Ocean genomics: microbial physiological ecology across marine ecosystems

Sonya Dyhrman
Professor, Department of Earth and Environmental Sciences
Senior Staff, Lamont-Doherty Earth Observatory
Columbia University
I have my dream job
Sonya Dyhrman

California
PhD Scripps

Krumlov - 2016

Massachusetts
Postdoc WHOI

Scientific Staff WHOI

New York: Faculty Columbia University
Columbia runs RV Langseth
Sonya – random (fun?) facts

C-CoMP at COP27
Dr. Sonya Dyhrman
The Key Role of Microbes in the Surface Ocean
November 16th, 2022
Key points for today

- Highlight how we ask and answer questions in the ocean with genomic tools

- Eukaryote physiological ecology
  - *Metatranscriptomes*

- Microbial interactions
  - *16S Amplicons*
  - *Metagenomes*
  - *Metatranscriptomes*

- Share my enthusiasm for marine microbes and ‘omics
The ocean makes our planet habitable
The vast unseen microbial populations play a critical role in ocean function
Marine Microbes - fundamental to ocean ecosystem function

- Marine microbes...
  - Produce and consume green house gases
  - Supply the marine food web
  - Recycle organic matter
  - *make the planet habitable*

Image courtesy C-MORE

phyto- = photosynthetic
-plankton = floating organisms
Phytoplankton play a profound role in the earth system

Phytoplankton drive half of global primary production
Seasonal chlorophyll distributions in the sea - highlights the global significance of phytoplankton
Phytoplankton underpin ocean ecosystem function

Haptophytes

Diatoms

N\textsubscript{2} Fixers

Chisholm 2000 Nature
The chemical – microbe network drives the fate of a lot of carbon.
Ocean conditions are changing and this will change the distribution and activities of microbes
Changes in temperature \( \text{CO}_2 \), pH, etc.
Models predict future shifts in microbial communities

What are the rules that govern the chemical – microbe network?

Who is there?
What are they doing?

If we learn the rules for how the system operates – then we have a better chance of predicting future ocean responses

‘Omic-enabled advances allowing to query cells in their environment in a species-specific way

Graphic: WHOI
Challenges and opportunities

- Long standing challenges:
  - Populations are dilute and they move around!
  - Salt!
  - Some keystone groups not in culture
  - Few genomes or transcriptome sequences (particularly for the eukaryotes)

- New opportunities:
  - Increases in genomes/transcriptomes for eukaryotic taxa
  - New array of bioinformatics and sequencing tools

A decade of meta ‘omic and multi ‘omic approaches!
100 Diatom Genomes

Diatoms are the most species-rich algal class with estimated ≥ 100,000 species. They contribute to 20% of global carbon fixation and oxygen production. However, with genomic information from only about 10 diatoms, our ability to harness their unique biology is very limited. Consequently, this project will sequence 100 diatom species for providing unique insights into their roles as key players for capturing carbon dioxide and as the foundation of diverse aquatic food webs. These insights will also be critical for advancing diatom-based biotechnology and synthetic biology platforms.

Proposer: Thomas Mock, University of East Anglia (UK)
Proposal: 100 Diatom Genomes Project

Scanning electron micrographs of diverse diatoms. (Credits: Diana Sarno, Marina Montesori, Nicole Poulsen, Gerhard Dieckmann)
Challenges and opportunities

• Long standing challenges:
  • Populations are dilute and they move around!
  • Some keystone groups not in culture
  • Few genomes or transcriptome sequences for eukaryotic taxa

• New opportunities
  • Increases in genomes/transcriptomes for eukaryotic taxa
  • *New array of bioinformatics and sequencing tools*

A decade of meta ‘omic and multi ‘omic approaches!
Vignettes

- Co-existing in a sea of competition: tracing drivers of phytoplankton physiological ecology in the open ocean

- Microbiome interactions: Tracking the microbiome of a keystone N₂ fixer
Thank you Dyhrman Lab family past and present
Acknowledgements

- Dyhrman Lab and many extraordinary collaborators
- NSF
- DOE
- Simons Foundation
- World Surf League
- Angell Family Foundation
Leveraging metatranscriptome data to study the chemical – microbe network

- Cultures offer controlled conditions – link genes to functions and study their regulation
- Field samples offer context
- Metatranscriptomes are species-specific
- Metatranscriptomes offer high resolution tracking of finely tuned responses
Microeukaryote metratranscriptomes with RNA-Seq

3x20L

Filter (>5μm)

Extract RNA (and standards)

rRNA depletion

De novo assembly

Reference mapping

Unselected

De novo assembly

Reference mapping

Selected

Extract RNA (and standards)

3x20L

Filter (>5μm)

rRNA depletion
<table>
<thead>
<tr>
<th>Quality Control</th>
<th>Removal of Sequences</th>
<th>Functional Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trimmomatic</td>
<td>Salmon</td>
<td>BLAST</td>
</tr>
<tr>
<td>FastQC</td>
<td>Kallisto</td>
<td>DIAMOND</td>
</tr>
<tr>
<td></td>
<td>Bowtie2</td>
<td>HMMer</td>
</tr>
<tr>
<td></td>
<td>RSEM</td>
<td></td>
</tr>
<tr>
<td>Assembly/Reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trinity</td>
<td></td>
<td>KEGG</td>
</tr>
<tr>
<td>TransABYSS</td>
<td></td>
<td>KOG</td>
</tr>
<tr>
<td>rnaSPADES</td>
<td></td>
<td>GO</td>
</tr>
<tr>
<td>MEGAHIT</td>
<td></td>
<td>PFam</td>
</tr>
<tr>
<td>Read Mapping</td>
<td>Protein Prediction and Clustering</td>
<td></td>
</tr>
<tr>
<td>Salmon</td>
<td>TransDecoder</td>
<td></td>
</tr>
<tr>
<td>Kallisto</td>
<td>OrthoFinder</td>
<td></td>
</tr>
<tr>
<td>Bowtie2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RSEM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normalization &amp; DE</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EdgeR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DESeq2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TPM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Variance stablization</td>
<td></td>
</tr>
</tbody>
</table>

**Taxonomic Annotation**
- BLAST
- DIAMOND
- MMSeq2
- MMETSP
- EukProt
- EukZoo
- PhyloDB
- MarFERRReT (2024)

**Cohen et al. (2020)** *Front. Mar. Science*
Sequencing and analytical workflows are constantly improving

Harriet Alexander @nekton4plankton · Apr 27

So check it out!

GitHub here: github.com/AlexanderLabWH...
Documentation here: eukrhythmic.readthedocs.io/en/latest/
jEUKbox here: eukrhythmic.readthedocs.io/en/latest/

Feedback / comments welcome!
Sequencing and analytical workflows are constantly improving.

Krikinos et al. 2023 BMC Bioinformatics
• Assembly captures a swath of reads in high abundance – highly diverse groups (red)

• Reference mapping captures a swath of reads in low abundance groups (blue)

• Generally equally good at recapitulating patterns when references are from local isolates (orange and yellow)
Metatranscriptome intercalibration effort – phase 1

- Large diameter filter sampled with CLIO (June 2023)
- 2x filter pieces to ~ 7 groups
- Evaluate:
  - Extraction
  - Library prep
  - Workflow intercomparison
- Will not answer all questions regarding intercalibration – but it will be a valuable starting point for building intercomparable datasets.

Thank you N. Cohen and H. Alexander for leadership and M. Saito for samples
Limitation

- What limits carbon cycling? – tracing physiological ecology in an expanding oligotrophic ocean with metatranscriptomics

Competition

- A day in the life – tracing competitive interactions with multi-’omics
Complex community dynamics driven in part by resources.
How do resources drive phytoplankton distributions and activities?
North Pacific Subtropical Gyre

This resource poor (oligotrophic) region is arguably the largest biome on the planet

Image credits: SeaWIFS Global Chlorophyll, H. Alexander
Blooms of keystone species can shift ecosystem state and carbon flow

Net Heterotrophy

- Abundant microbiome
- Microbial loop
- Low abundance Keystone microbiome

Net Autotrophy

- Diatoms
- Nitrogen Fixers *Trichodesmium*
- Haptophytes
- Increased Export

Nutrient input, or other forcings

Graphic: H. Alexander
*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 eukaryotic marine phytoplankton to have multiple strains sequenced, identifying pan genome

*Gephyrocapsa* (Bendif et al., 2023; Filatov et al., 2021)
Hawaii Ocean Experiment: Dynamics of Light and Nutrients

Wilson et al. 2015 GBC

center for microbial oceanography: research and education

(c-more) linking genomes to biomes

(HOE-DYLAN)
HOE: DYLAN: Water sampling
**HOE:DYLAN: Semi-factorial resource additions**

- **Nitrate/nitrite**: ~40 µM
- **Phosphate**: ~2.6 µM
- **Silicate**: ~100 µM
- **Fe**: ~0.6 nmol/kg
- **B_{12}**: <0.1 pM

**in situ**  |  Control  | Deep seawater amendment  | (+N, Si, Fe, B_{12})  | (+P, Si, Fe, B_{12})  
---|---|---|---|---
IS  | C  | DSW  | +N  | -P  | -N  | +P

**N addition**
Sampling and pipeline

Collect samples from St. ALOHA

>5µm

6 day incubations

Isolate mRNA

AAAAAA

AAAAAA

AAAAAA

Submit samples to sequencing center

RSEM Mapping

DE with edgeR

Custom reference database of *E. huxleyi* reference sequence

Target 60 million 100 bp paired-end reads

6 day incubations

Custom reference database of *E. huxleyi* reference sequence
E. huxleyi physiological response suggests N control

Log$_2$ Average Abundance

Significantly Differentially Abundant Genes

In situ transcriptome responses track with culture proteome

Nitrogen limits *E. huxleyi* growth and calcification in NPSG

Nitrogen is a driver of *E. huxleyi* dynamics
Emiliania huxleyi has many diverse isolates.

Cultured strains are highly diverse - isolated from a broad temperature range and displaying considerable physiological diversity.
“Everything is everywhere, but the environment selects”

Does the variable component of the pan genome underpin which strain dominates a given environment?
Sequences from *E. huxleyi* strains

- CCMP1516
- CCMP379
- CCMP374
- CCMP370
- PLYM219

Cluster into orthologous groups to identify core and variable OGs (OrthoMCL)

Map field data and examine shifts in expression (RSEM, edgeR)

<table>
<thead>
<tr>
<th>Gene content of the <em>E. huxleyi</em> species complex</th>
<th>Core</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of OGs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alexander et al. unpub.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Tracking orthologous groups at St. ALOHA

Use diagnostic OGs to track strain distribution

Alexander et al. unpub.
Strain distribution *in situ* and in the incubations

Alexander et al. unpub.
Expression of the variable portion of the pan-genome

Expression of variable gene set significantly increased following N-addition. Alexander et al. unpub.
Ongoing work to define the pan genome

Aim 1: Define the pan genome

Core genes

Strain-specific genes

Aim 2: Identify ecotype sorting in the environment

Aim 3: Characterize pan genomic response to environment

Culture experiments

Transcriptome sequencing

Core

Strain-specific

Harriet Alexander Lead PI

https://alexanderlabwhoi.github.io/
Expanding the pan genome

~15 new whole genome sequences being finalized (Nanopore/Illumina) and RNASeq libraries to support gene prediction
Future work to define the pan genome

Aim 1: Define the pan genome

Aim 2: Identify ecotype sorting in the environment

Aim 3: Characterize pan genomic response to environment

Transcriptome sequencing

Culture experiments

Core
Strain-specific

Alexander et al. unpub.
How do genomic traits underpin competition outcomes?

Haley, Palmer et al. unpub.
Summary

- **Metatranscriptome RNA-Seq:**
  - Species-specific analyses are possible even for low abundance keystone populations.

- Nitrogen limits *E. huxleyi* activity, and influences calcification

- Variable genes in the pan-genome may influence ecotype strain sorting.

- Move towards predictive responses of the chemical–microbe network in a changing ocean
Core questions... adventures in eukaryote metatranscriptomics

Limitation

• Nitrogen limits carbon export and cycling in a keystone haptophyte

• The environment selects

Competition

• A day in the life – tracing competitive interactions with multi-’omics
MAHALO: Special thanks to Daniel Muratore and Joshua Weitz for synthesis efforts
Sampling light driven responses in the NPSG

Legend:
- L: Lipid
- M: Metabolite
- 0.2: >0.2 μm transcript
- 5: >5 μm transcript
Microbes running on time

Diel Cycling of Marine Microbial Metabolism in Complex Communities Partitions Scarce Nutrients and Enables Emergent Biogeochemical Function

Matthew Harke and 3 others View all
Research Scientist, Gloucester Marine Genomics Institute

Jointly applying new multi ‘omic techniques

Kuijawsinski et al. (2011) Annual Review Marine Science
Sampling every 4hrs for many days ....
Sampling and pipeline

- Isolate mRNA
- Submit samples to sequencing center
- Sequence using RNAseq and map reads to reference
- Target 80 million 100 bp paired-end reads
- BWA HTSeq
- RAIN – periodicity
- WGCNA - network

MMETSP ~400 transcriptomes
Genome references and other unigene databases

MMETSP = Marine Microbial Eukaryotic Transcriptome Project

[Links]
https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559
Visualizing diel patterns

~65,000 analyte time series

Different taxa have distinct diel transcription patterns

Carbon fixation and respiration are highly coordinated.

Carbon fixation and respiration are highly coordinated


Carbon metabolism synchronized across all phytoplankton


All photosynthesis, carbon fixation, and TCA cycle processes are synchronized across bacteria and eukaryotes!

Becker, Harke et al. (2021) ISME J
Asynchronous transcription of N uptake and metabolism between microeukaryotes and between taxonomic domains suggest competition for limiting resource is driving temporal niche partitioning

• **Metatranscriptomes:**

• Light drives synchronous metabolic cascades which underpin carbon cycling in the NPSG

• Light drives asynchronous metabolism of nitrogen – a limiting resource in the NPSG

• Temporal niche partitioning of N uptake and metabolism may decrease competition for limiting resources
Nitrogen limits carbon export and cycling in a keystone hapotophyte.

The environment selects.

Temporal niche partitioning supports a complex community in a resource poor environment.