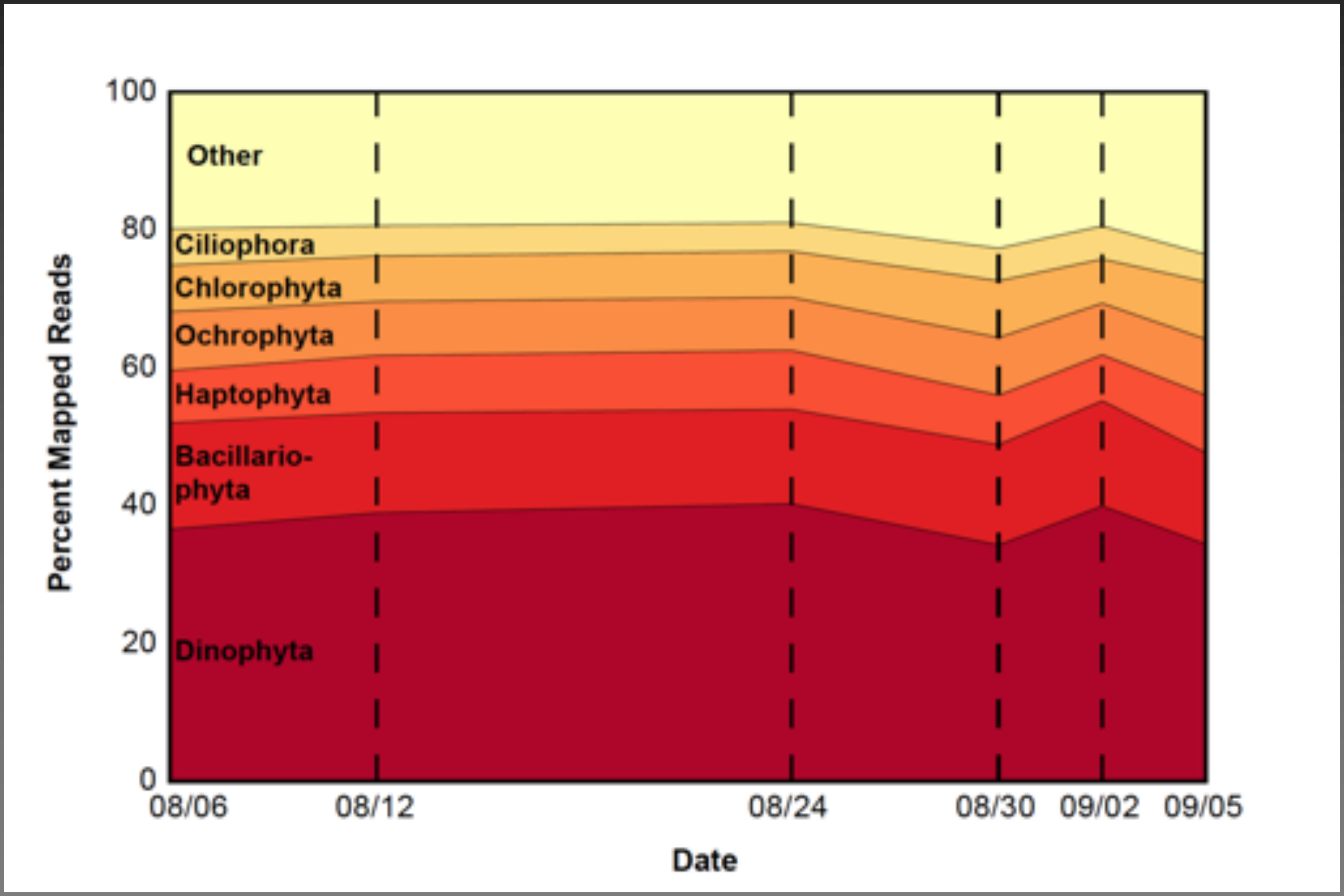
Read mapping against genomes v. MMETSP at St. ALHOA



Alexander et al. 2015b *PNAS*

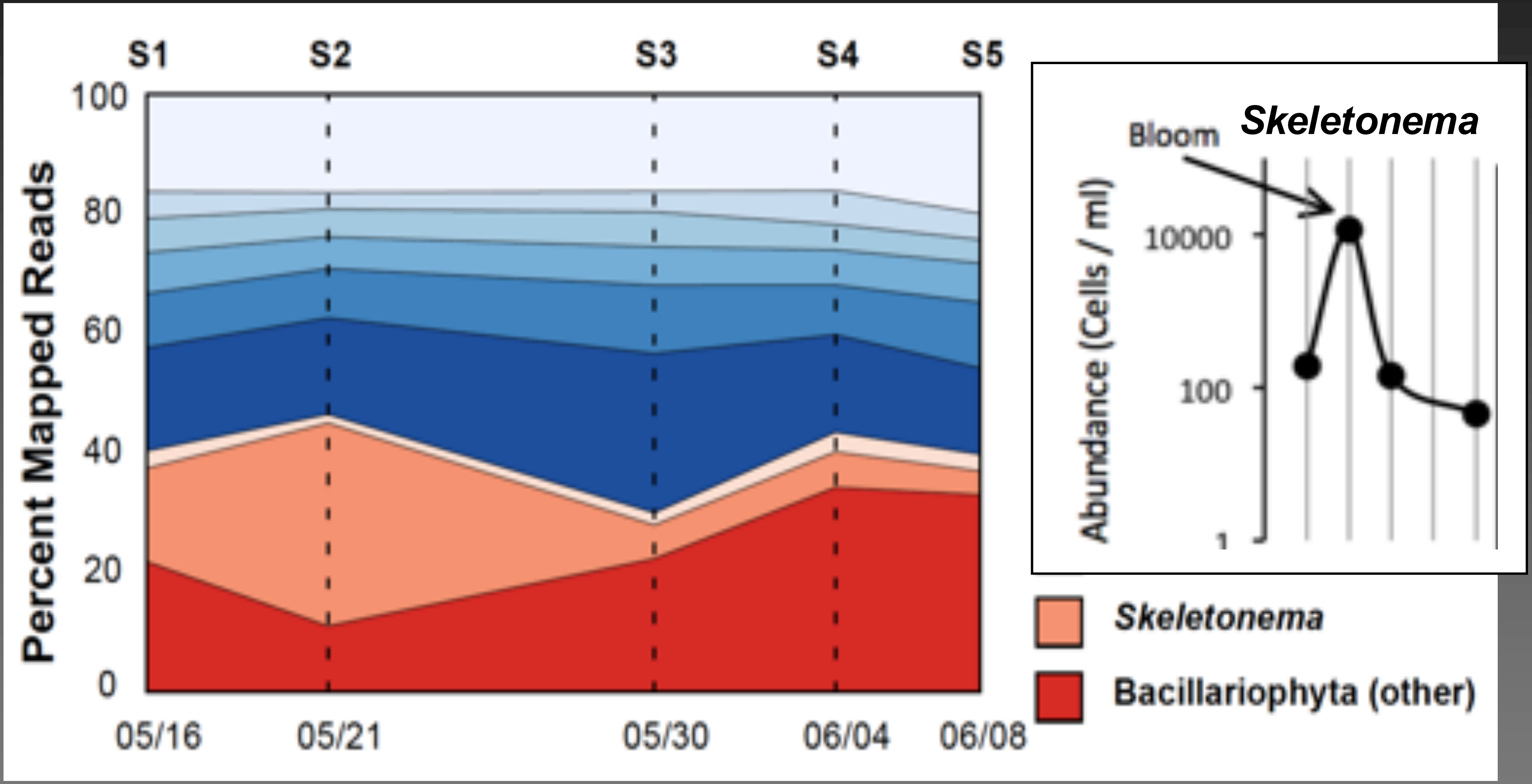
Sequence read identification is substantially improved over using phytoplankton genomes which do not capture the same diversity.

Taxonomic distribution of reads



Distributions is highly stable, much more so than what was observed in coastal system.

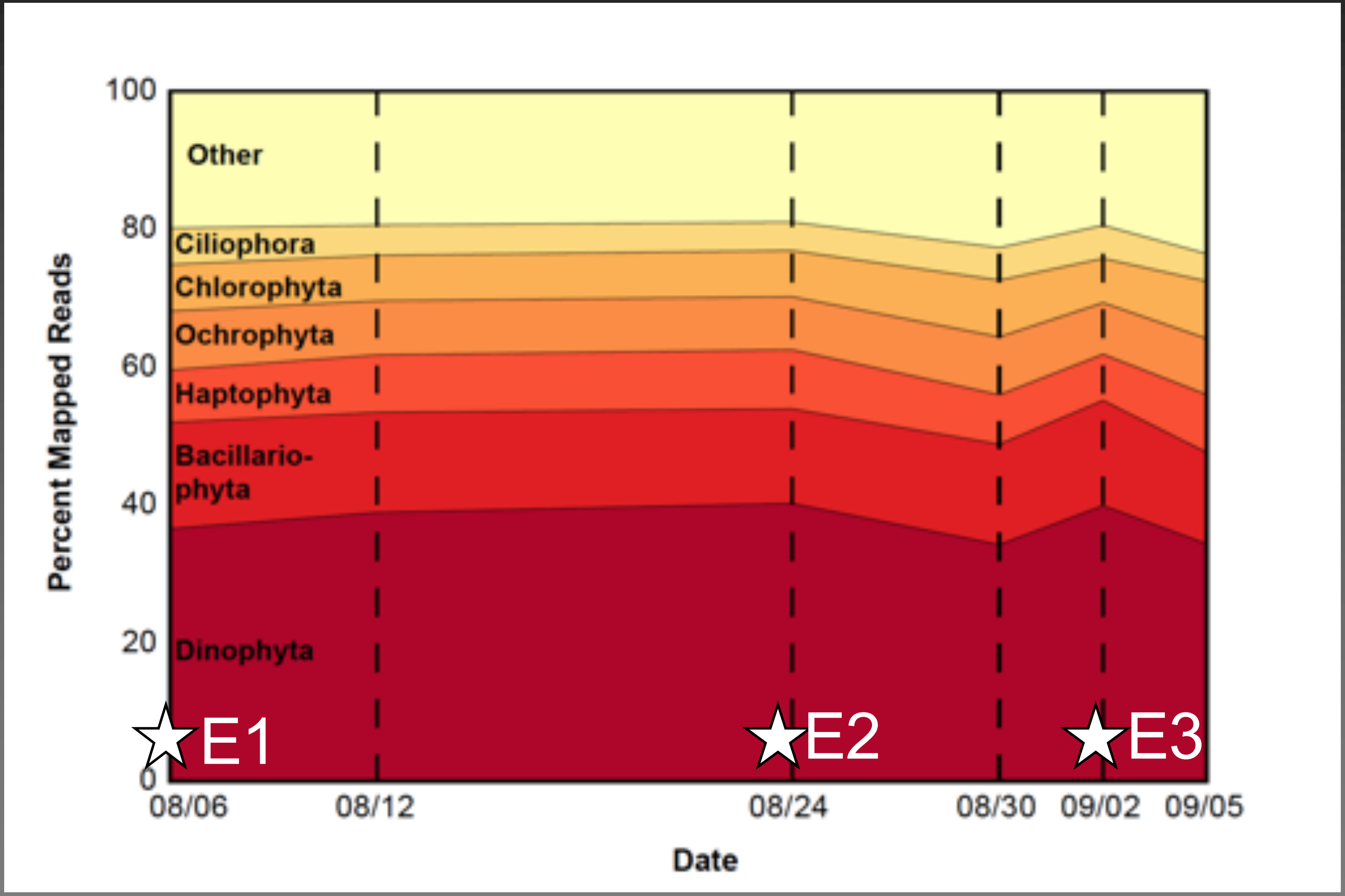
Taxonomic distribution of reads - coastal bloom



Alexander et al. 2015 *PNAS*

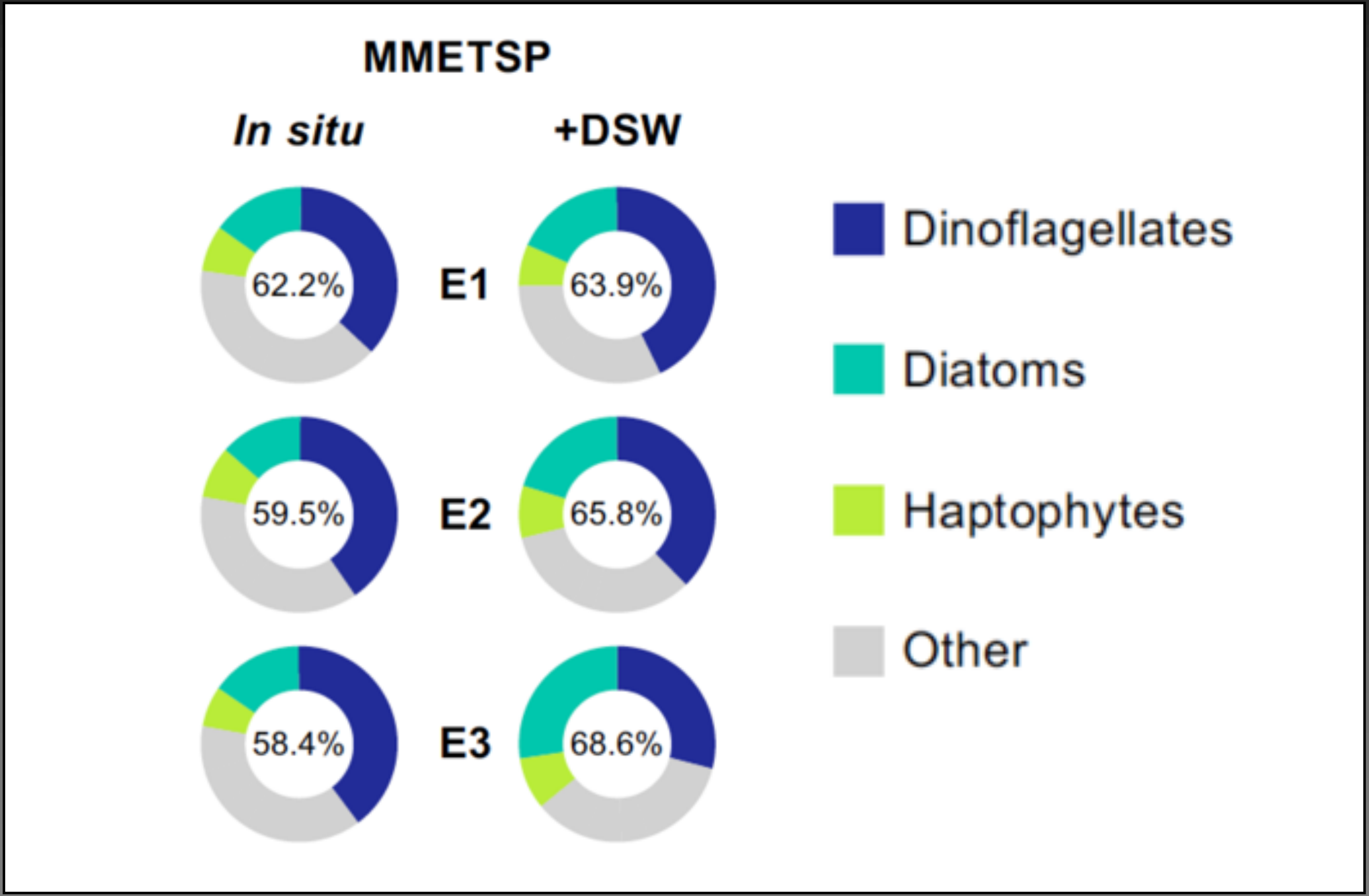
Taxonomic distribution of reads is more variable in a coastal system

Taxonomic distribution of reads



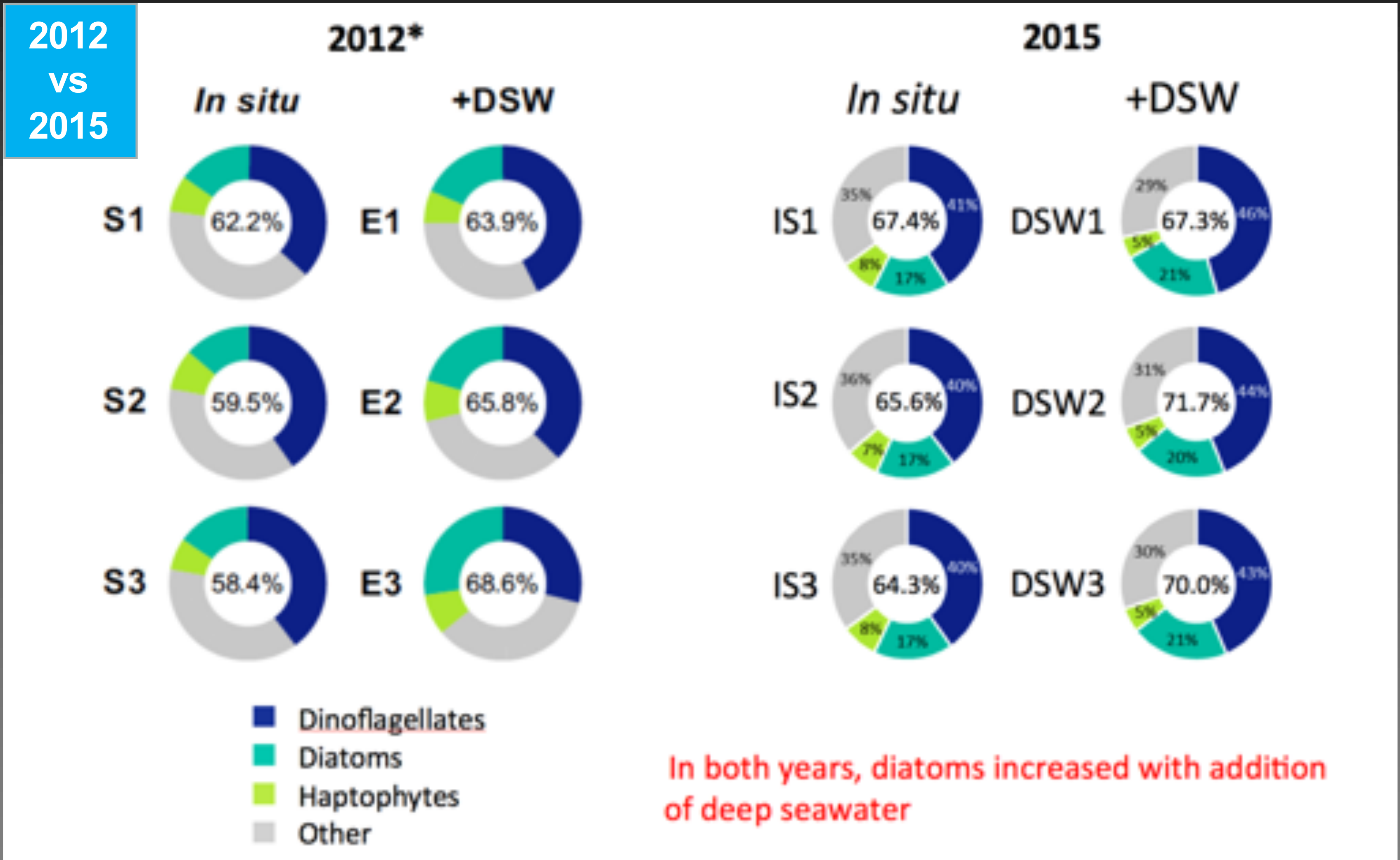
Alexander et al. 2015b *PNAS*

Increase in diatoms during simulated blooms



Read identification is robust, most are dinoflagellates, but diatoms increase in simulated blooms (+DSW)

Simulated bloom response was reproducible 3 years later

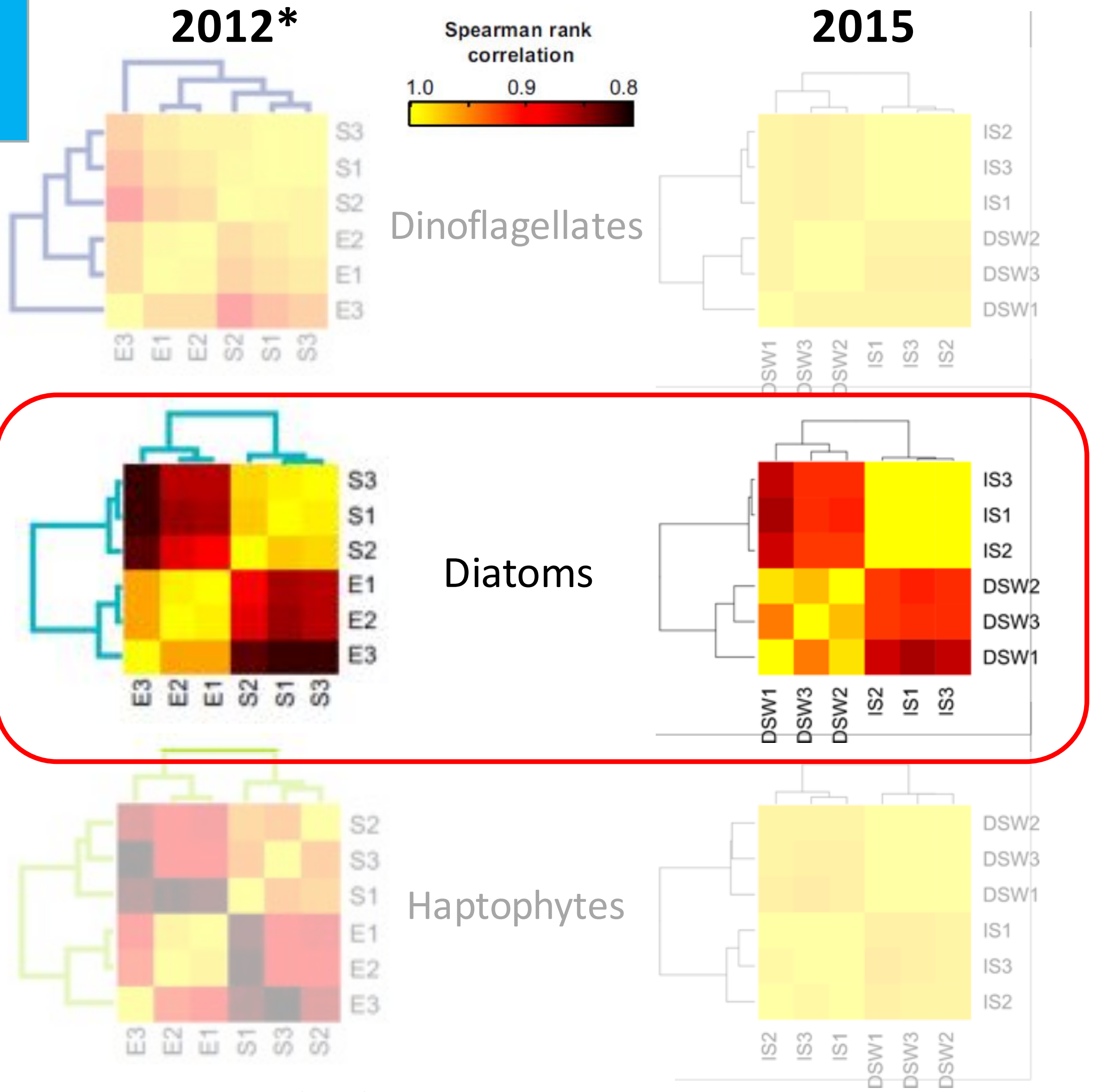


In both years, diatoms increased with addition of deep seawater

Read identification increases with larger database, remarkable similarity in response between years

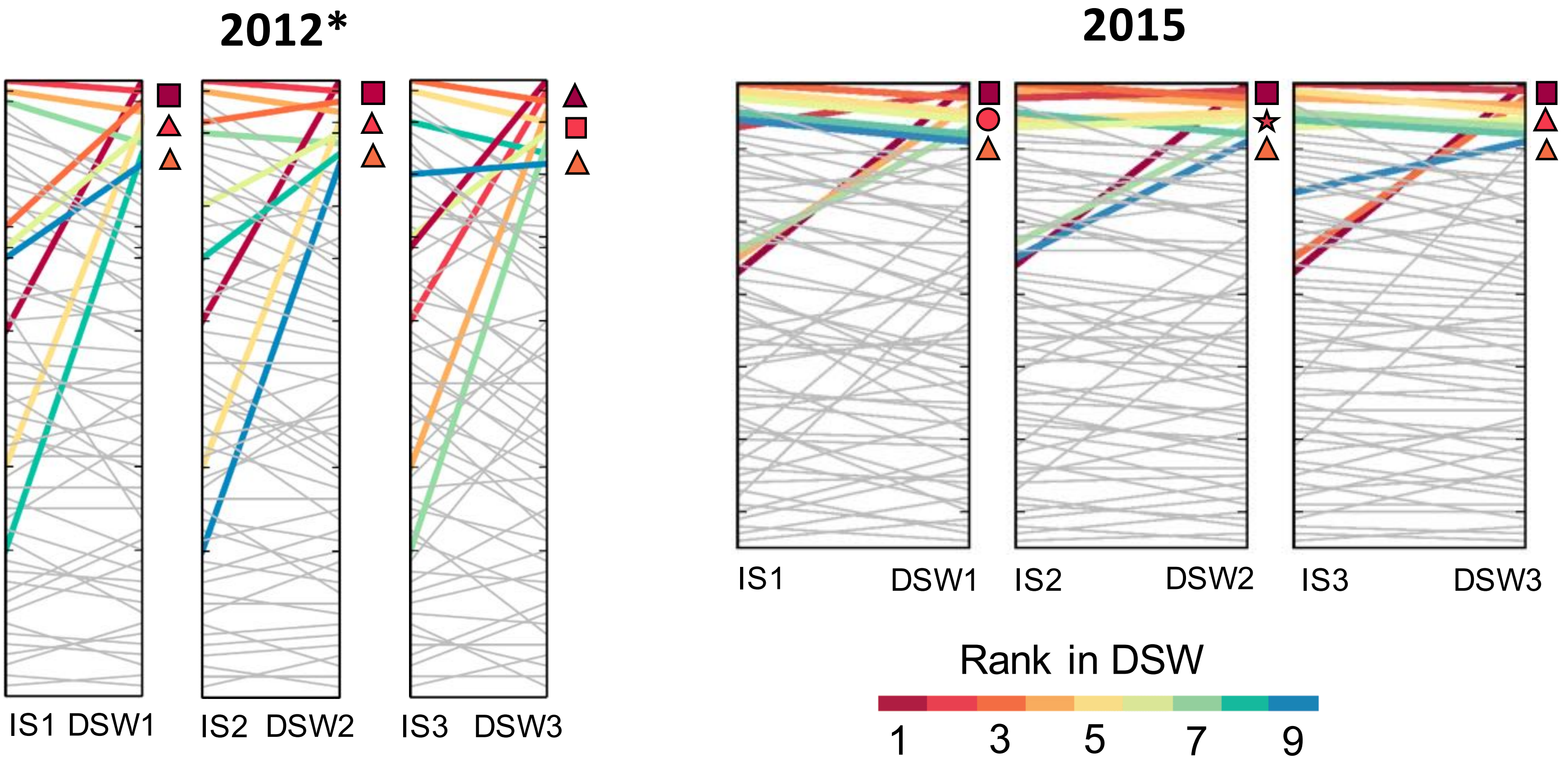
Community composition shifts are reproducible between years

2012
vs
2015



E = +DSW S = *in situ*

Its even the same species changing…

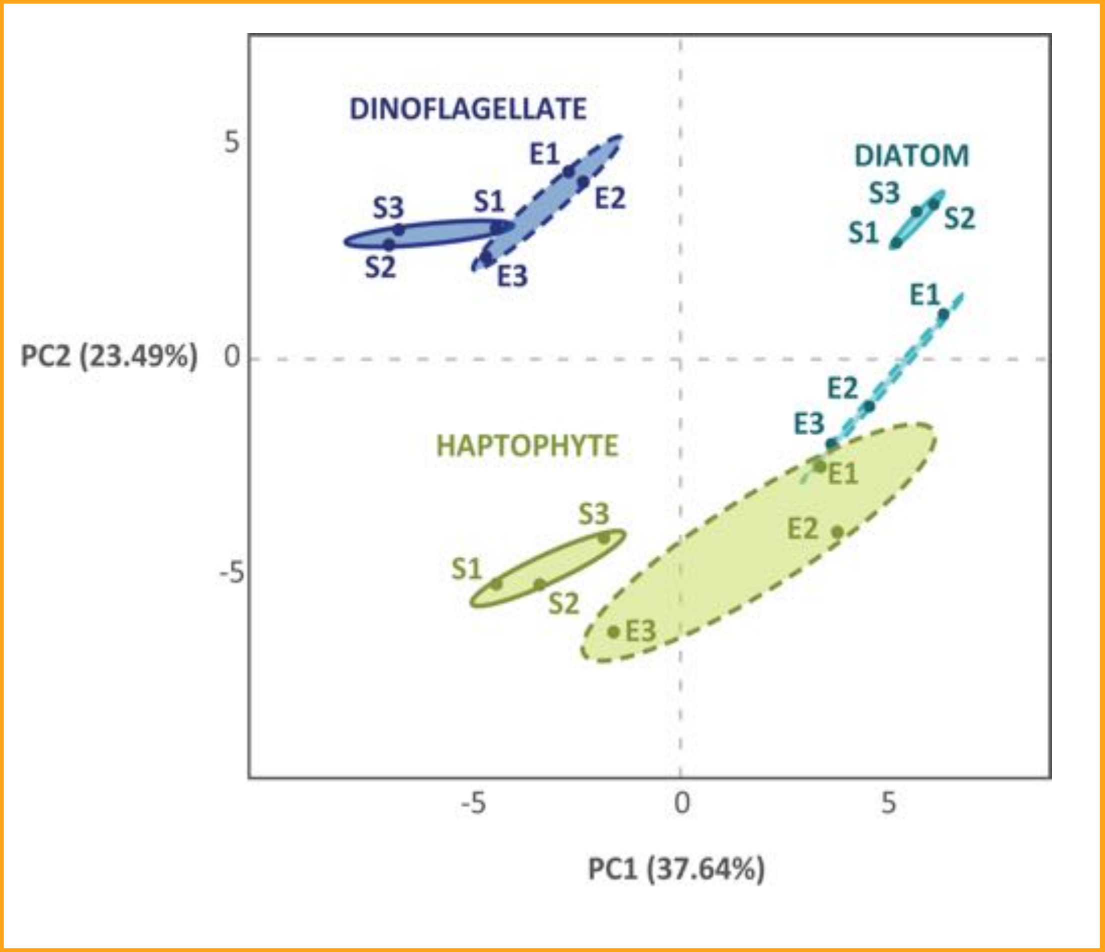


□ *Cylindrotheca sp.*
△ *Pseudo-nitzschia sp.*
○ *Chaetoceros sp.*
☆ *Ditylum sp.*

In both years, *Cylindrotheca* sp. and *Pseudo-nitzschia* sp. dominate with DSW addition

2012 vs 2015

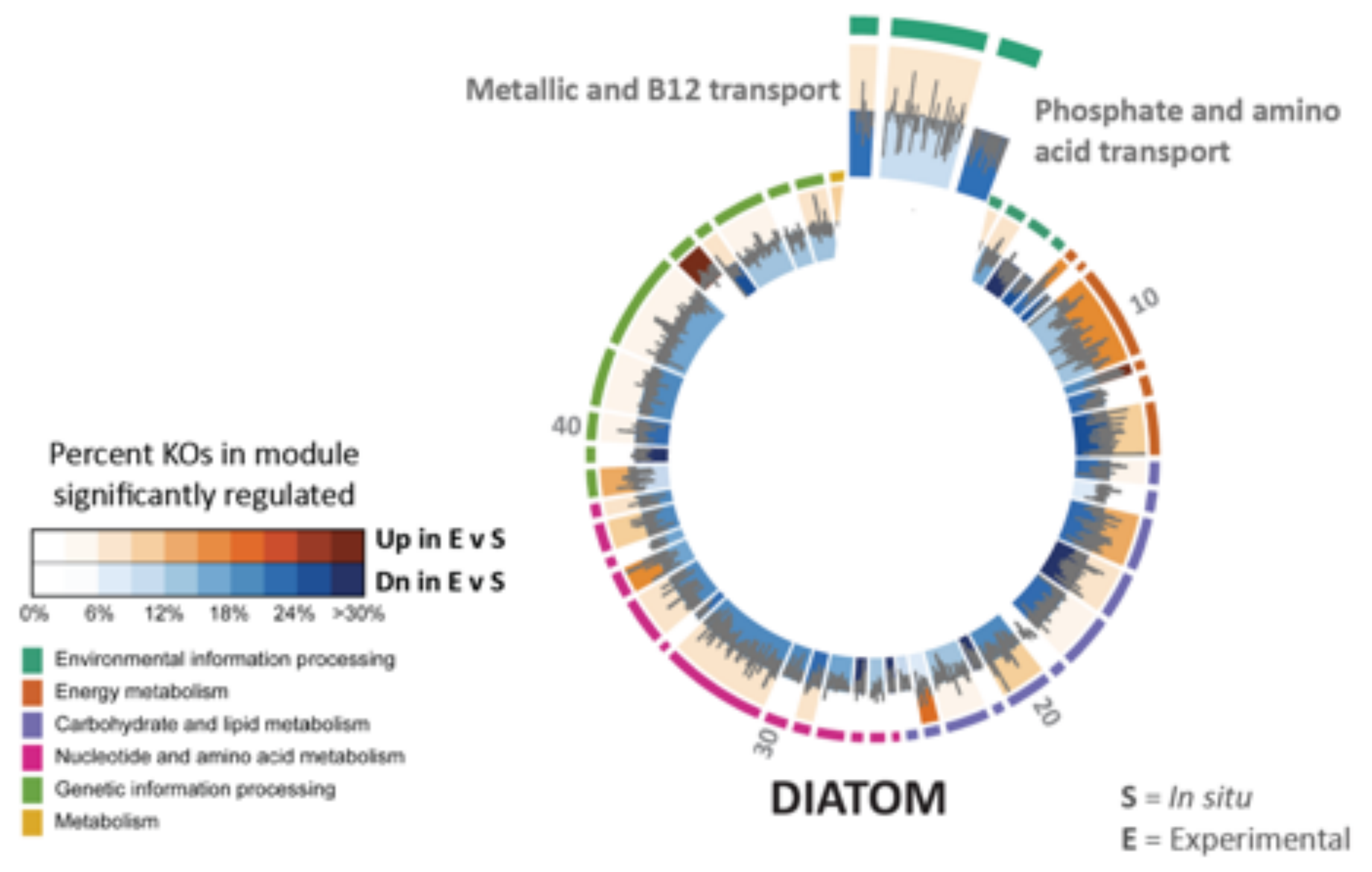
Patterns of metabolism



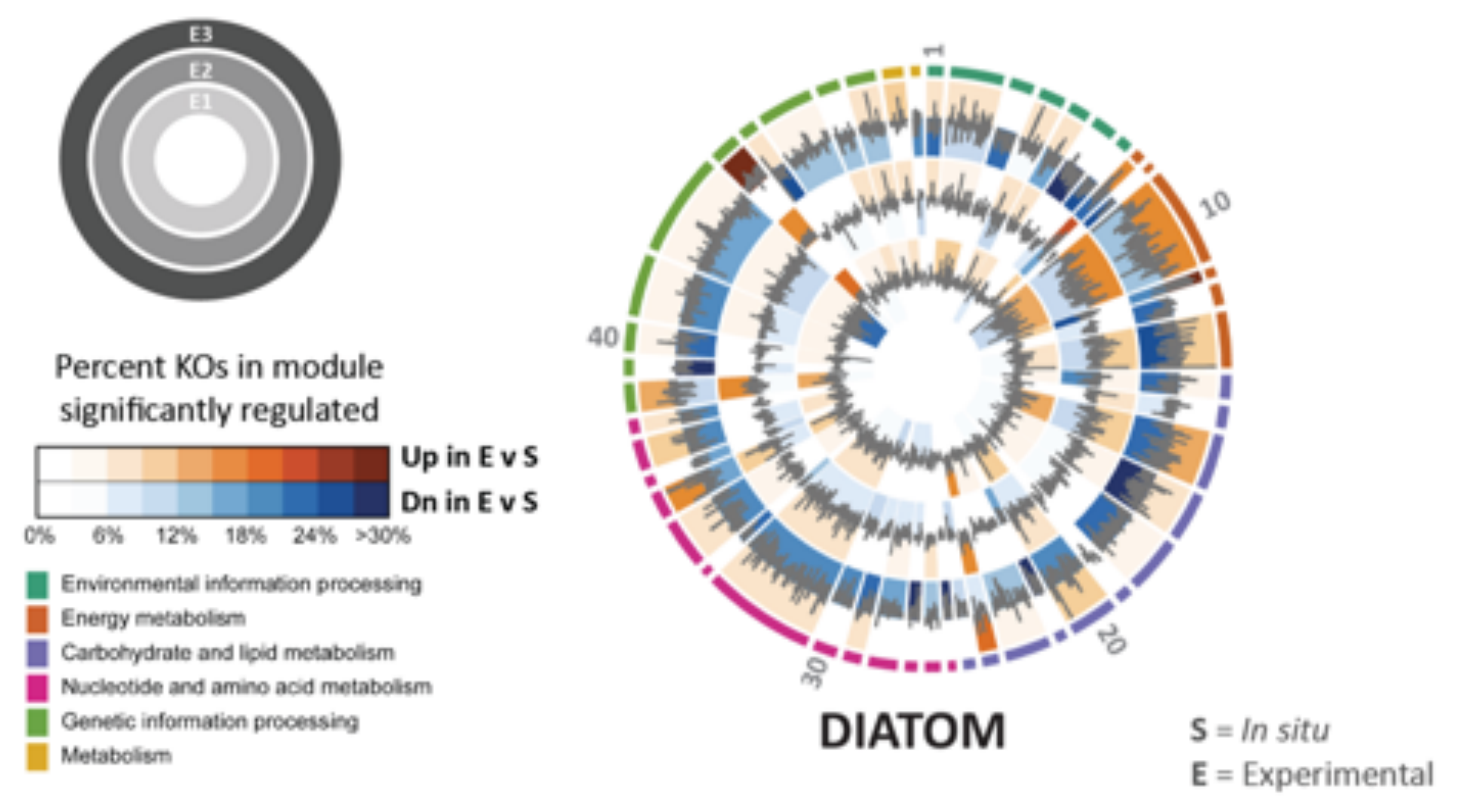
Alexander et al. 2015b *PNAS*

Species are distinct from each other with 95% confidence, and all but dinoflagellates significantly shift in the simulated blooms

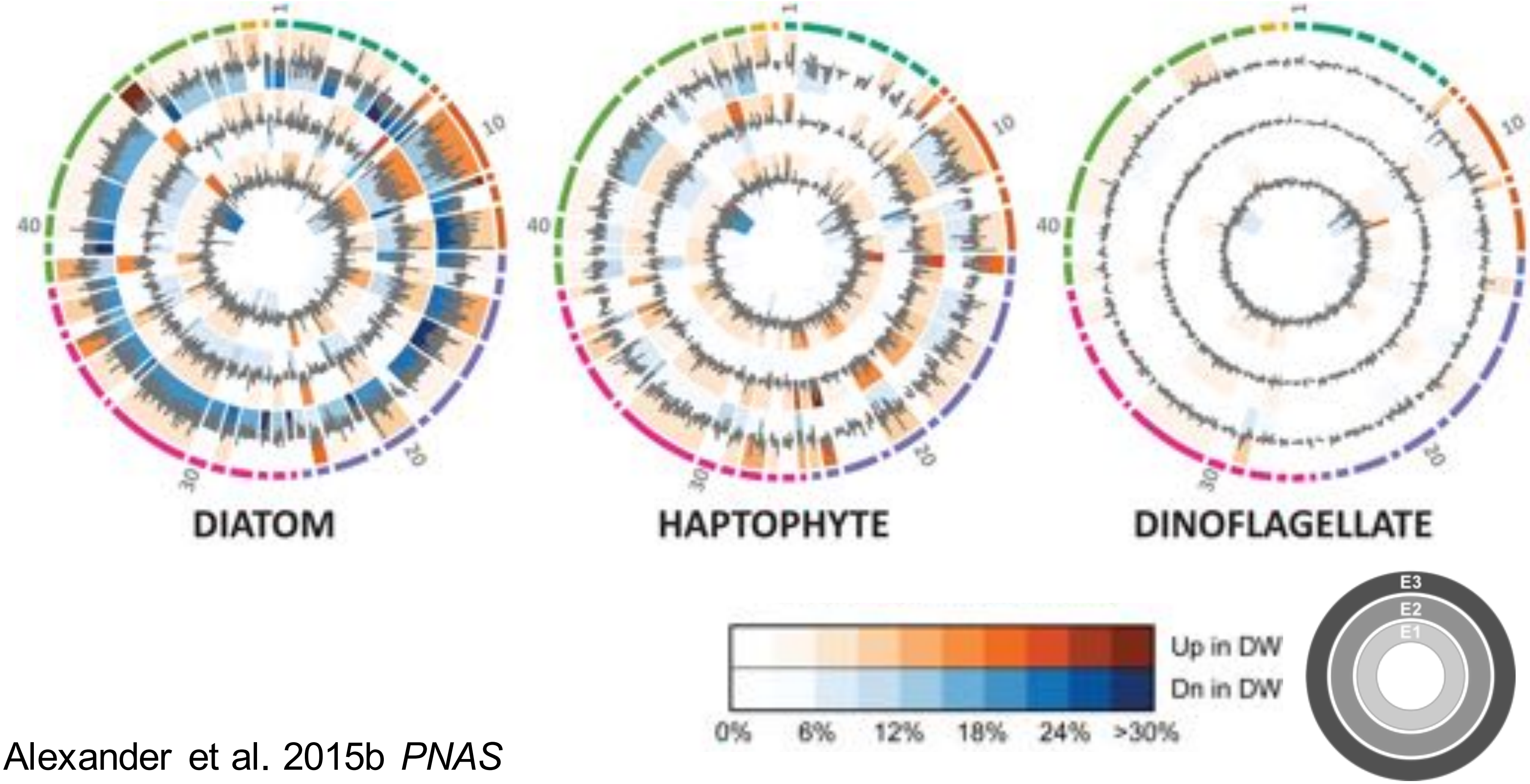
Transcriptional shifts in diatoms



Transcriptional shifts in diatoms



Transcriptional response during simulated blooms

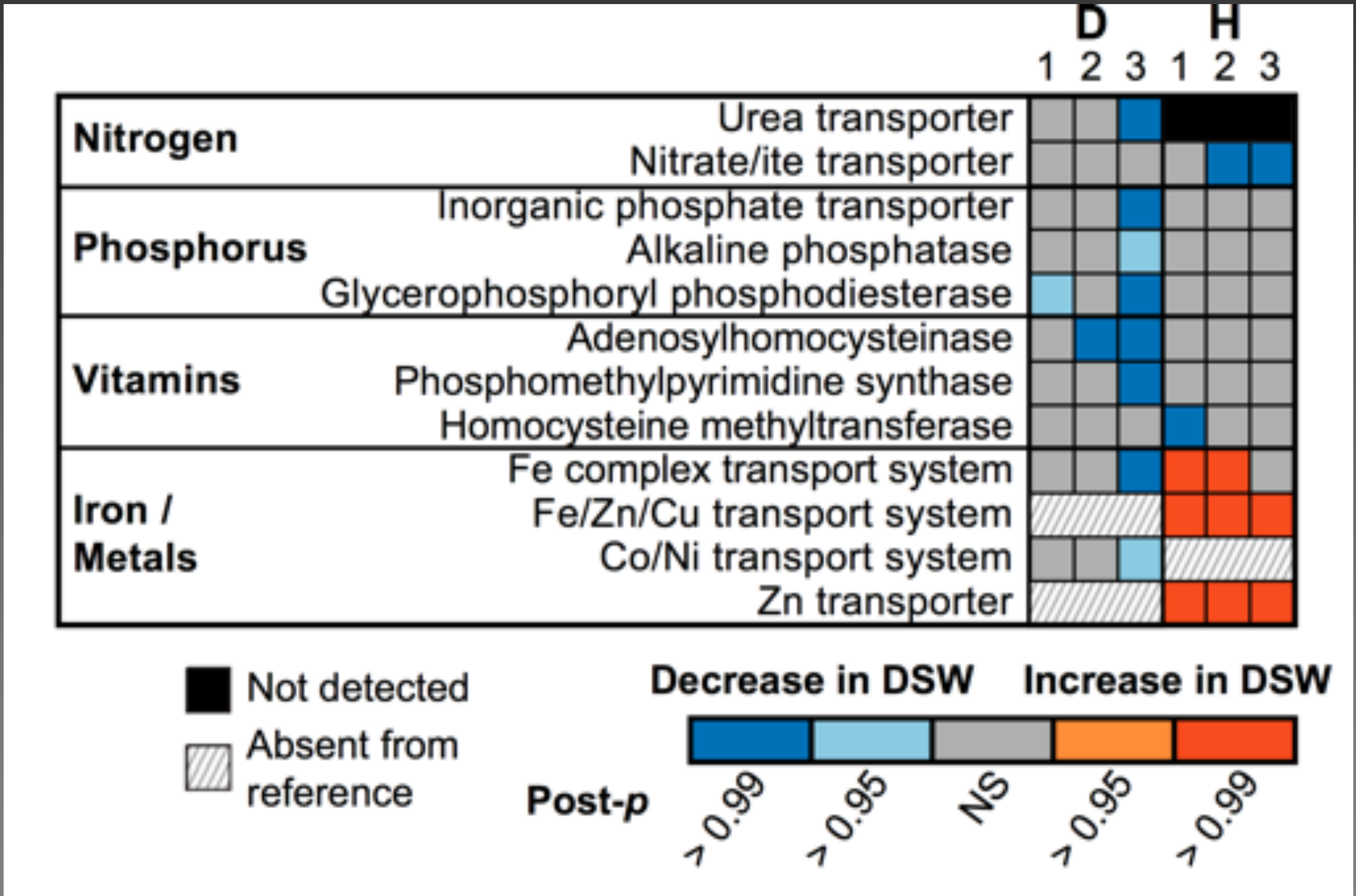


Alexander et al. 2015b *PNAS*

Functional groups have distinct and reproducible transcriptional response during blooms driven by nutrient input

Diatoms appear to be released from nutrient limitation

Fold change of genes related to resource acquisition



Alexander et al. 2015b *PNAS*

Transcriptional responses underscore functional group traits

Dinoflagellates

FEW significant shifts → Alternative trophic modes?

Diatoms

BROAD transcript decreases reflecting release from limitation
TARGETED transcript increase related to growth

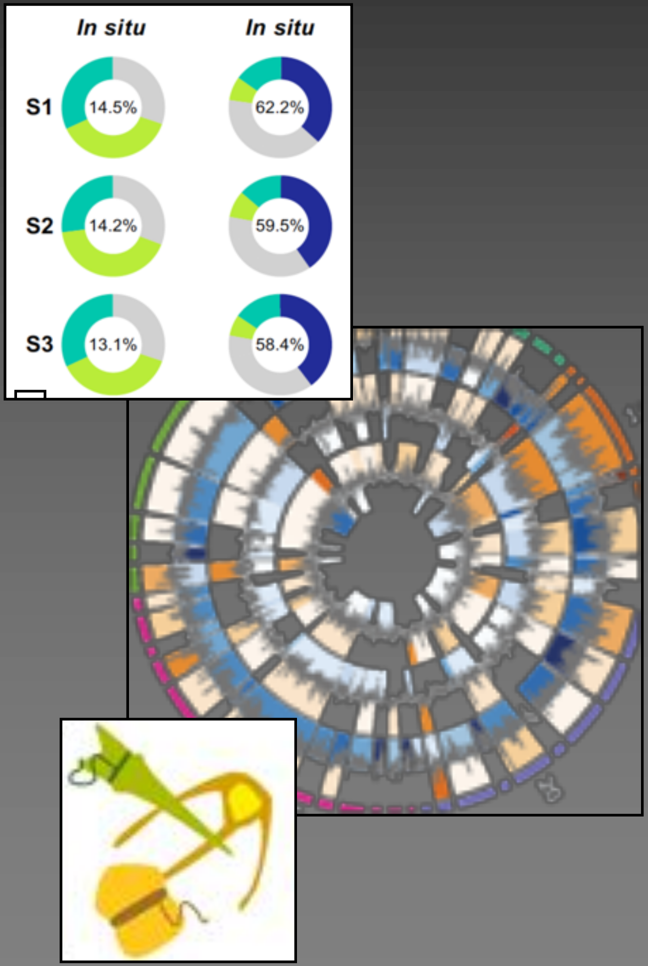
→ "the scavengers"
r-selected

Haptophytes

MUTED increase and decreases → "the survivalists"
K-selected

Summary – Nutrient responses

- **Transcriptomes**: The better the reference database (MMETSP) the better resolution
- **Metatranscriptomes**: Phytoplankton groups appear to be limited by resource availability in the NPSG with remarkable year to year consistency
- **Metatranscriptomes**: Dinoflagellates have little transcriptional response to nutrient input - will need other methods
- Diatoms are particularly responsive nutrient input which likely underpins their dominance in blooms



Core questions

Nutrient responses

- Proof of concept: How do different groups respond to nutrient pulses?

Light responses

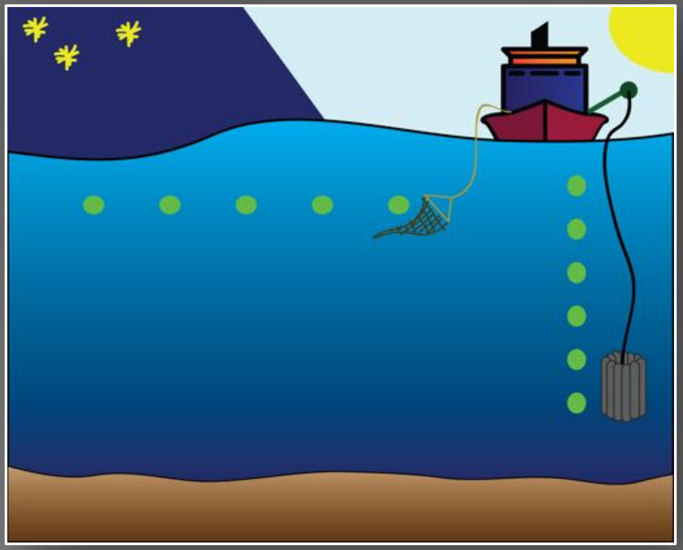
- How does light drive metabolic cascades across phyla?

Interactions

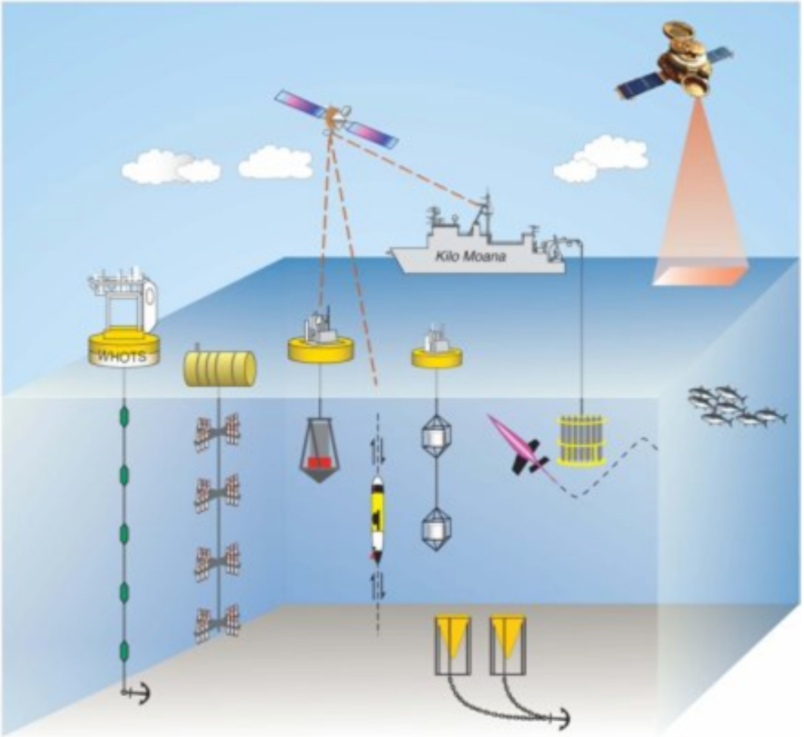
- Are there patterns of coordinated expression that underpin host-symbiont physiology?
- How does light drive interactions between the microbiome and host?

HOE:DYLAN

MAHALO



Hawaii Ocean Experiment: Dynamics of Light (MAHALO)

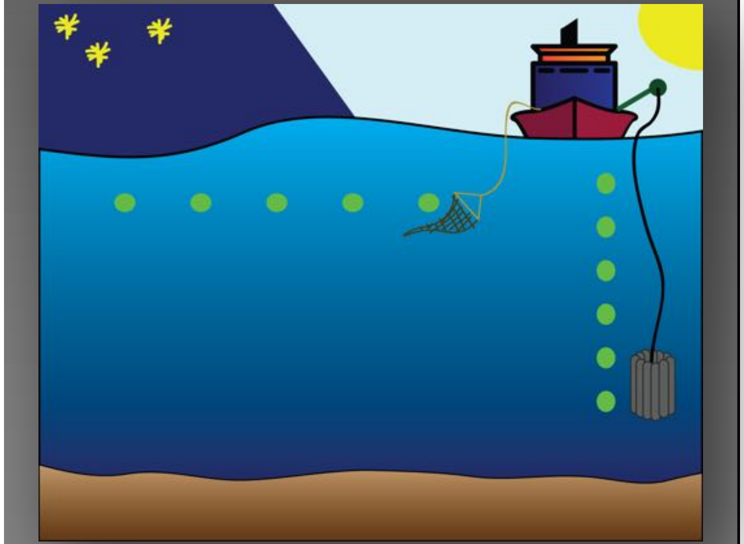


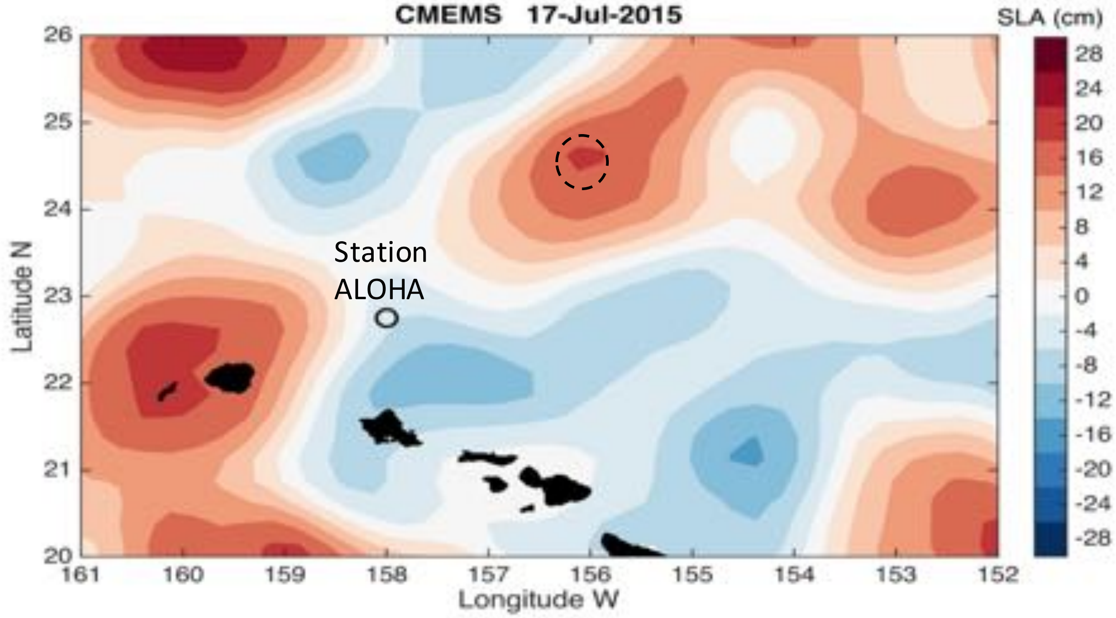
MAHALO: Special thanks to Daniel Muratore and Joshua Weitz for synthesis graphics

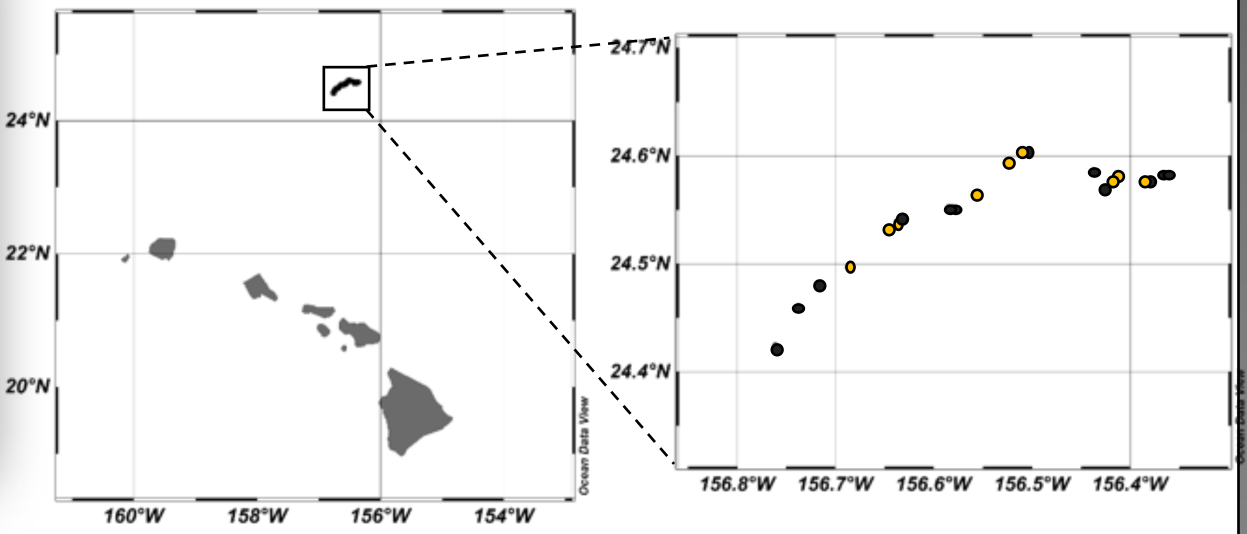
SCOPE
SIMONS COLLABORATION ON OCEAN PROCESSES AND ECOLOGY
Inputs
Microbes
Models
Outputs
SCOPE.SOEST.HAWAII.EDU

Sampling light driven responses in the NPSG

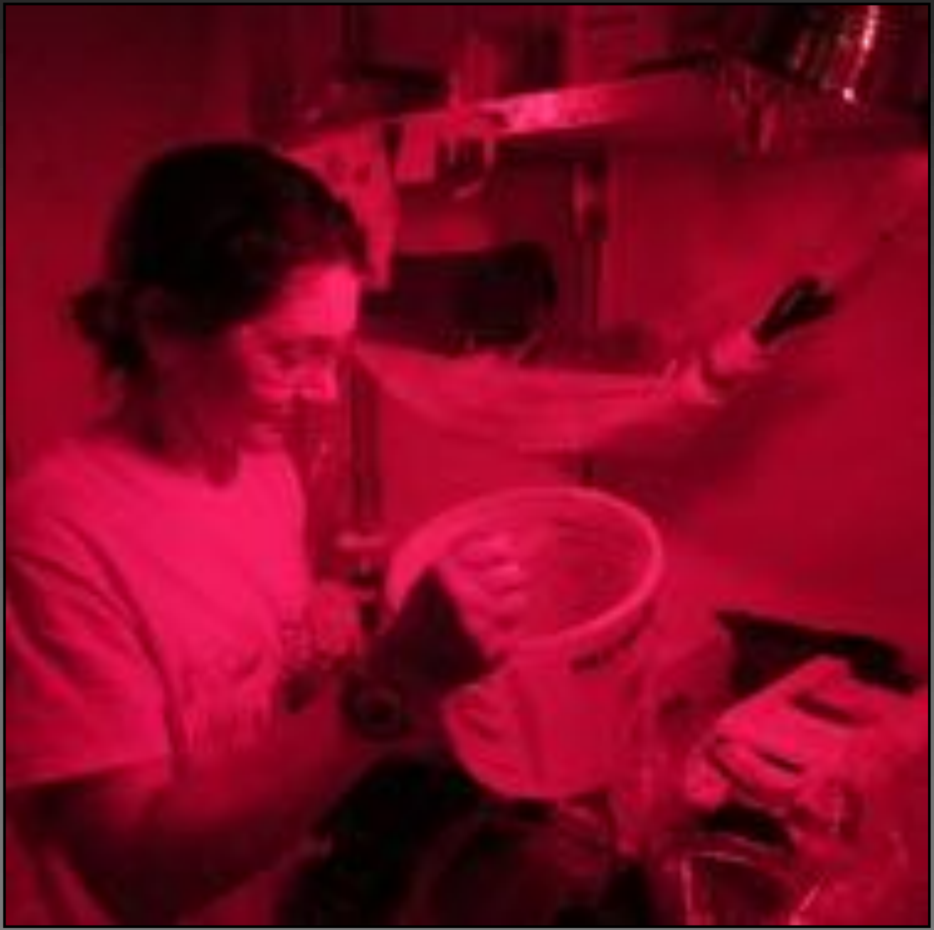
R/V Kilo Moana



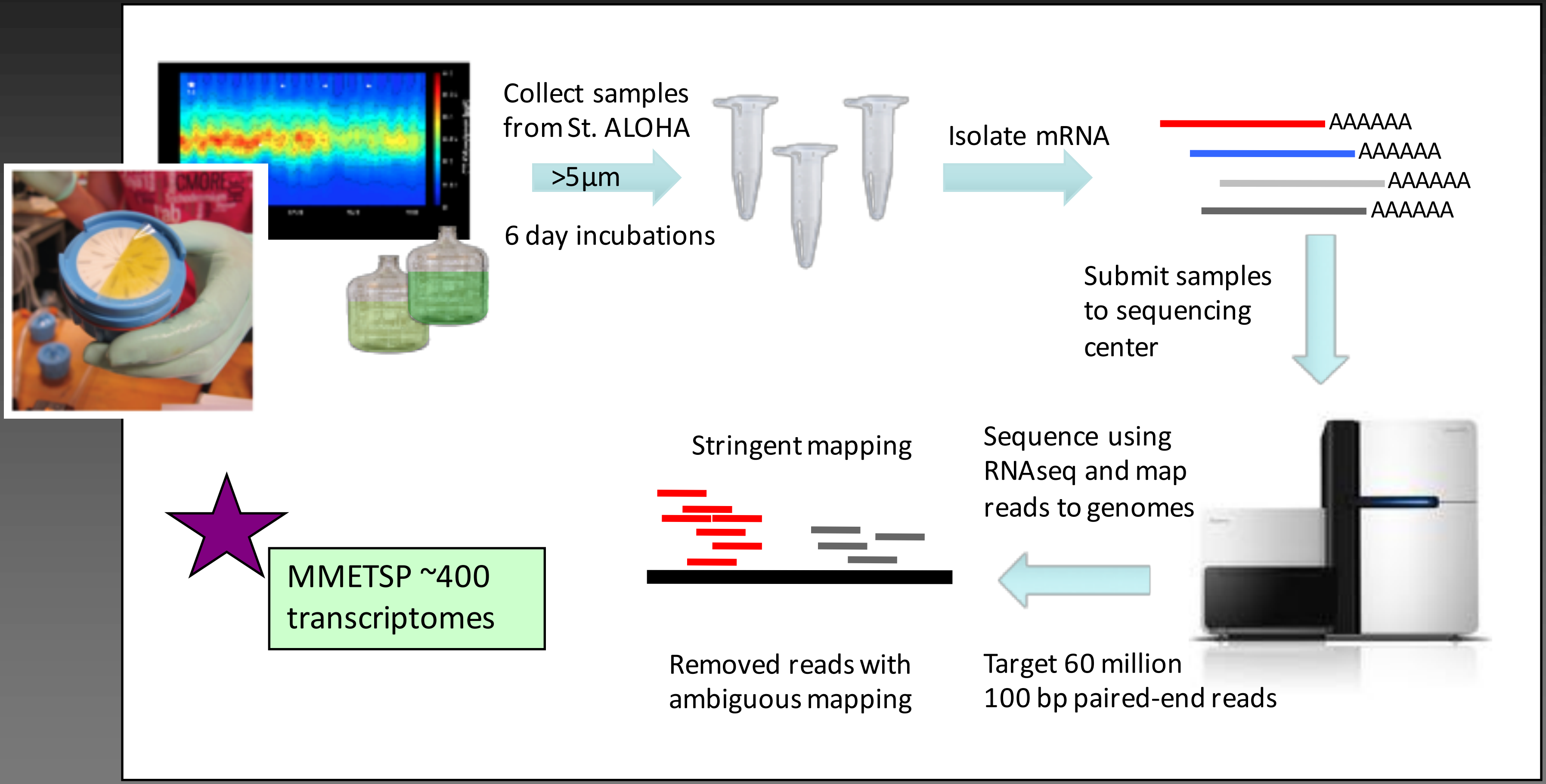




Sampling every 4hrs for 3.5 days



Sampling and pipeline



Collect samples from St. ALOHA

6 day incubations

Isolate mRNA

Submit samples to sequencing center

Sequence using RNAseq and map reads to genomes

Target 60 million 100 bp paired-end reads

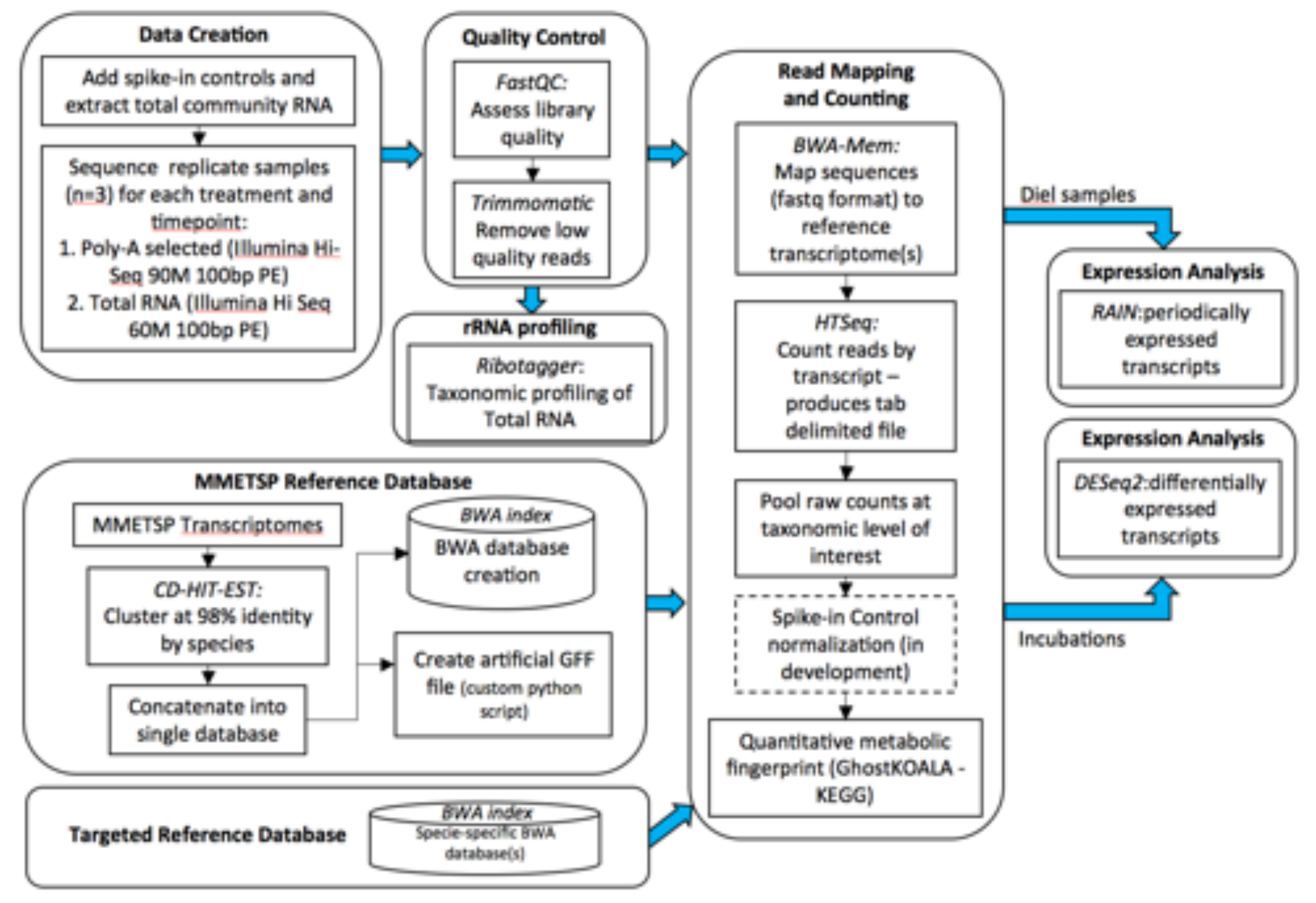
Stringent mapping

Removed reads with ambiguous mapping

★ MMETSP ~400 transcriptomes

★ MMETSP = Marine Microbial Eukaryotic Transcriptome Project

Sequencing and analysis strategy



Data Creation

Add spike-in controls and extract total community RNA

Sequence replicate samples (n=3) for each treatment and timepoint:

1. Poly-A selected (Illumina Hi-Seq 90M 100bp PE)
2. Total RNA (Illumina Hi Seq 60M 100bp PE)

Quality Control

FastQC: Assess library quality

Trimmomatic Remove low quality reads

rRNA profiling

Ribotagger: Taxonomic profiling of Total RNA

MMETSP Reference Database

MMETSP Transcriptomes

CD-HIT-EST: Cluster at 98% identity by species

Concatenate into single database

BWA index BWA database creation

Create artificial GFF file (custom python script)

Targeted Reference Database

BWA index Specie-specific BWA database(s)

Read Mapping and Counting

BWA-Mem: Map sequences (fastq format) to reference transcriptome(s)

HTSeq: Count reads by transcript – produces tab delimited file

Pool raw counts at taxonomic level of interest

Spike-in Control normalization (in development)

Quantitative metabolic fingerprint (GhostKOALA - KEGG)

Diel samples

Expression Analysis

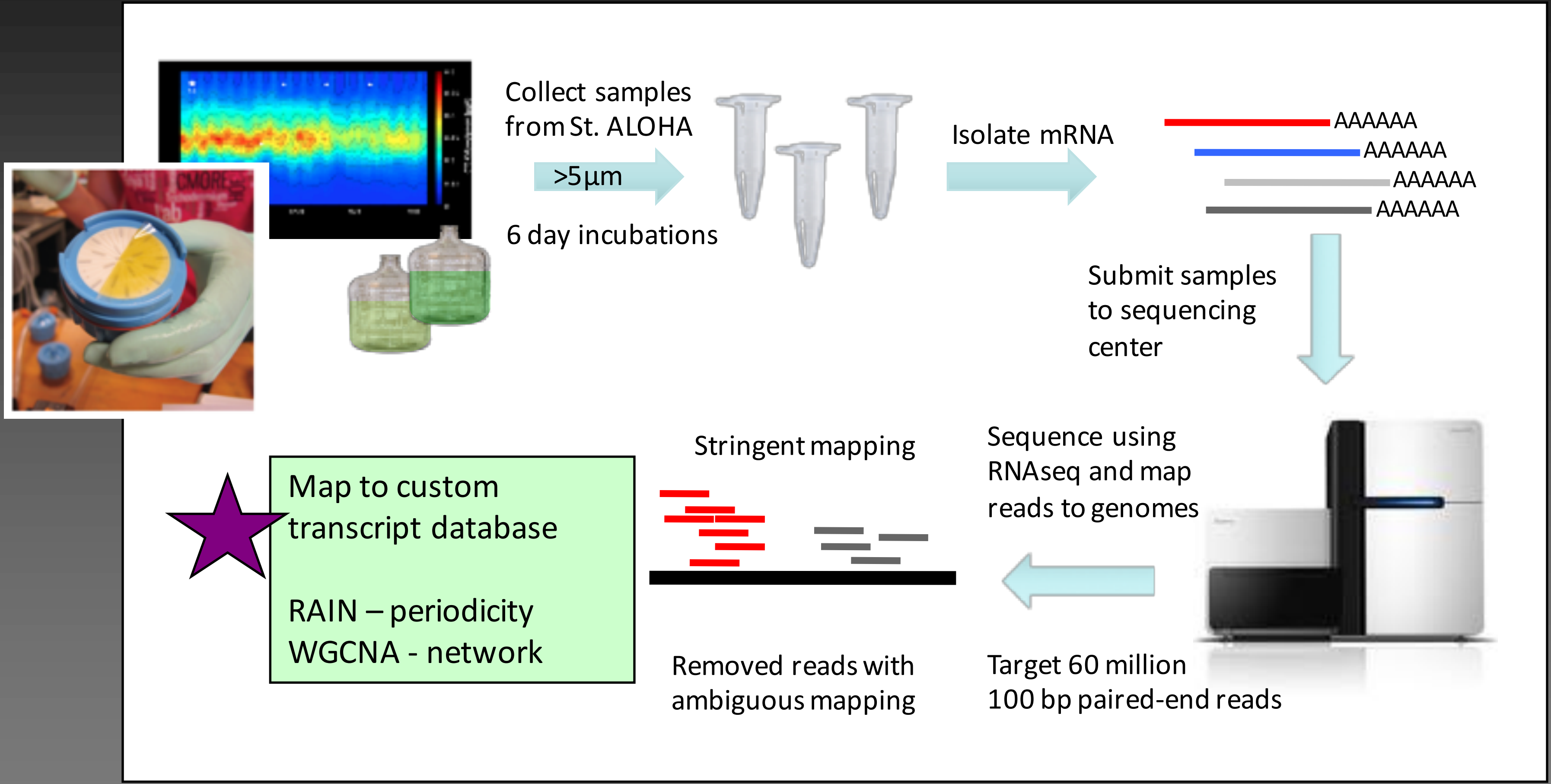
RAIN:periodically expressed transcripts

Incubations

Expression Analysis

DESeq2:differentially expressed transcripts

Sampling and pipeline



Collect samples from St. ALOHA

>5µm

6 day incubations

Isolate mRNA

Submit samples to sequencing center

Sequence using RNAseq and map reads to genomes

Target 60 million 100 bp paired-end reads

Stringent mapping

Removed reads with ambiguous mapping

Map to custom transcript database

RAIN – periodicity
WGCNA - network

https://omictools.com/rhythmicity-analysis-incorporating-nonparametric-methods-tool

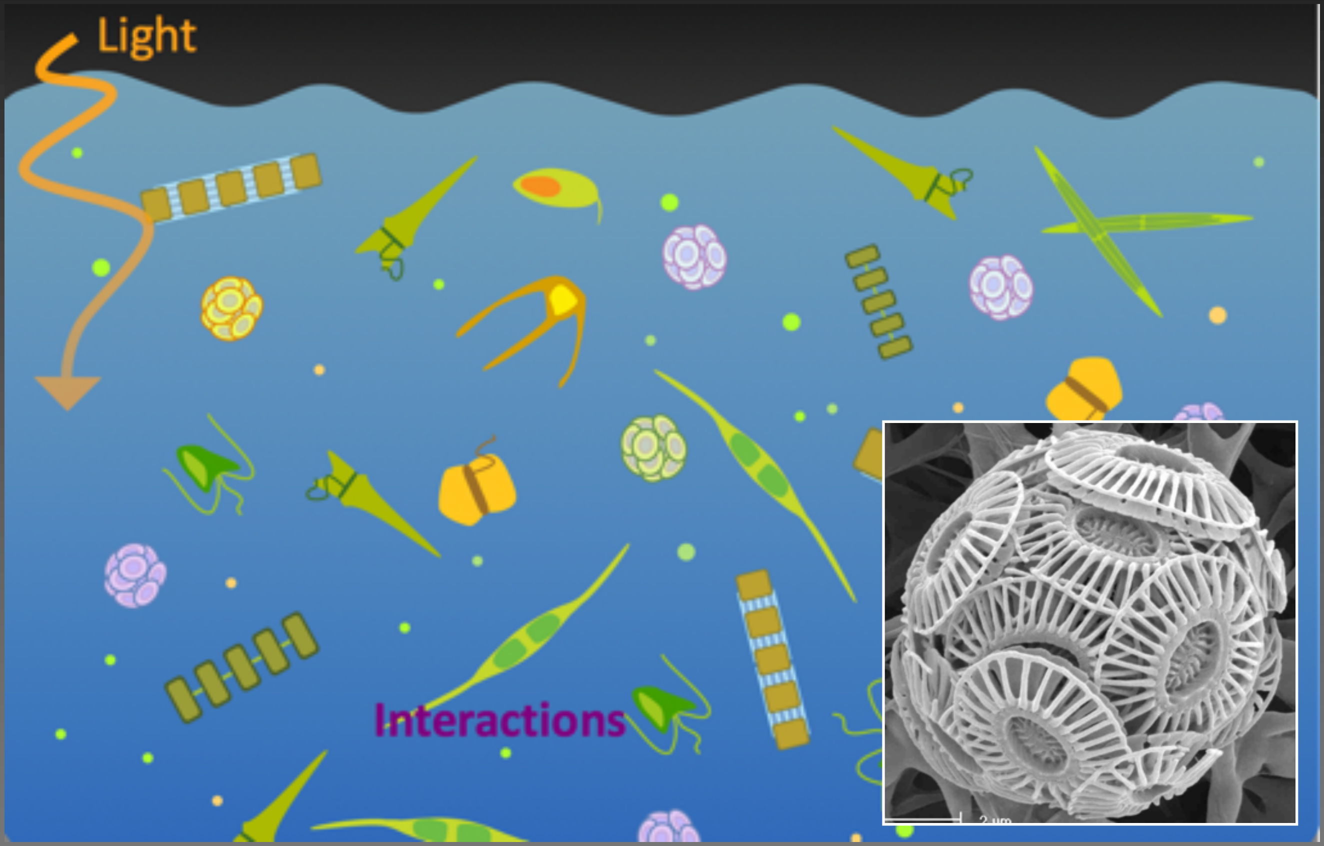
https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559

NMDS plot shows phased peaks in transcription

Different photosynthetic phyla have different patterns

Some overlap of functions between phyla

Diving beyond phyla to the species level …



Tracking strain specific responses in the calcifying haptophyte:
Emiliania huxleyi (Ehux)

Emiliania huxleyi: a cosmopolitan, globally significant species

- Calcification - critical role in global carbon cycle and strongly linked to climate driven ocean acidification
- Source of paleoproxies for climate reconstructions
- Form dense blooms, drivers largely unknown
- First marine phytoplankton to have multiple strains sequenced, identifying pan genome

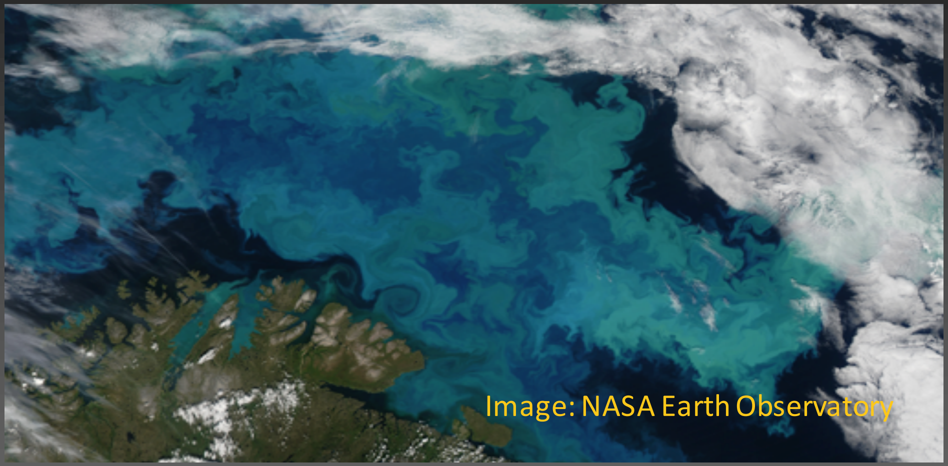
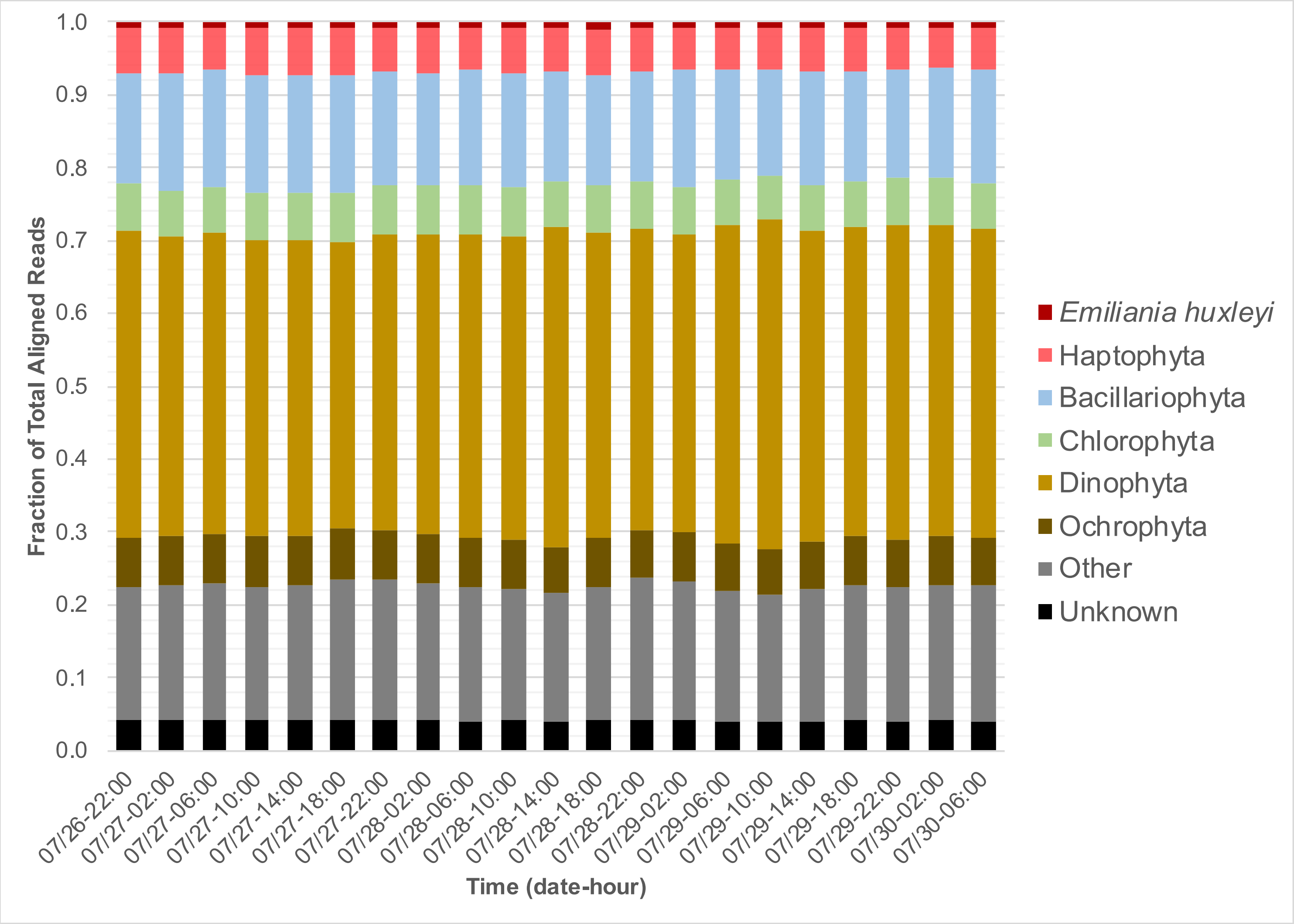
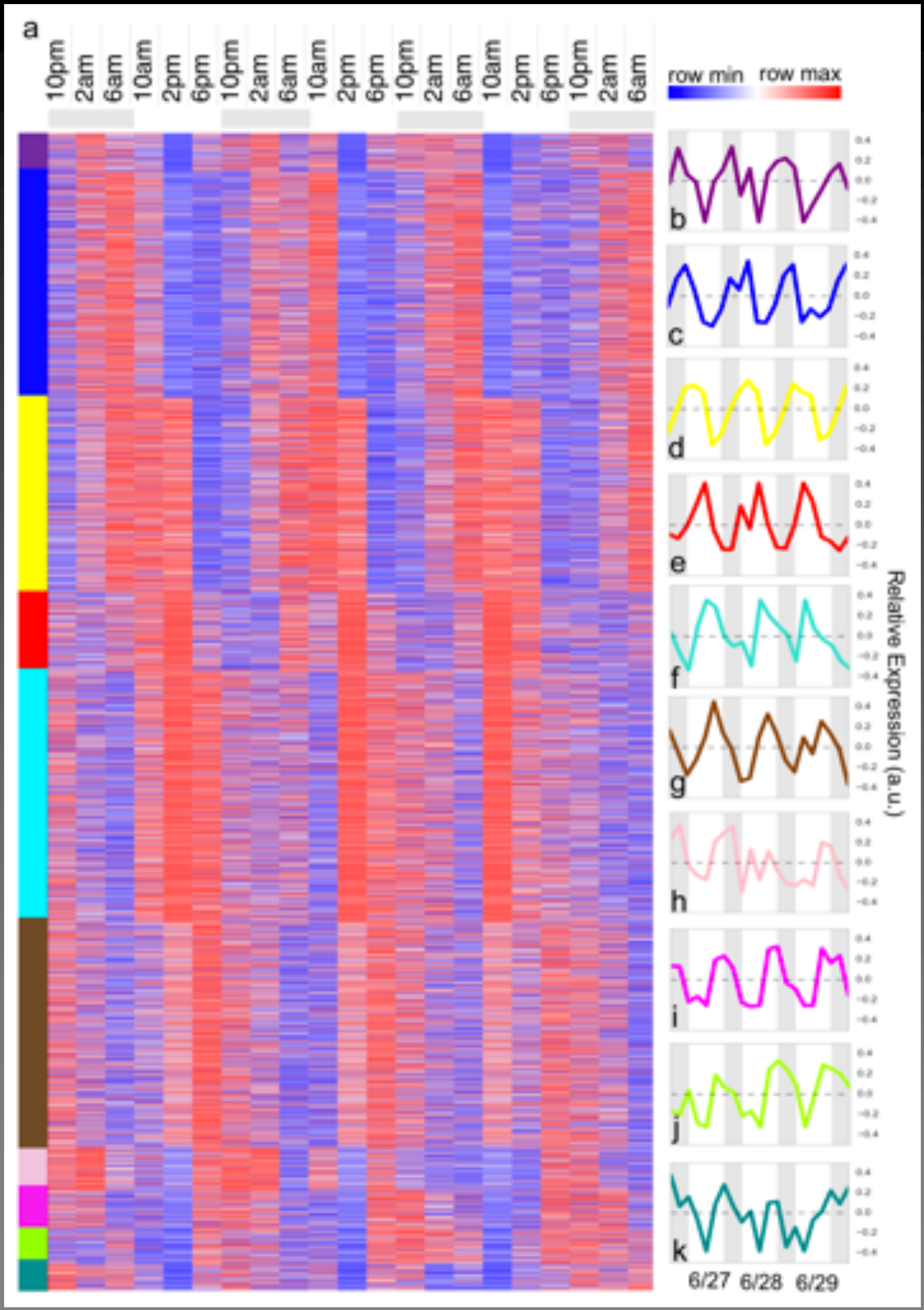


Image: NASA Earth Observatory

Data richness by phyla – *E. huxleyi* dominates haptophyte signal



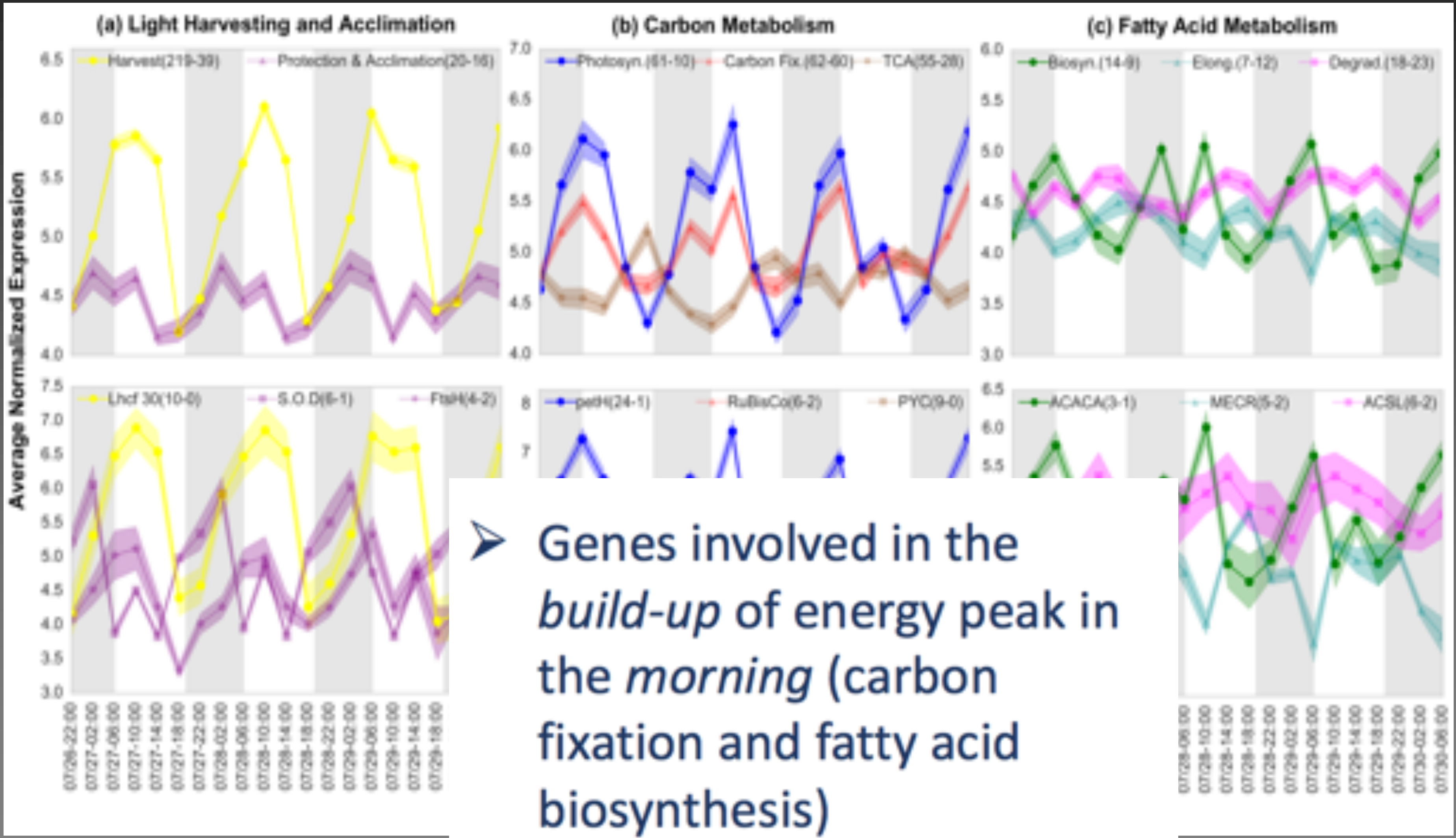
Hernandez et al. (in prep)



- 25% of total genes exhibit diel pattern
- Most peak in morning and evening

Hernandez et al. (in prep)

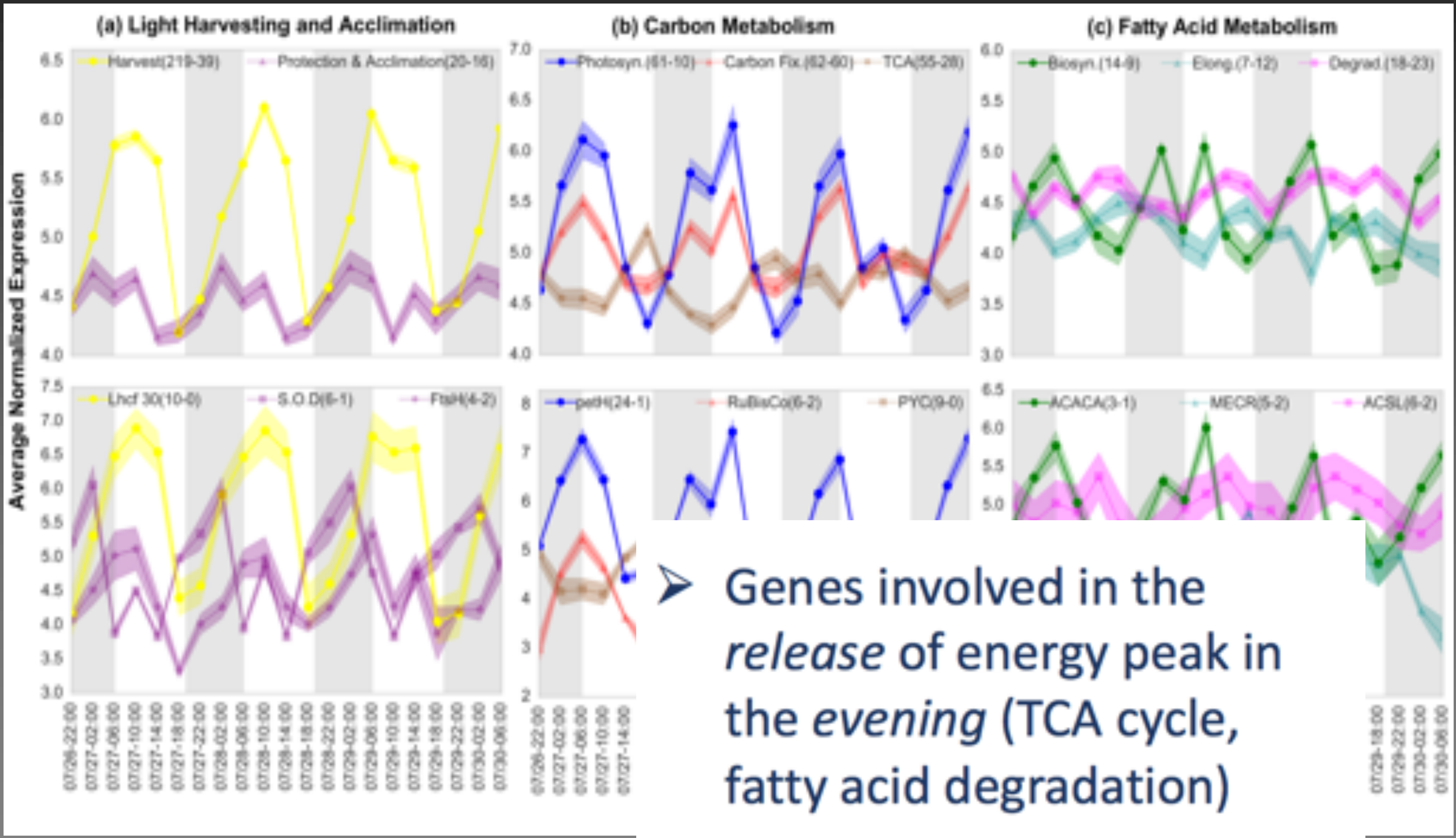
Diel partitioning of metabolism



- Genes involved in the *build-up* of energy peak in the *morning* (carbon fixation and fatty acid biosynthesis)

Hernandez et al. (in prep)

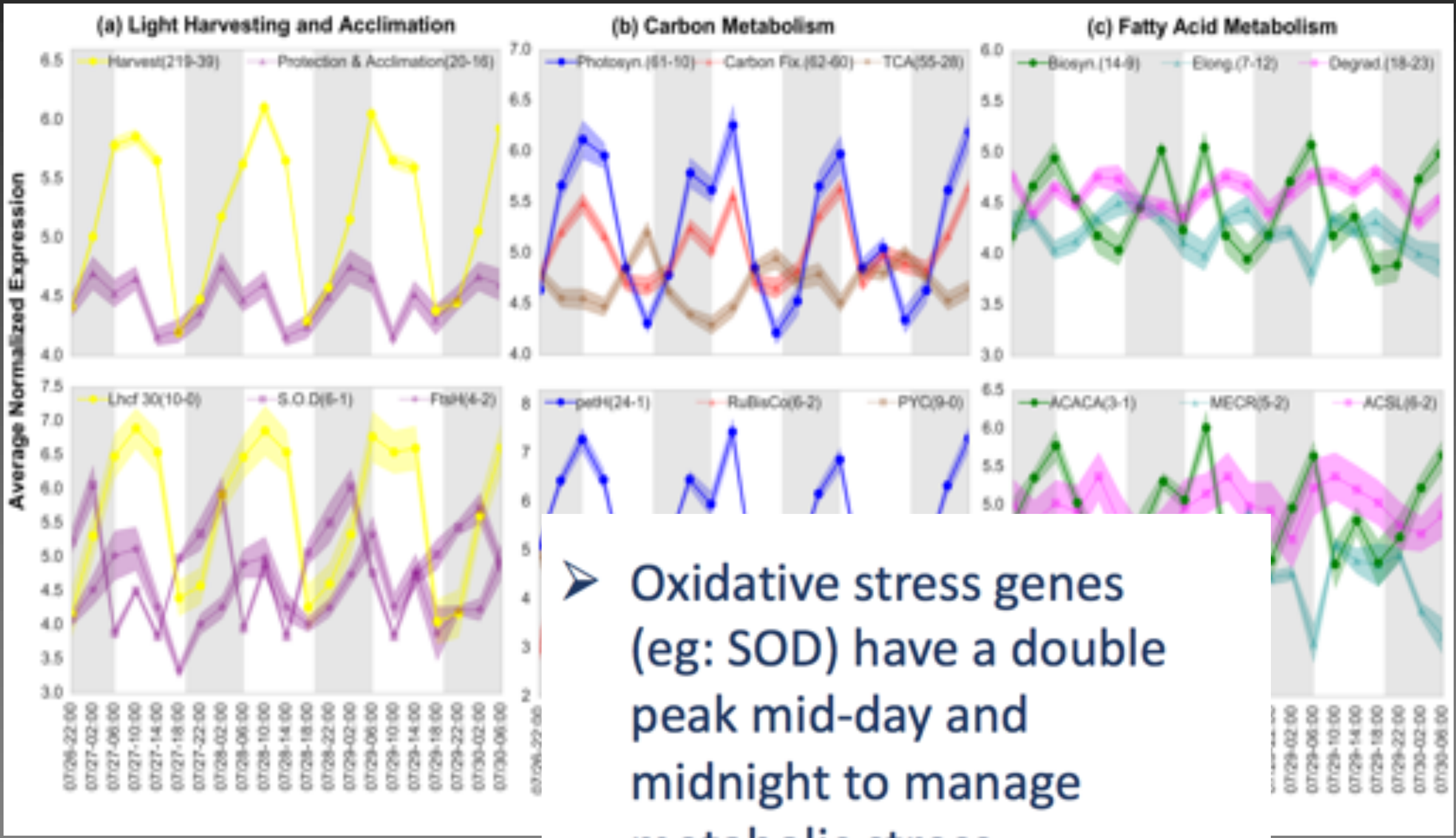
Diel partitioning of metabolism



- ➢ Genes involved in the *release* of energy peak in the *evening* (TCA cycle, fatty acid degradation)

Hernandez et al. (in prep)

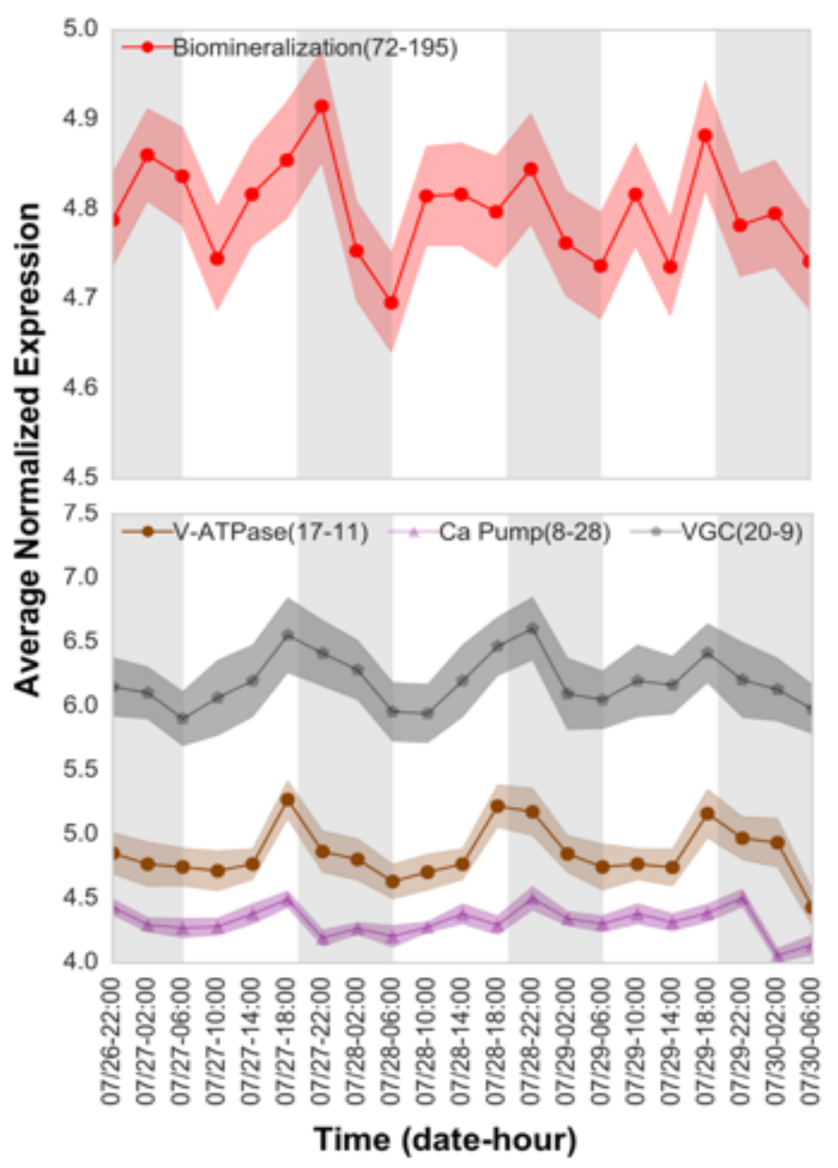
Diel partitioning of metabolism



- Oxidative stress genes (eg: SOD) have a double peak mid-day and midnight to manage metabolic stress

Hernandez et al. (in prep)

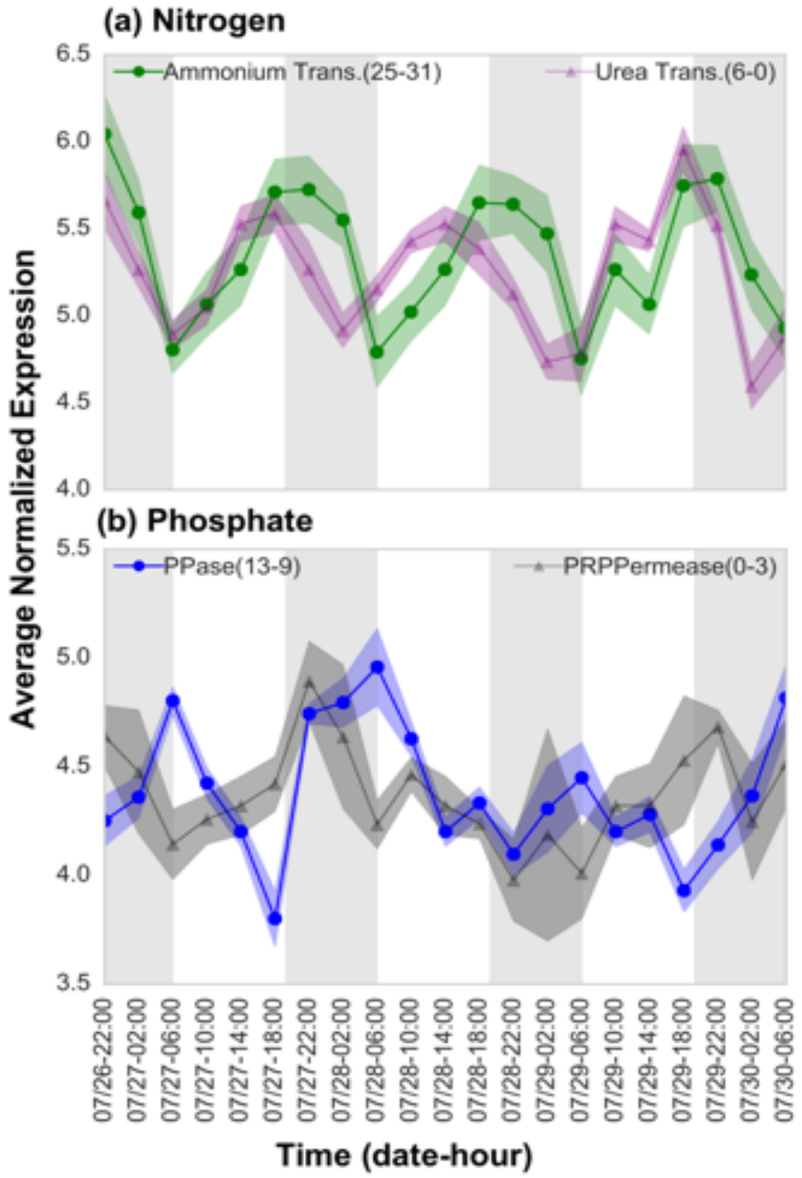
Patterns of putative biomineralization pathways



- Little evidence of phased biomineralization (calcification).
- Diel patterns in some putative biomineralization genes, but these pathways are poorly understood.
- Caution! Sampling time matters in comparing between samples/stations.

Hernandez et al. (in prep)

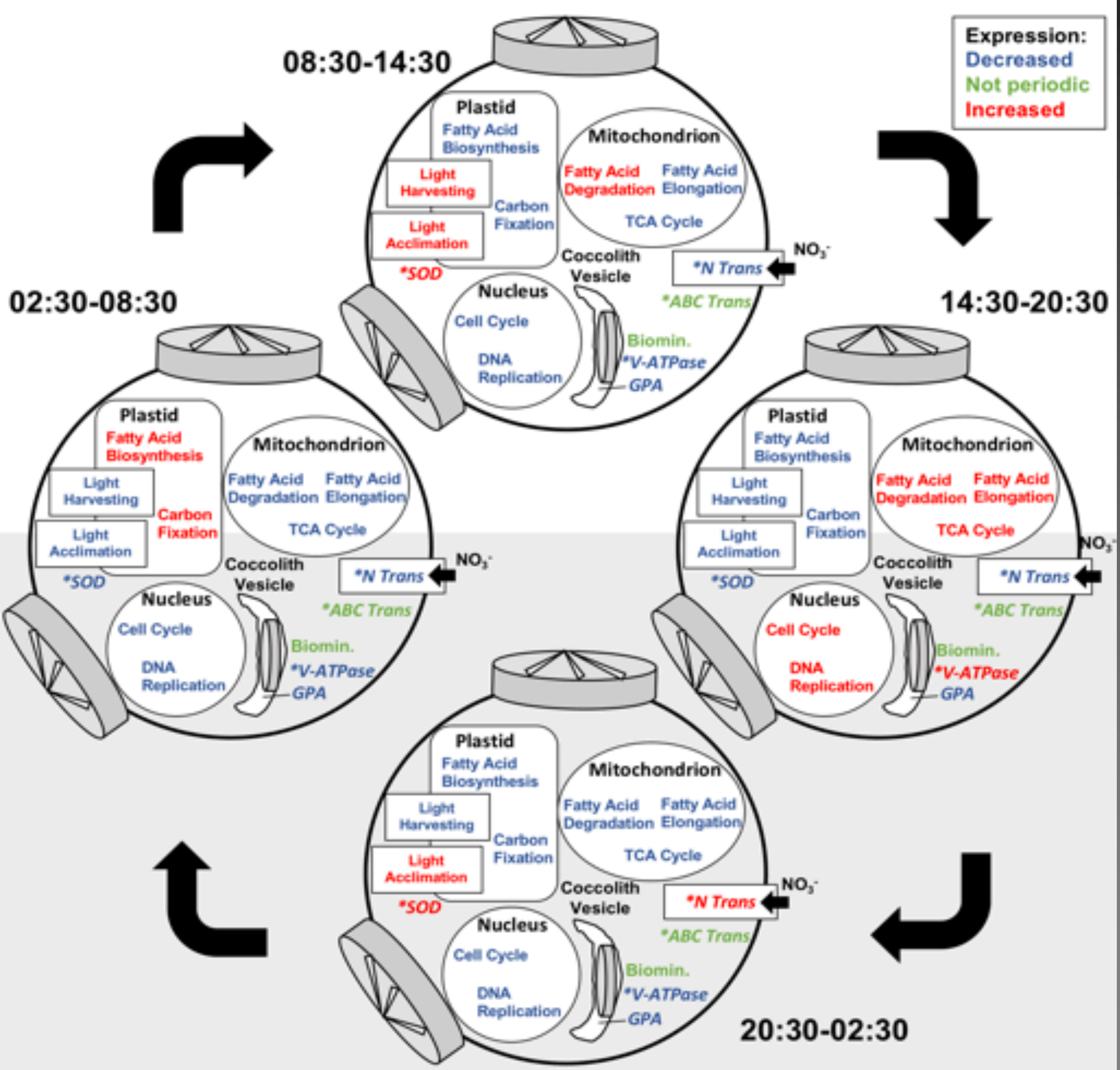
Patterns of resource acquisition



- Strong diel patterns in N-related transporters, fewer diel signals in P-related transport and metabolism
- Caution! Sampling time matters in comparing between samples/stations.

Hernandez et al. (in prep)

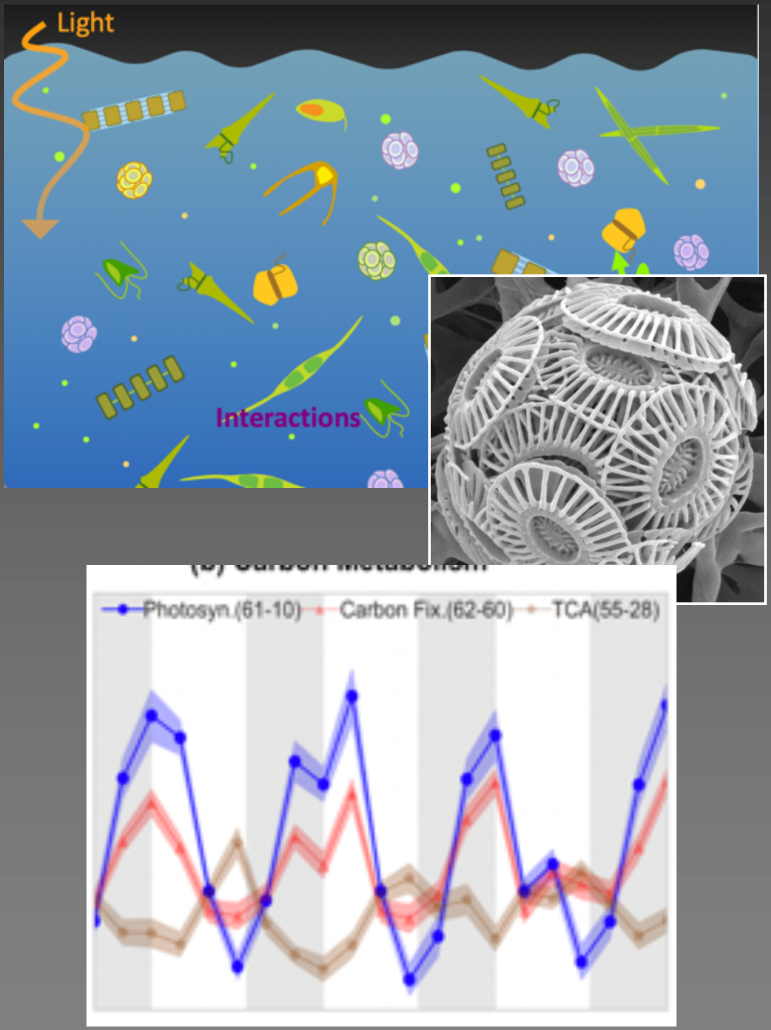
Light drives a reproducible cascade of metabolism



Summary – Light responses

How does light drive metabolic cascades across phyla?

- **Metatranscriptomes**: Light drives phyla specific metabolic cascades which underpin carbon cycling in the NPSG
- **Metatranscriptomes**: Functional responses related to photosynthesis are shared between phyla
- **Metatranscriptomes**: Species-specific resolution of *E. huxleyi* tracks a day in the life of key primary producers in the NPSG



Core questions

Nutrient responses

- Proof of concept: How do different groups respond to nutrient pulses?

Light responses

- How does light drive metabolic cascades across phyla?

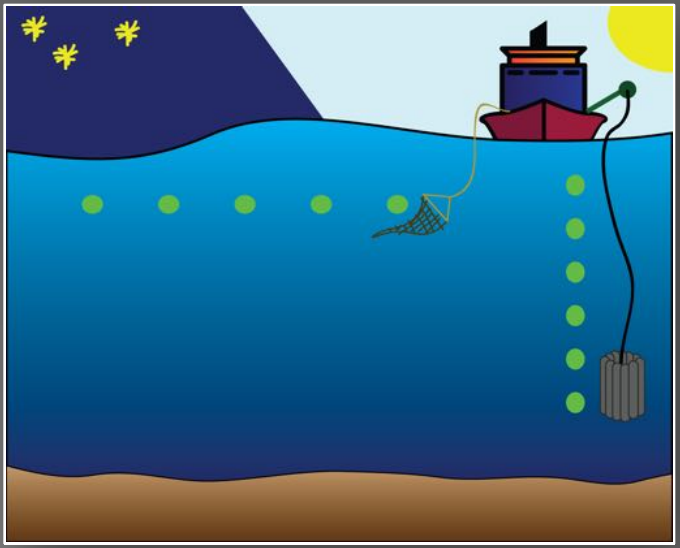
Interactions

- Are there patterns of coordinated expression that underpin host-symbiont physiology?

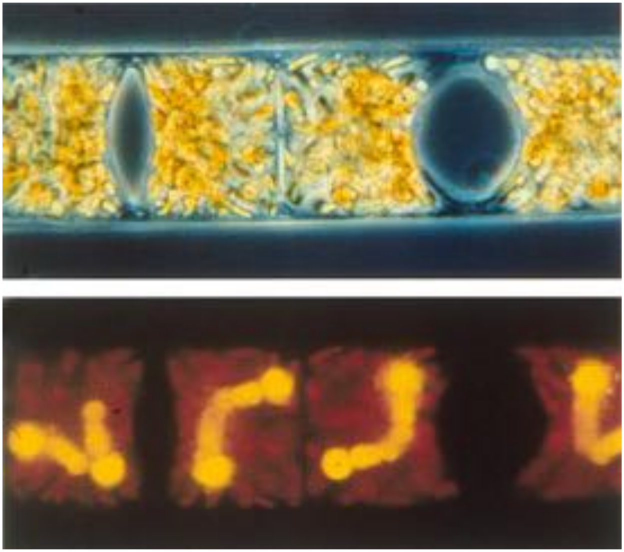
- How does light drive interactions between the microbiome and host?

HOE:DYLAN

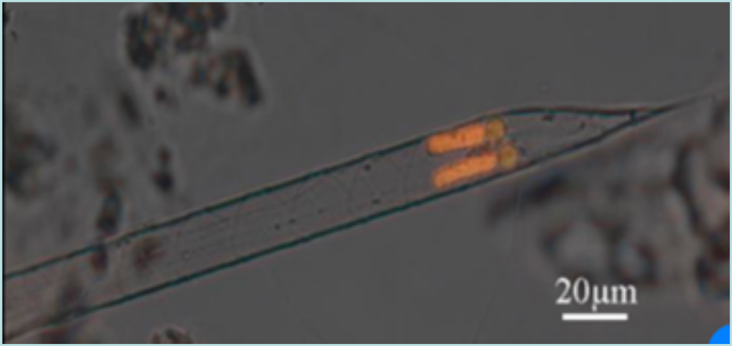
MAHALO



Using light to study interactions



- Some of the most important microbial eukaryotes live in symbioses or associations.
- Diatom diazotroph associations (DDAs) drive summer export of carbon to the deep sea, sequestering CO_2.
- Diazotroph supplies fixed N_2 to host.
- There are no long term reference cultures, resulting in limited knowledge about the activities of these symbioses in the field.



Are there patterns of coordinated expression that underpin host - symbiont physiology?

3x20L from 15m
(21 time points, every 4 h, 3.5 days)

Filter
(>5µm)

Extract RNA

Selected

Map to reference database of MMETSP diatoms
(Including *Rhizosolenia*)

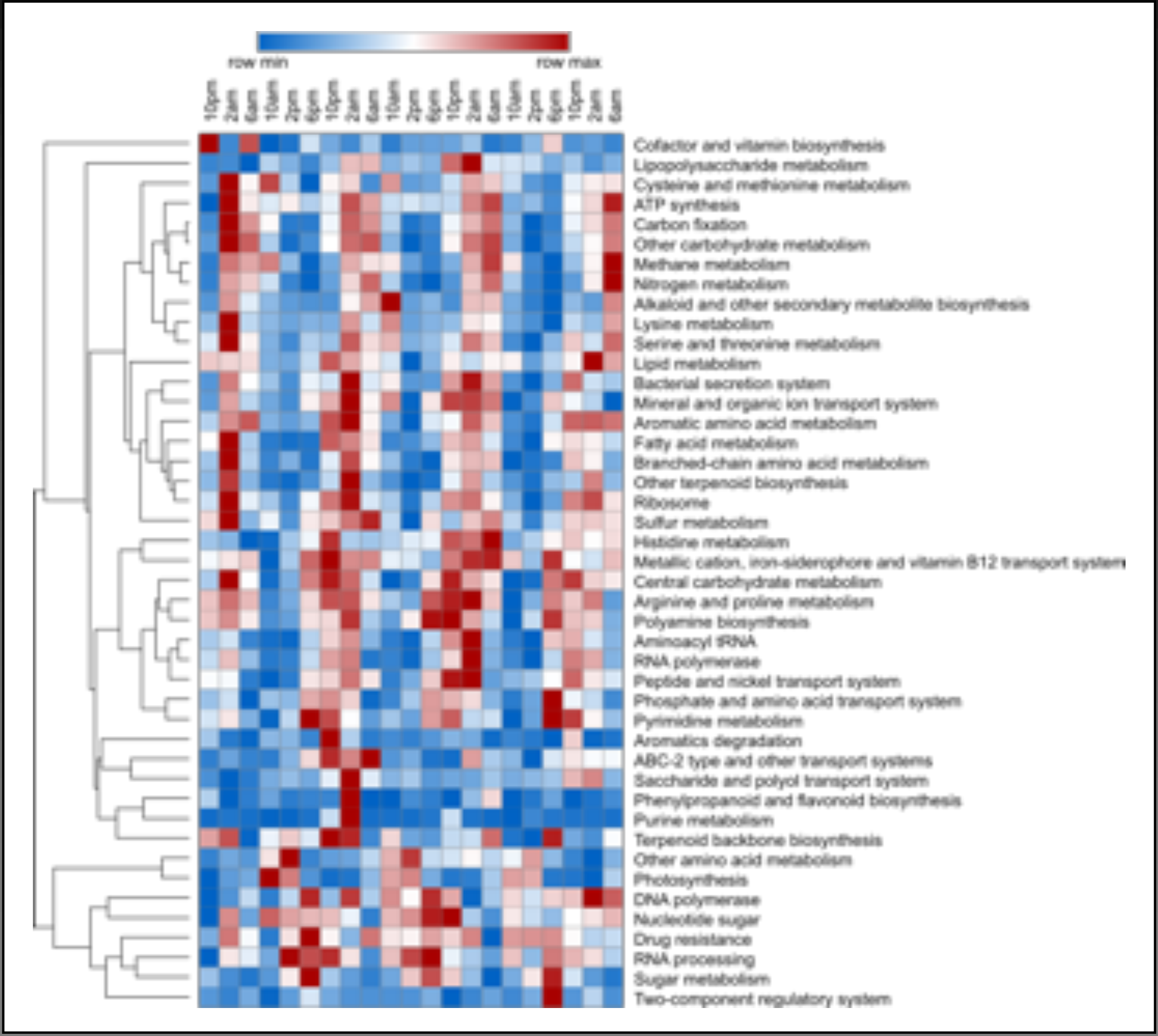
Unselected

Map to reference database of diazotrophs
(Including *Richelia*)

- Each sample generates between 4-6 Gb of sequencing data
- Current data set :
  - ~ 19 billion reads
  - ~ 2 terabytes of data

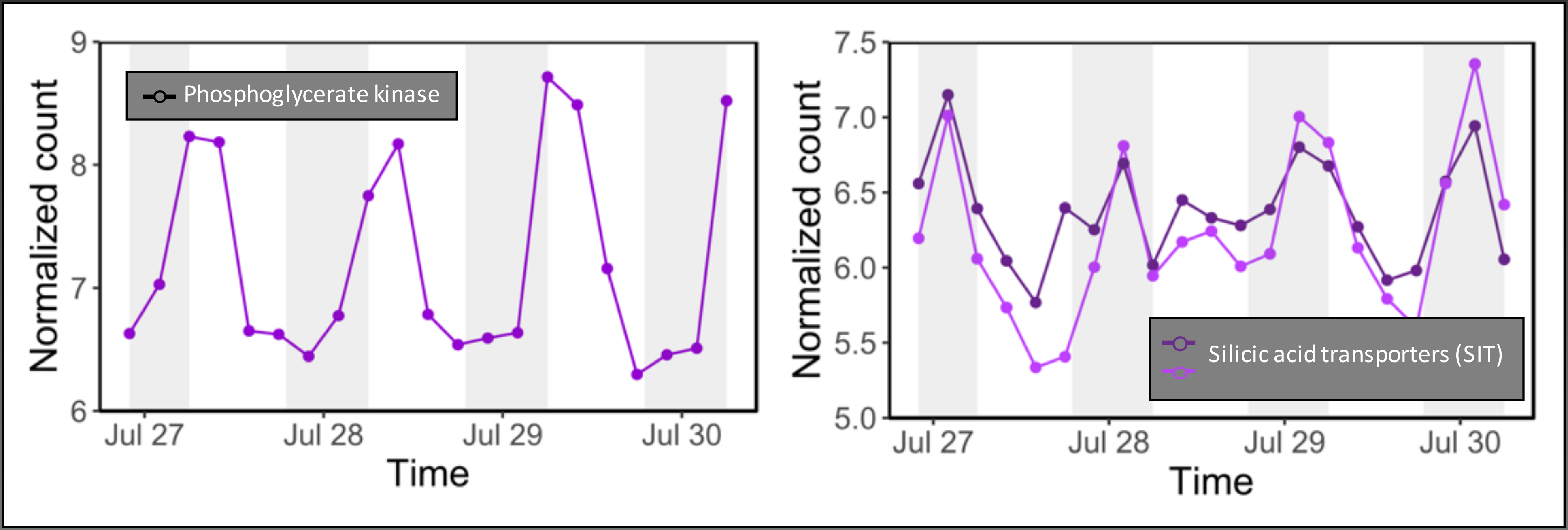
RAIN and WGCNA

Diel patterns in *Rhizosolenia host*



Harke et al. (2018) *ISMEJ*

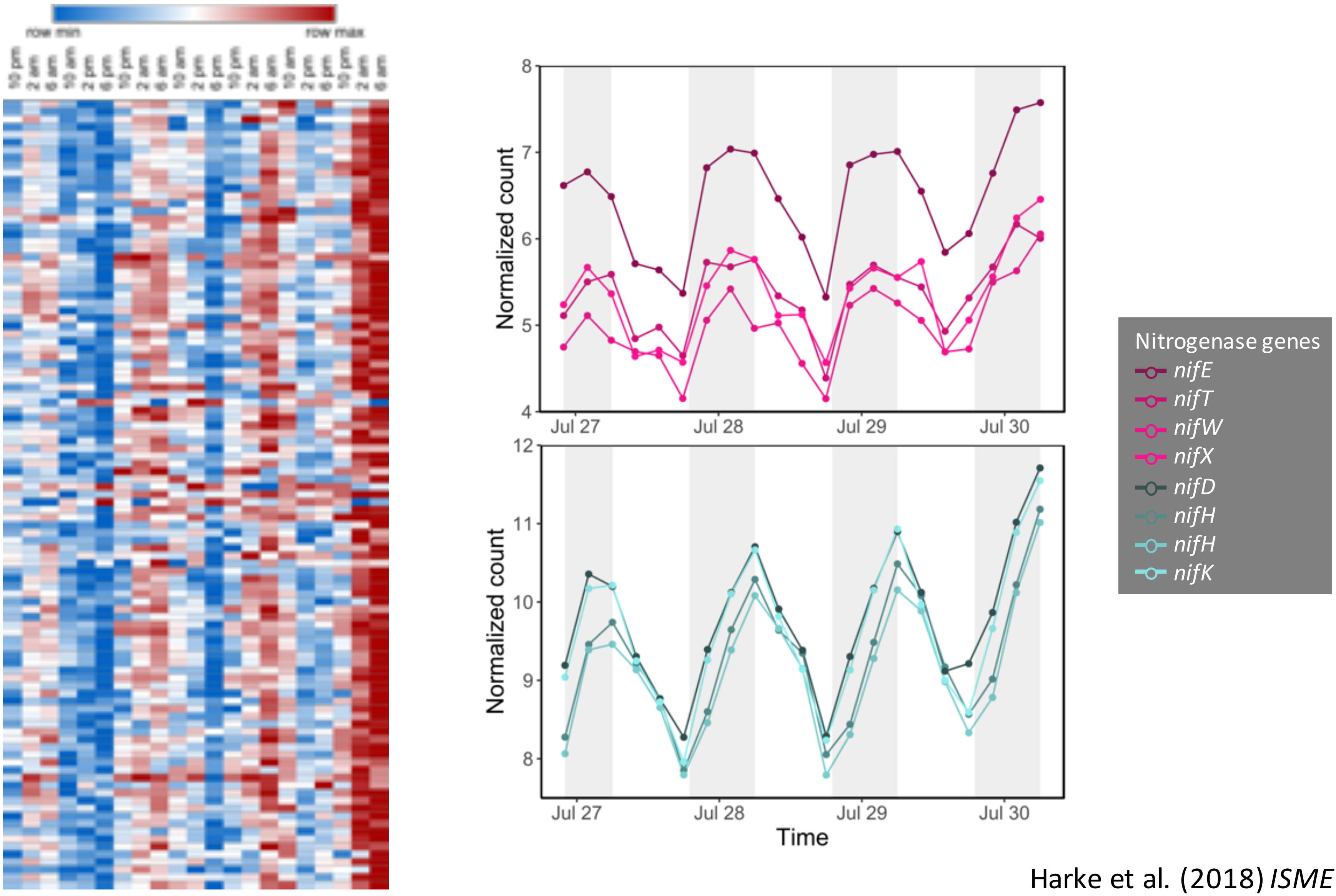
Significantly diel genes in *Rhizosolenia* host



Harke et al. (2018) *ISMEJ*

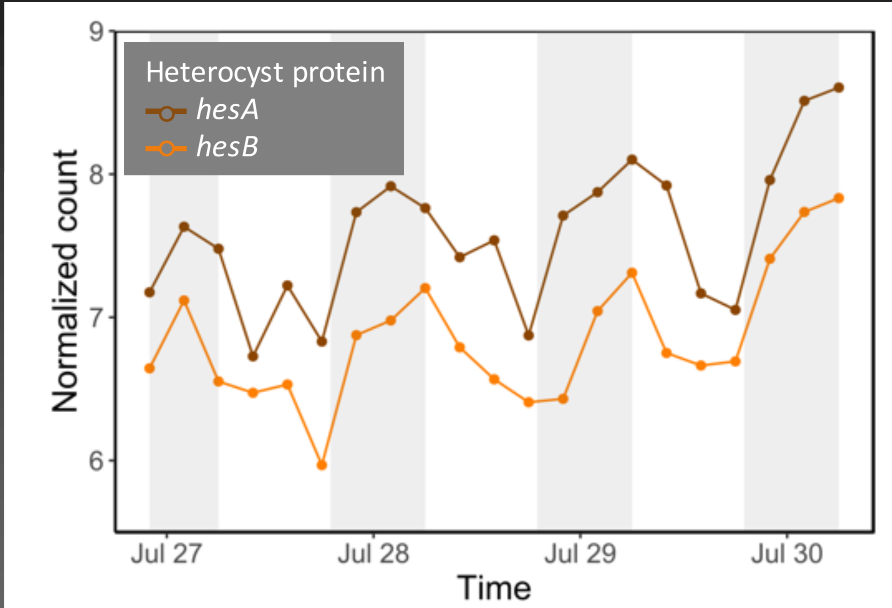
Carbon fixation and growth (Si uptake - cell wall formation) are tightly linked to the light-dark cycle

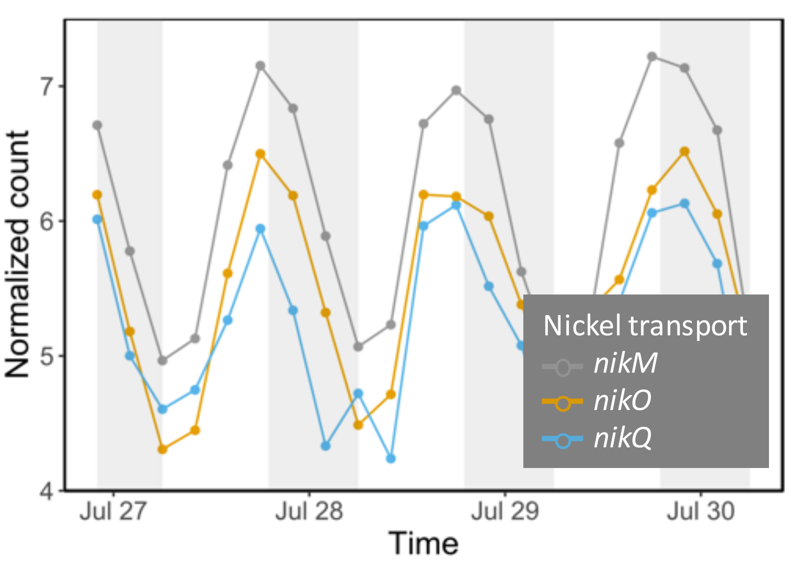
Richelia symbiont has strong diel patterns

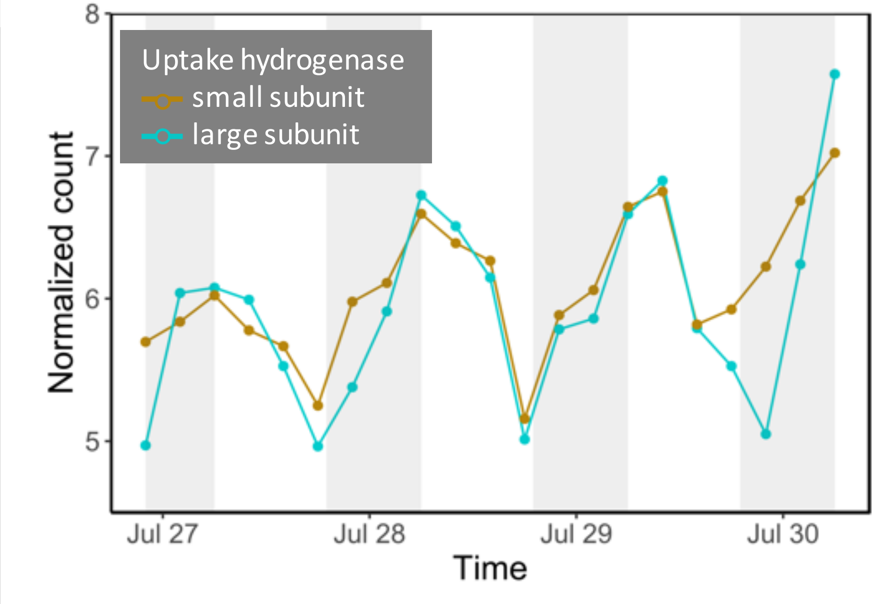


Harke et al. (2018) *ISMEJ*

Coordinated expression of genes related to N_2 fixation in *Richelia*



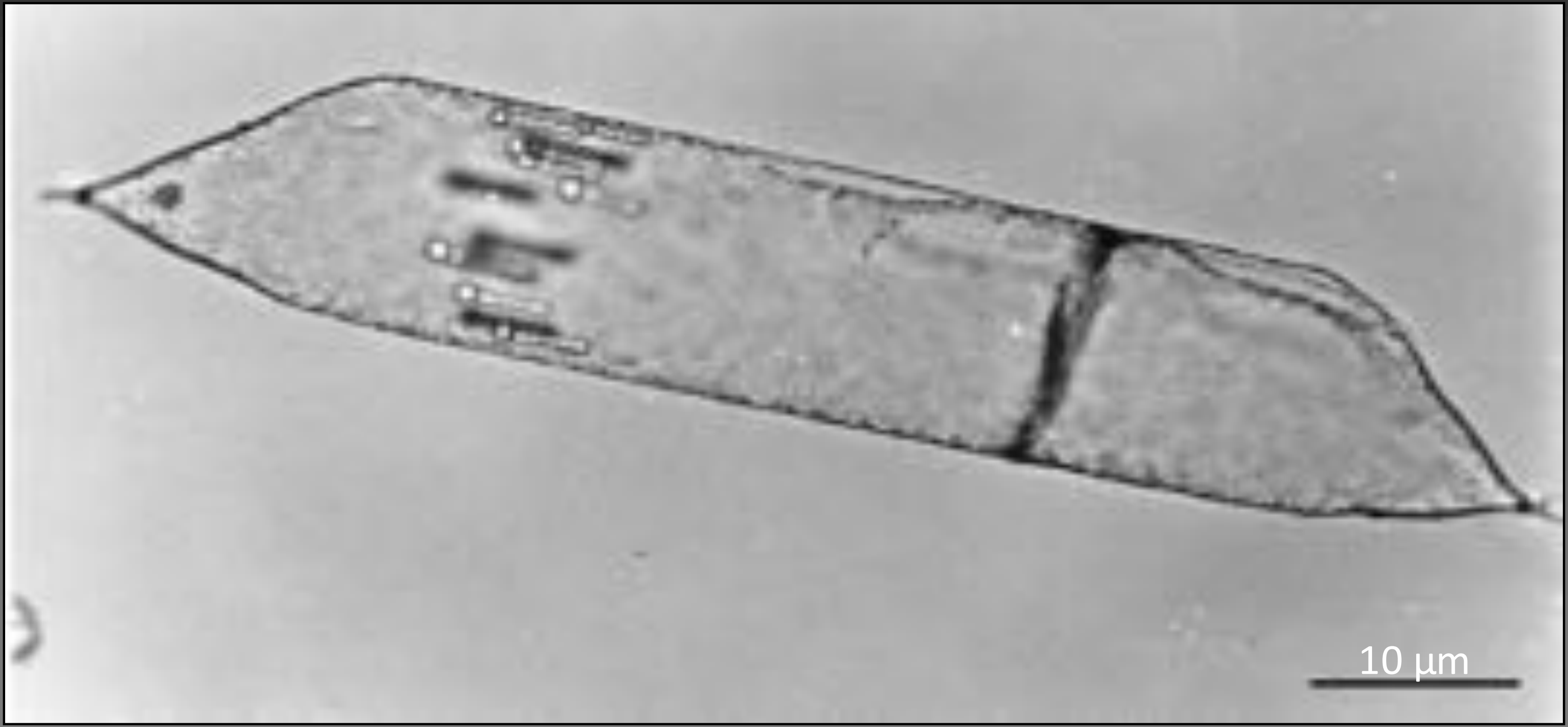




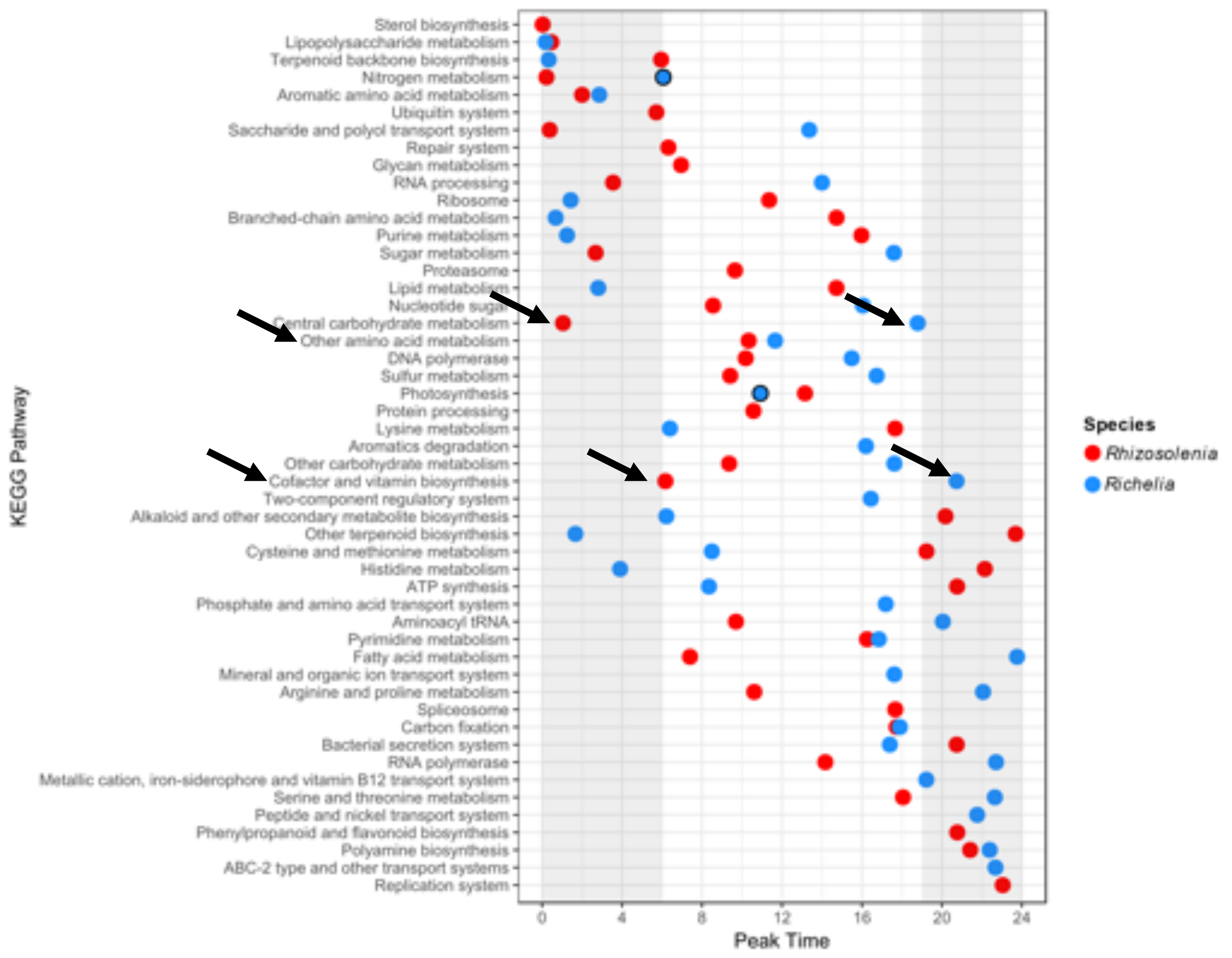
Harke et al. (2018) *ISMEJ*

Timing of expression suggests coordination of processes supporting N_2 fixation

Coexpression between host and symbiont

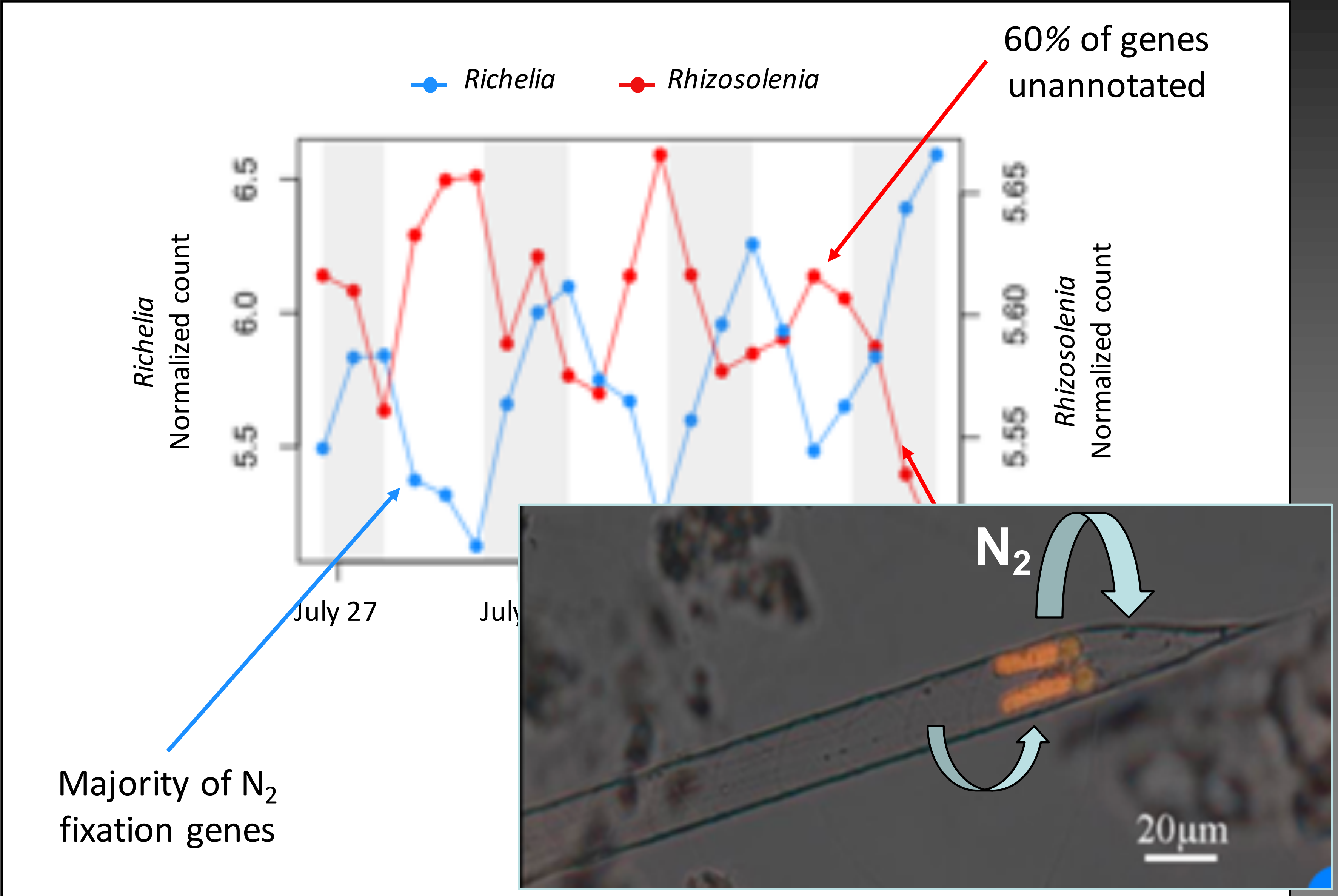


Cascade in metabolic timing between host and symbiont



Metabolic offset for many KEGG pathways - although both are photosynthetic, core metabolism is offset

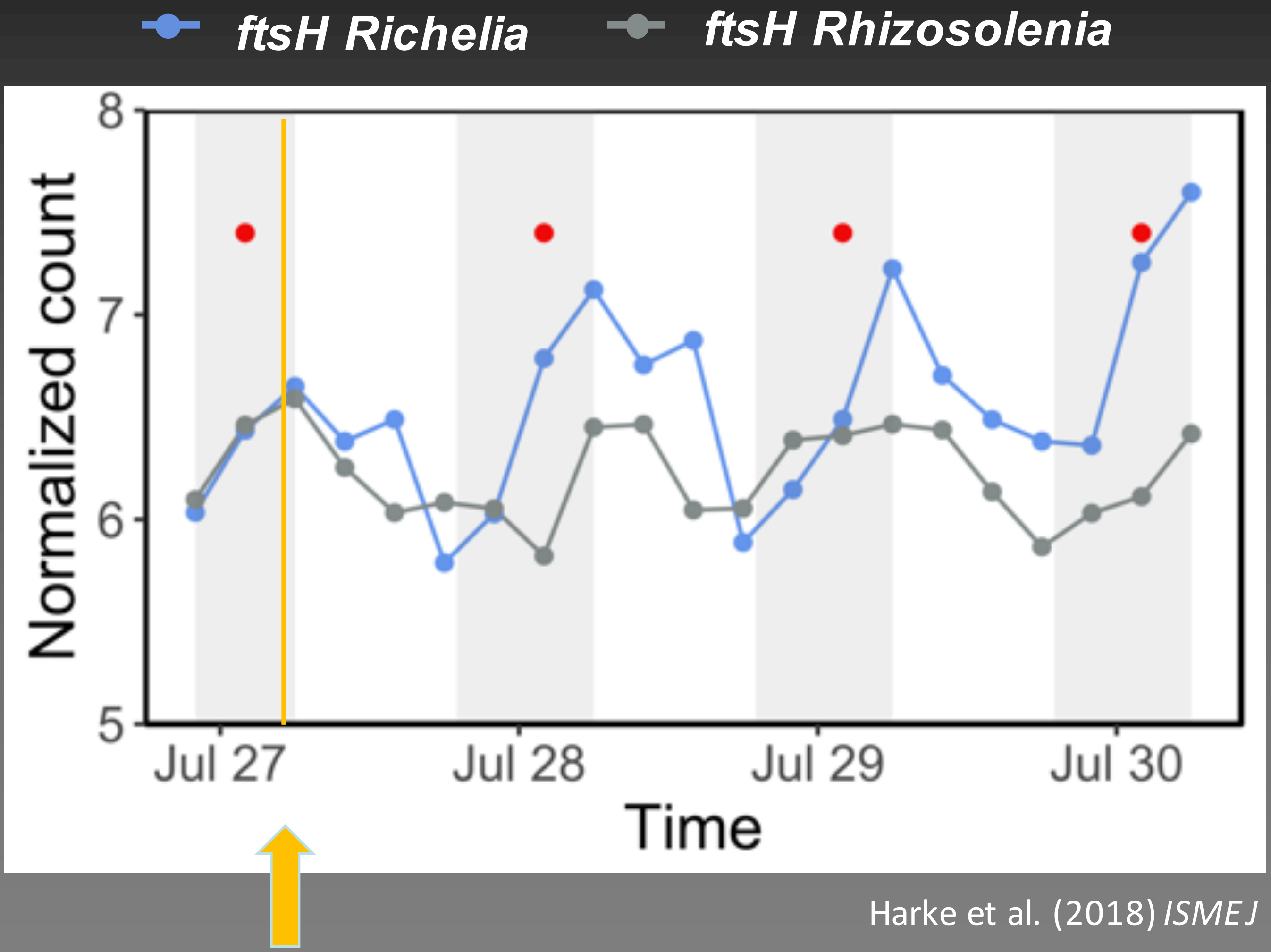
Coordinated expression between host and symbiont in N_2 fixation module using WGCNA



60% of genes unannotated

Majority of N_2 fixation genes

Is cell division coordinated to maintain association?

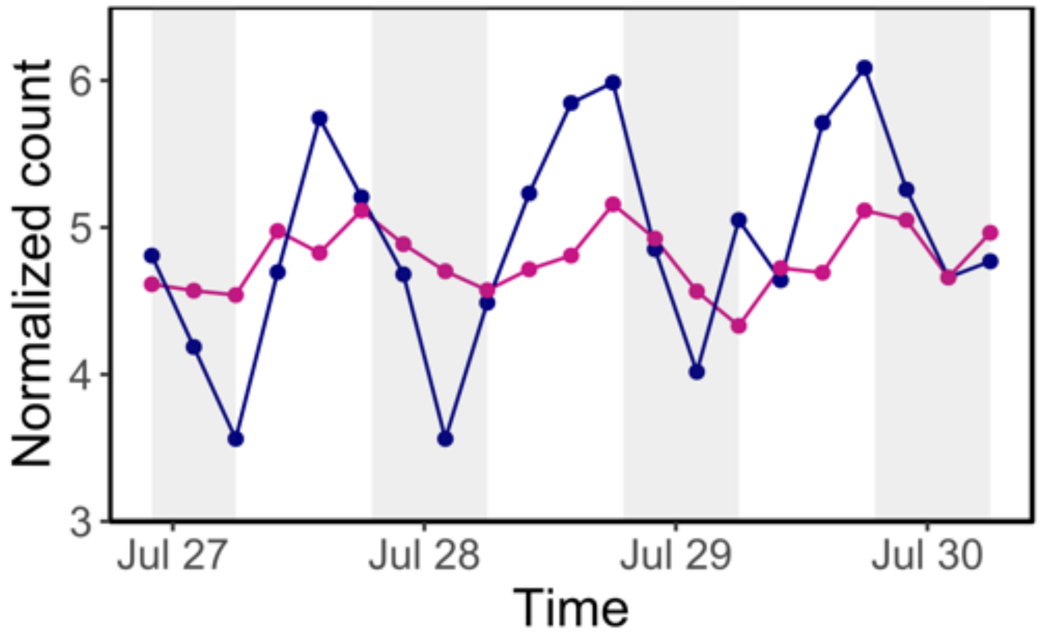


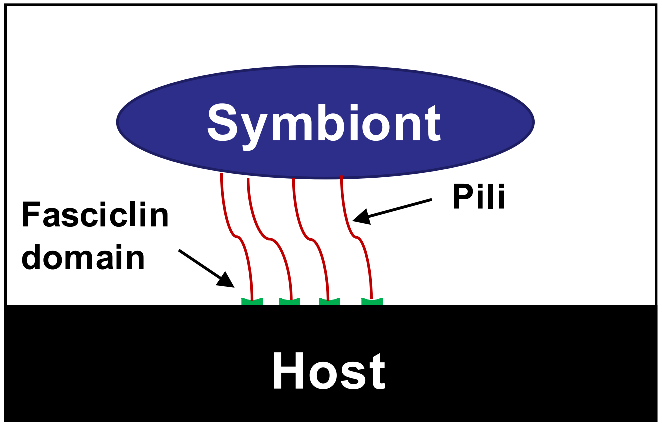
Harke et al. (2018) *ISMEJ*

Division timing

Villareal et al. 1989

How is the association maintained?





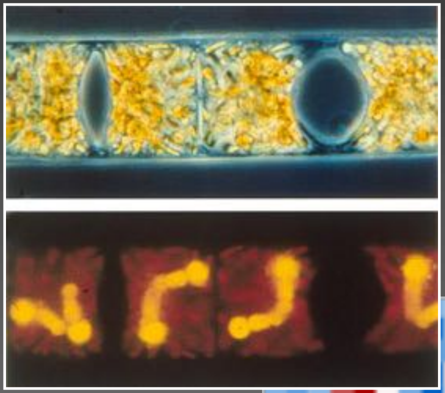
- **Genes involved in cell adhesion have similar peak expression timing**

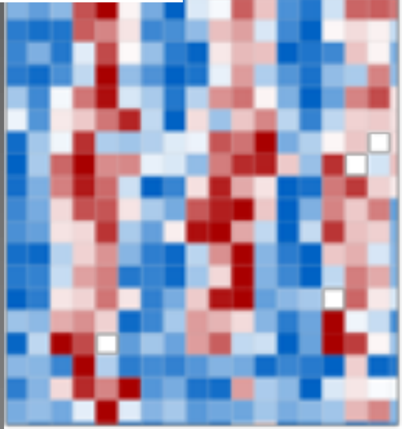
Harke et al. (2018) *ISMEJ*

Summary – Interactions coordinated responses in DDA

Are there patterns of coordinated expression that underpin host - symbiont physiology?

- **Selected Metatranscriptomes**: Host physiology is driven independently by light
- **Unselected Metatranscriptomes**: Symbiont physiology is driven independently by light.
- **Dual Metatranscriptomes**: Light drives coordinated patterns that underpin N exchange, cell division, and maintenance of symbiosis





Core questions

Nutrient responses

- Proof of concept: How do different groups respond to nutrient pulses?

Light responses

- How does light drive metabolic cascades across phyla?

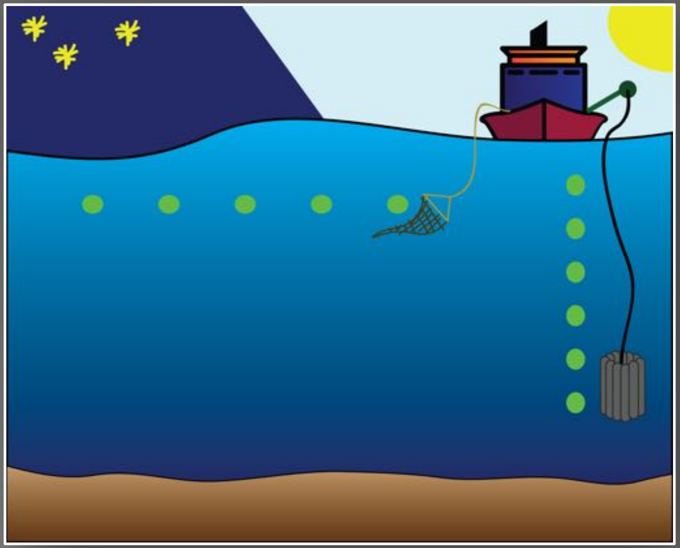
Interactions

- Are there patterns of coordinated expression that underpin host-symbiont physiology?

- How does light drive interactions between the microbiome and host?

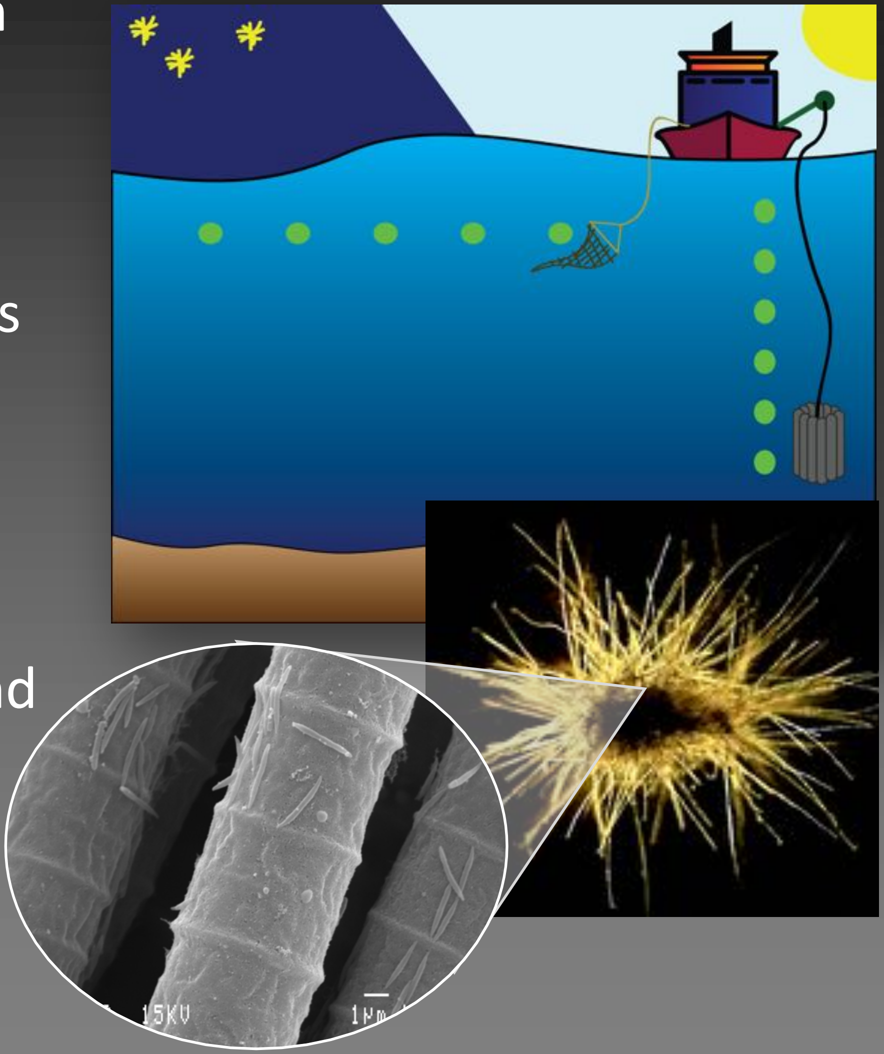
HOE:DYLAN

MAHALO



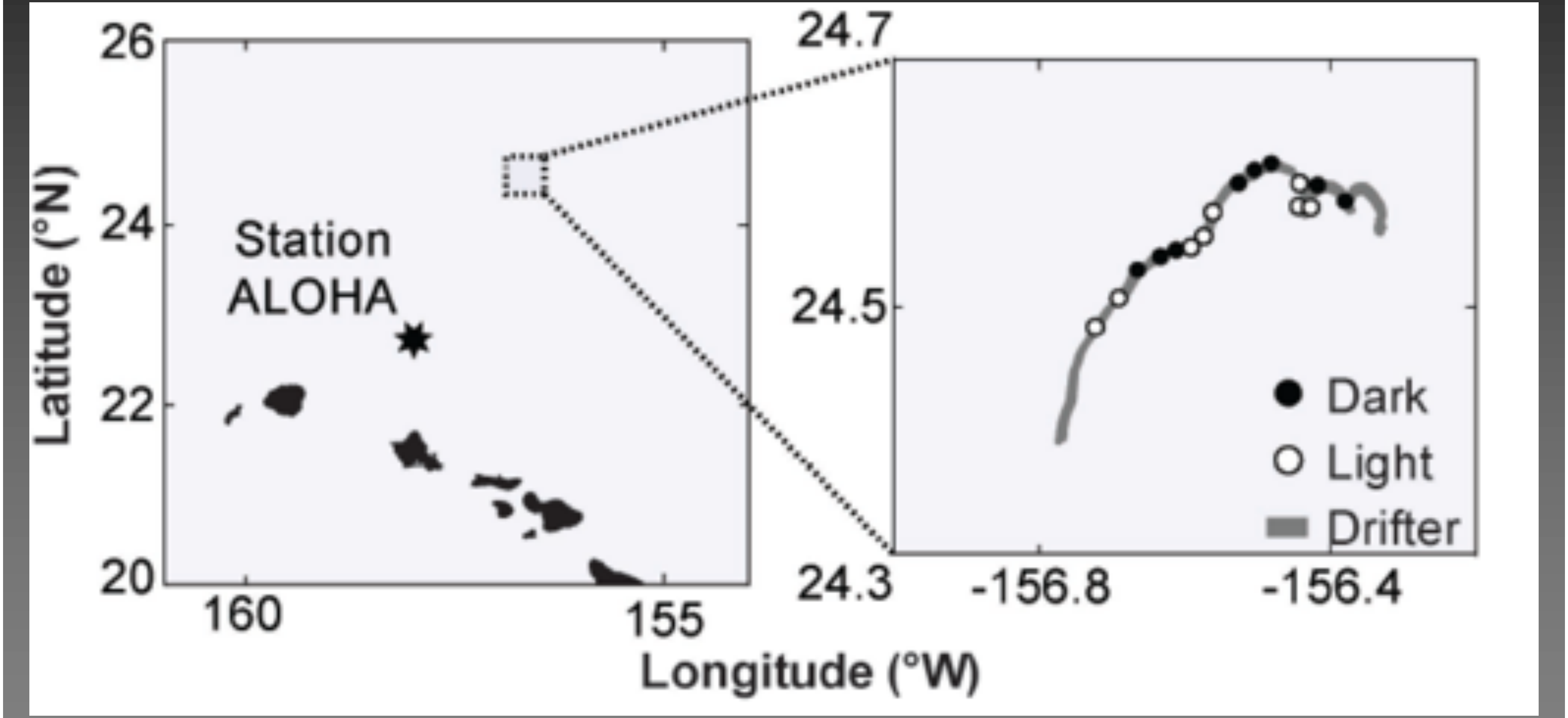
Using light to study interactions

- *Trichodesmium* plays a critical role in the ecology of the NPSG by fixing N_2 that is otherwise limiting.
- Little is known about the interactions between this colonial cyanobacteria and its heterotrophic epibiotic microbiome
- Do epibionts have diel responses, and what might they tell us about interactions between the host and microbiome?



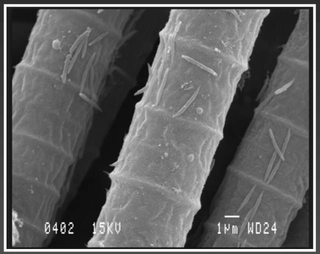
How does light drive interactions between the microbiome and host?

Studying the holobiont … using a time series of daily rhythms



Colony samples were taken every 4 hours for 72 hours, twice!

Light-driven dynamics in host and microbiome



Metatranscriptomes

↓

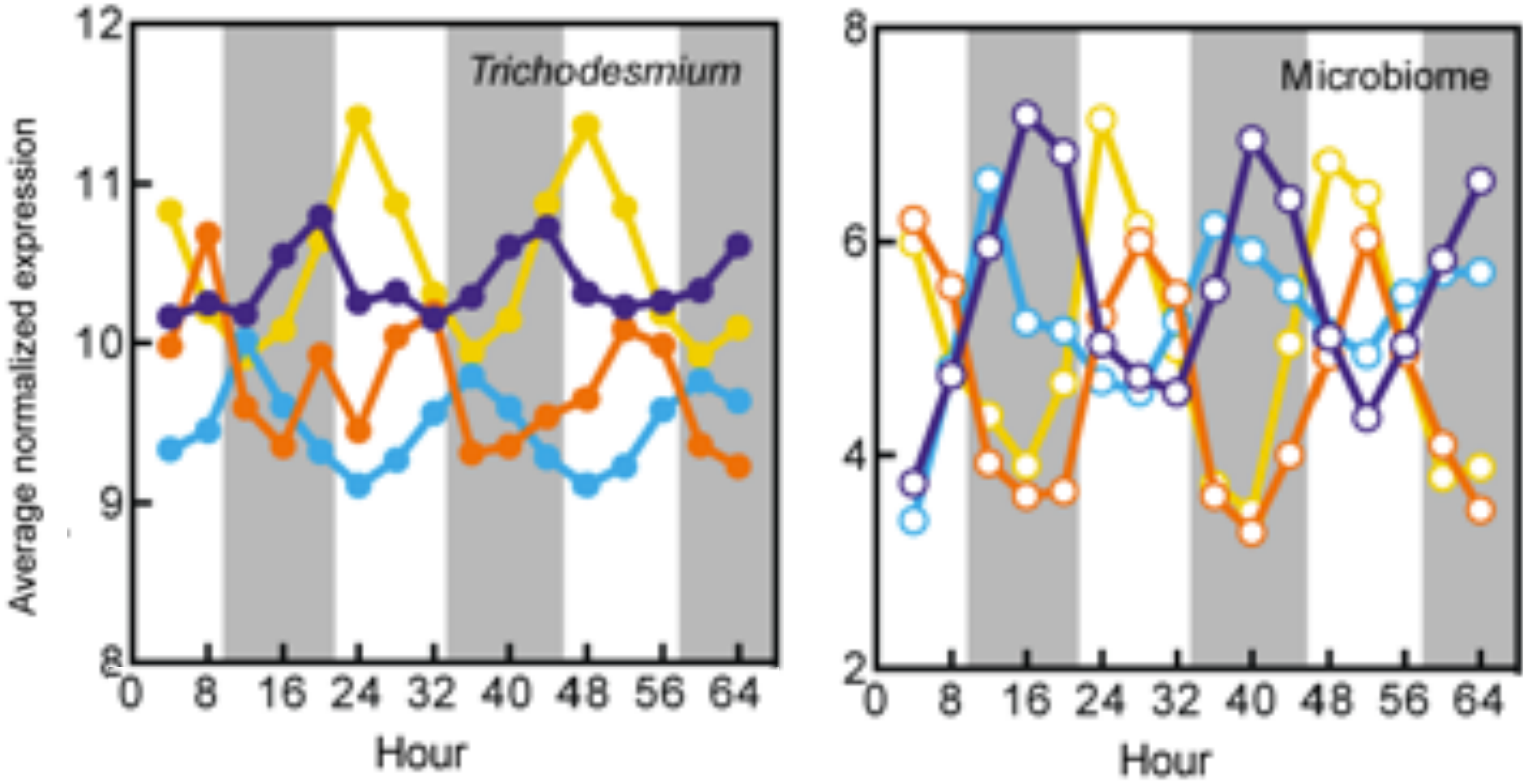
Orthologous group analysis

↓

Periodicity: RAIN

Co-expression: WGCNA

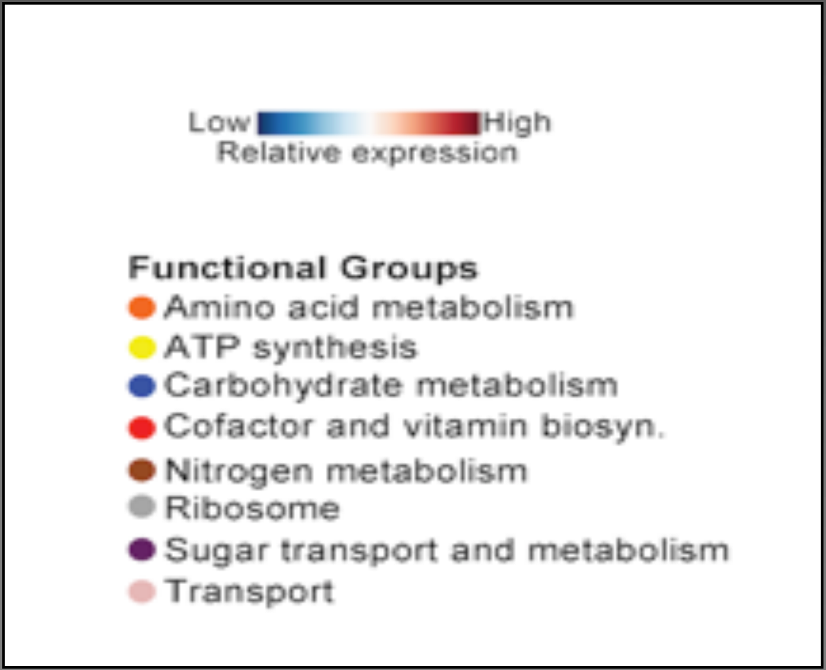
Diel modulation of transcripts in *Trichodesmium*



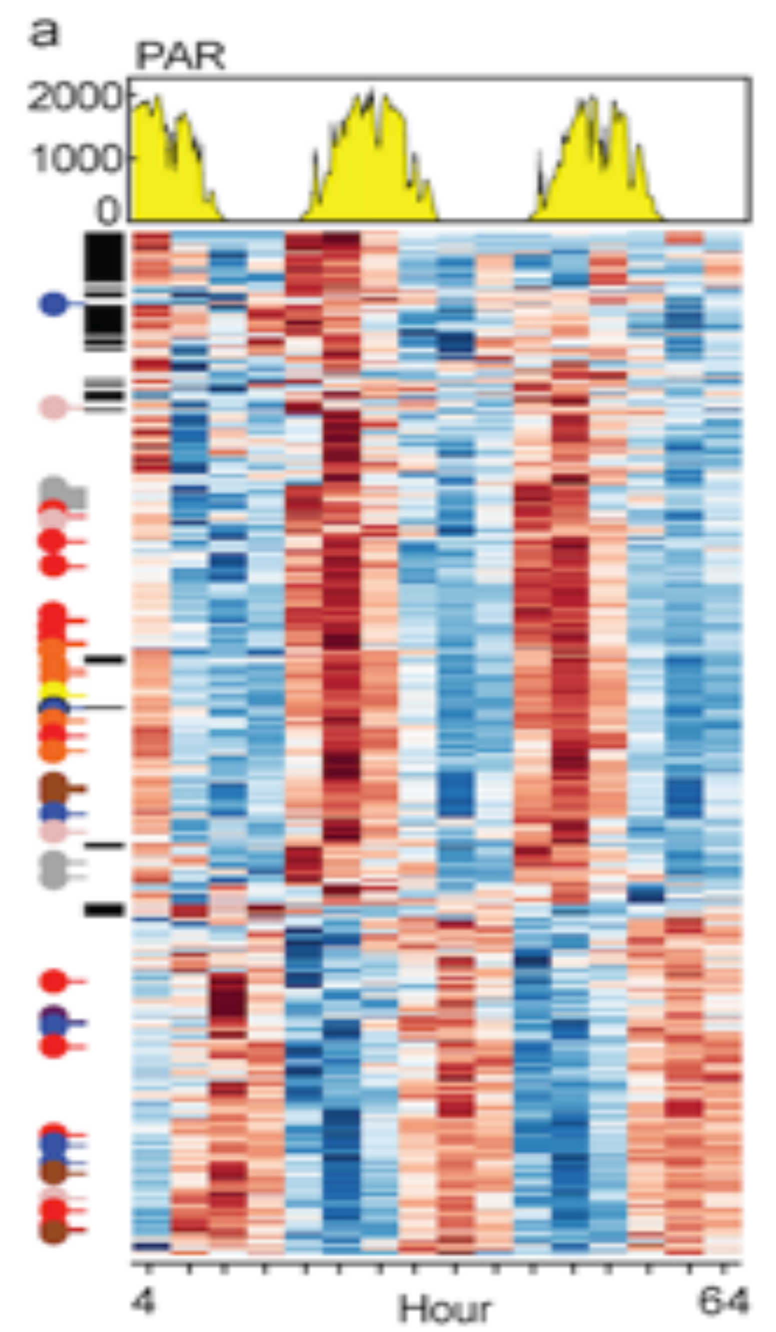
Frischkorn et al. (2018) *ISMEJ*

WGCNA co-expression network

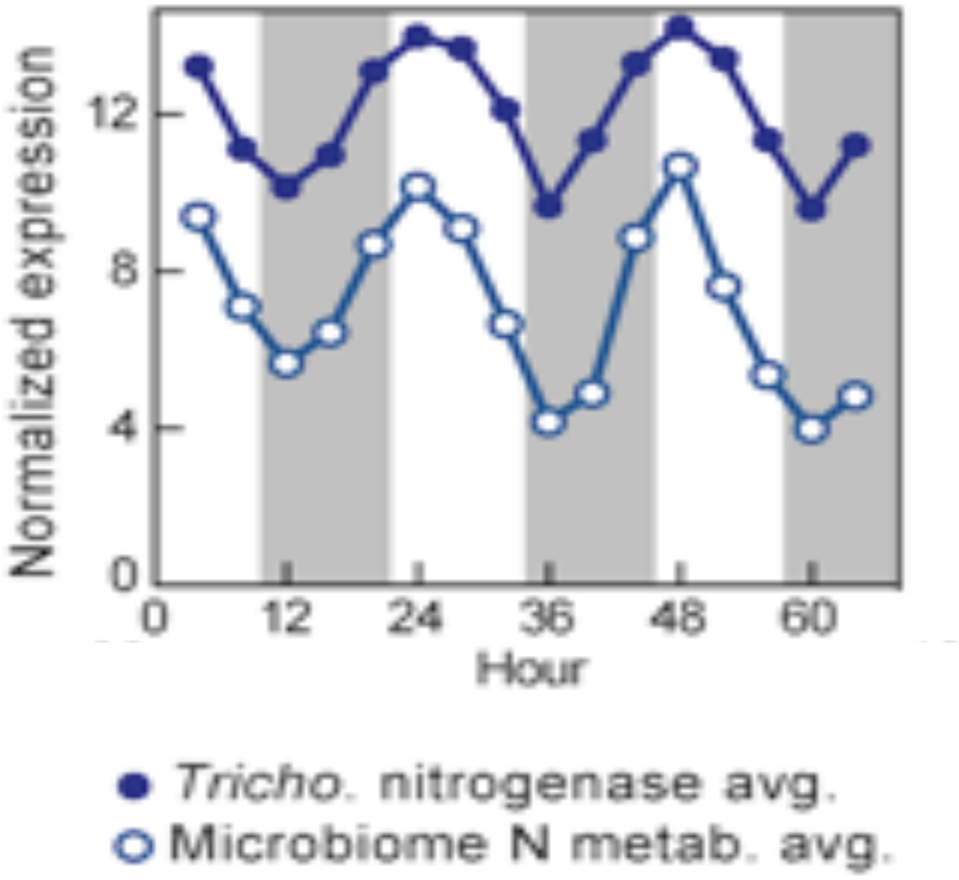
Module example: Significantly coordinated expression patterns between *Trichodesmium* and microbiome.

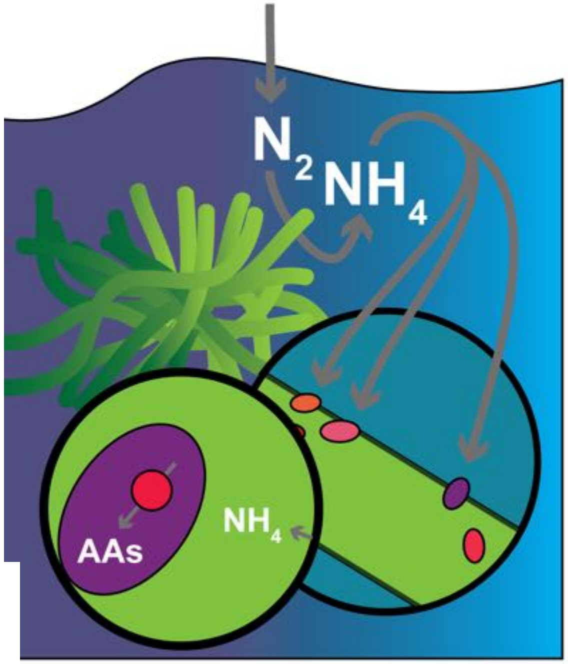


Frischkorn et al. (2018) *ISMEJ*



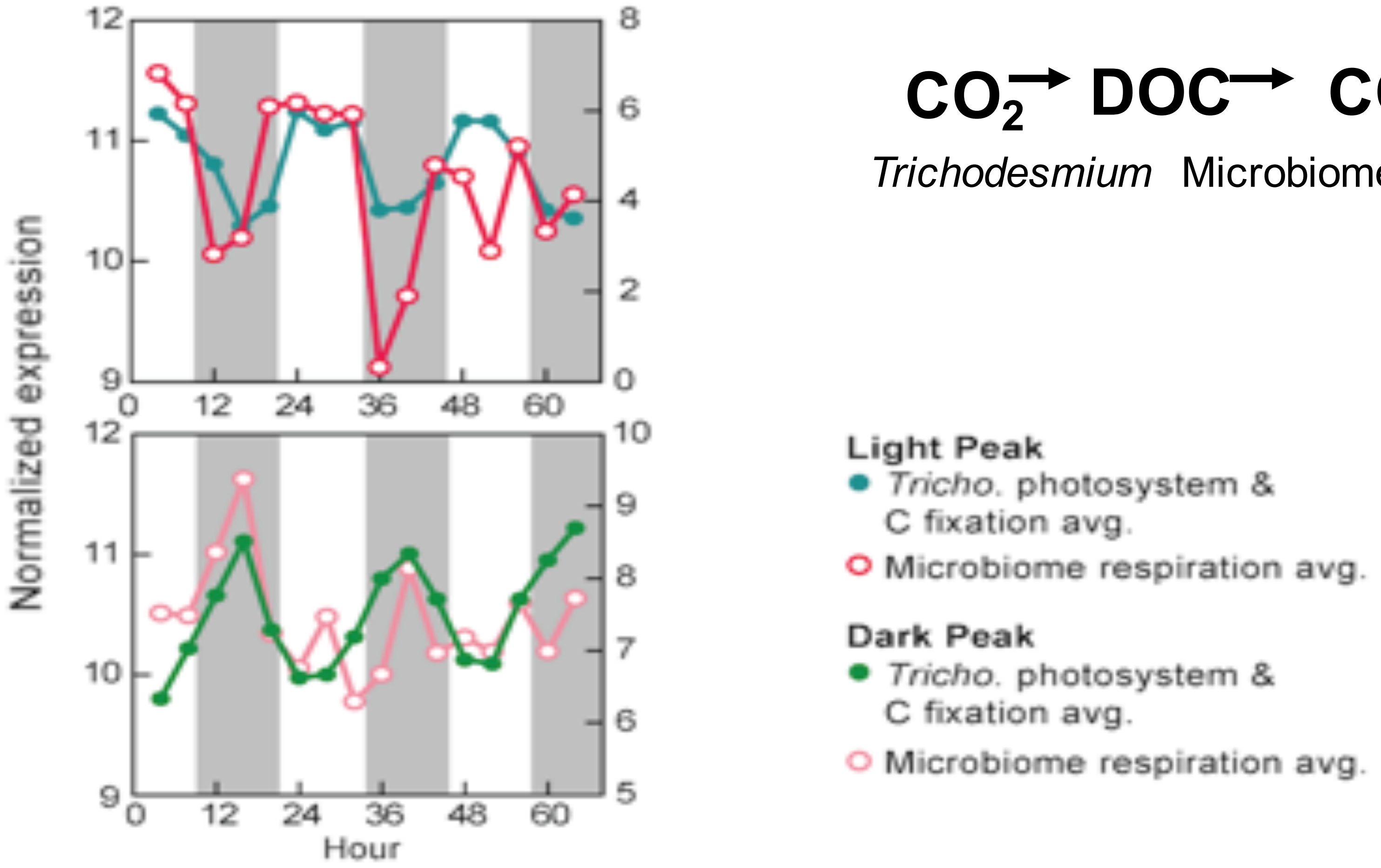
Diel modulation of transcripts suggest coordination: Nitrogen





Frischkorn et al. (2018) *ISMEJ*

Diel modulation of transcripts suggest coordination: Carbon



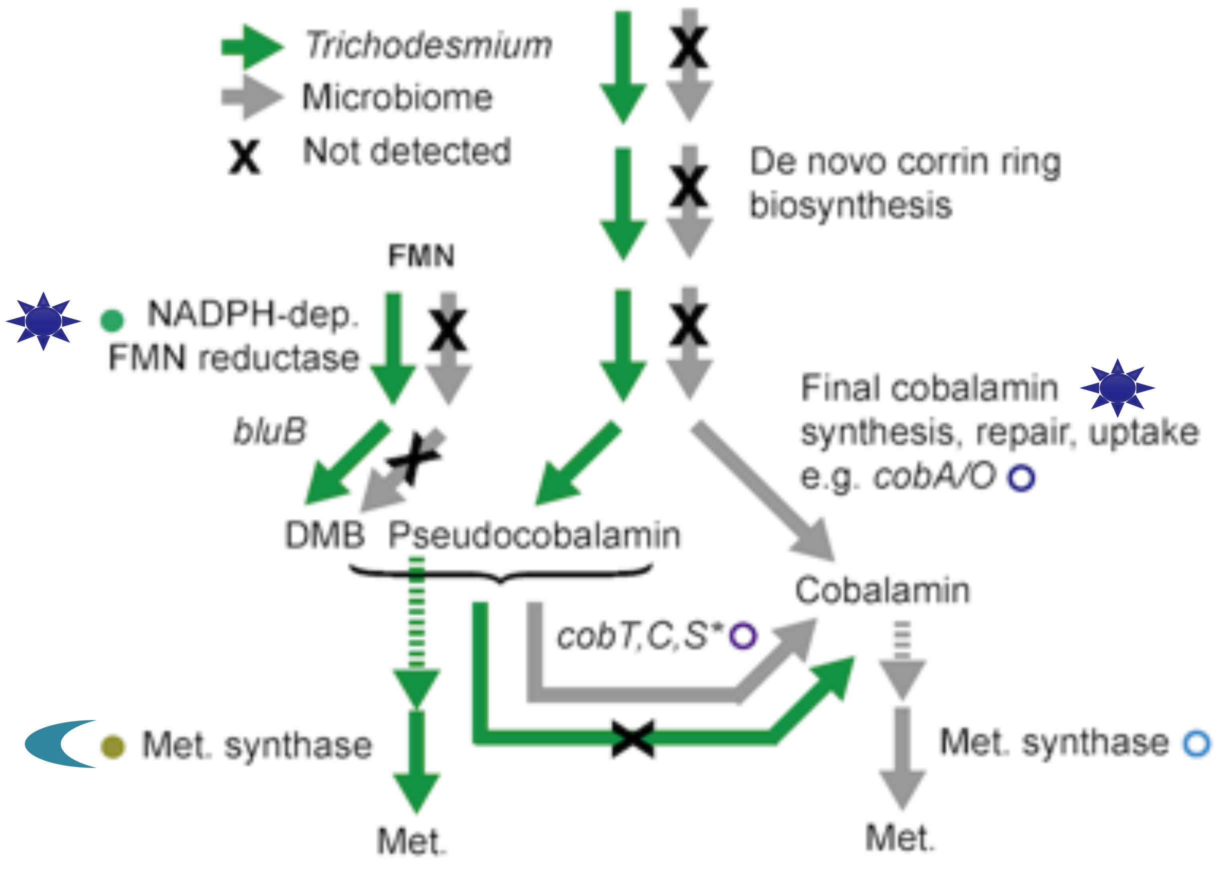
$CO_2 \rightarrow DOC \rightarrow CO_2$

Trichodesmium Microbiome

Frischkorn et al. (2018) *ISMEJ*

Diel modulation of transcripts suggest coordination: Vitamins

☾ Night peak ✷ Day peak



Trichodesmium
Microbiome
X Not detected

De novo corrin ring biosynthesis

FMN

✷ ● NADPH-dep. FMN reductase

bluB

Final cobalamin synthesis, repair, uptake ✷
e.g. *cobA/O* ○

DMB Pseudocobalamin

Cobalamin

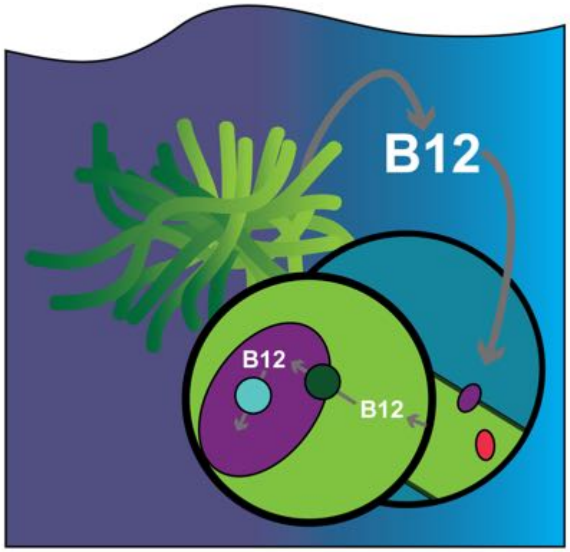
*cobT,C,S** ○

☾ ● Met. synthase

Met. synthase ○

Met.

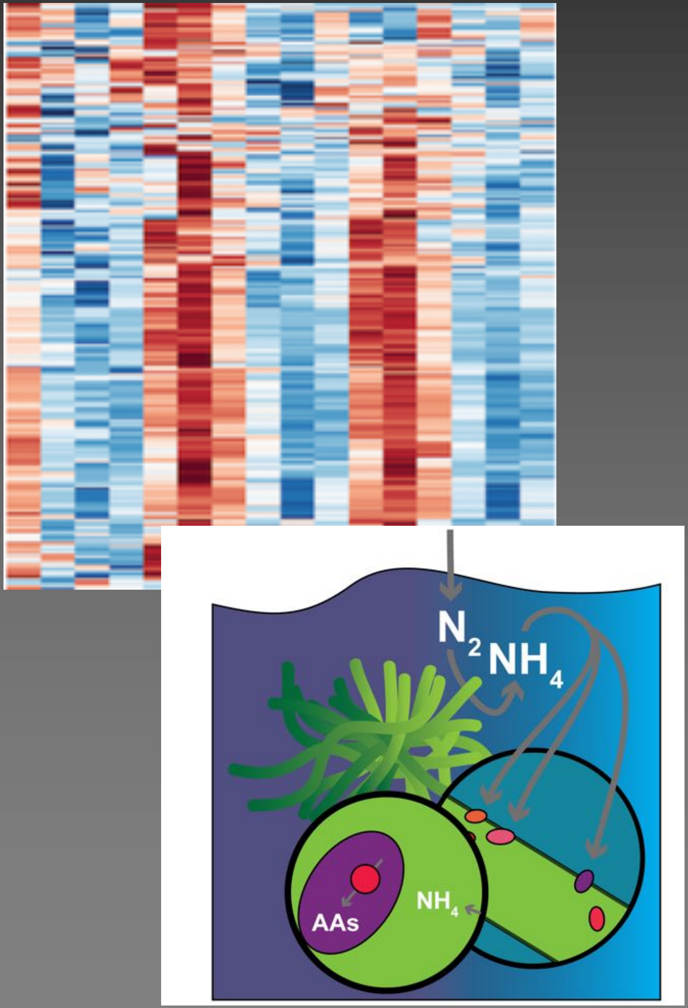
Met.



Summary – Interactions coordinated responses between host and heterotrophic microbiome

How does light drive interactions between the microbiome and host?

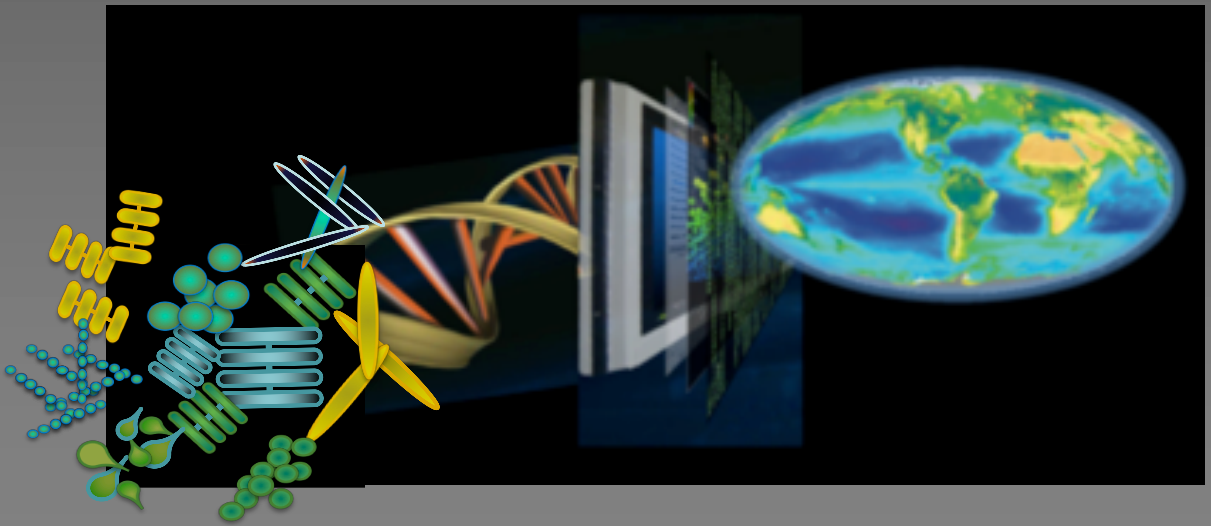
- **Metatranscriptomes**: Heterotrophic epibionts have strong diel patterns, which are coordinated with the patterns present in the host
- **Metatranscriptomes**: Dynamics suggest the exchange of nitrogen, carbon, and vitamins



Conclusions

Transcriptome-enabled approaches are providing new tools to identify the physiological ecology of single species and their role in ecosystem structure and function.

Lessons learned: We need more reference sequences and to think creatively about how we approach and mine sequence data. Need to sample at the same time of day to avoid diel variability.

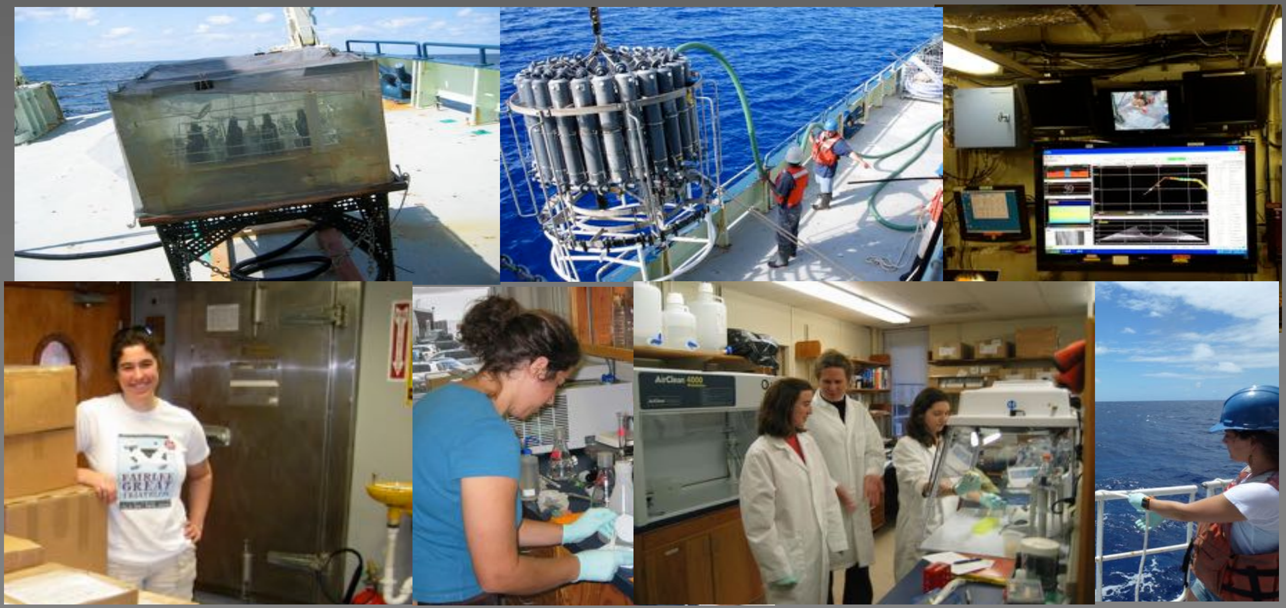


Acknowledgements

- Dyhrman Lab: Dr. Rachel Wisneiwski Jakuba, Dr. Elizabeth Orchard, Dr. Louie Wurch, Abigail Heithoff, Dr. Harriet Alexander, Dr. Kyle Frischkorn, Dr. Alena Sevucu, Dr. Julia Diaz, Dr. Monica Rouco Molina, Dr. Matt Harke, Dr. Gwenn Hennon, Hannah Joy-Warren, Maria Hernandez, Sheean Haley and many undergraduate interns.

- Collaborators: Ben Van Mooy, C-MORE partnership, Mike Lomas and the BATS Program, Jim Ammerman and the ATP^3 Program, Eric Webb, Claudia Benitez-Nelson, Ellery Ingall, Dan Repeta, Dennis McGillicuddy, John Waterbury, Cabell Davis, Mak Saito, Bethany Jenkins, Tatiana Rynearson, SCOPE

- C-MORE
- Simons Foundation
- NSF
- EPA
- NOAA
- DOE, JGI
- WSL
- Angell Foundation



Other things to talk to me about

- Experimental evolution
- Future ocean dynamics
- Harmful algal blooms
- Proteomics and metabolomics
- Working in Antarctica
- Learning in virtual environments

Whyville

Click on me when you're ready to hear about my lab.

My seedlings made the water not red anymore!

More seedlings = no red tides.

That's great!

Yeah