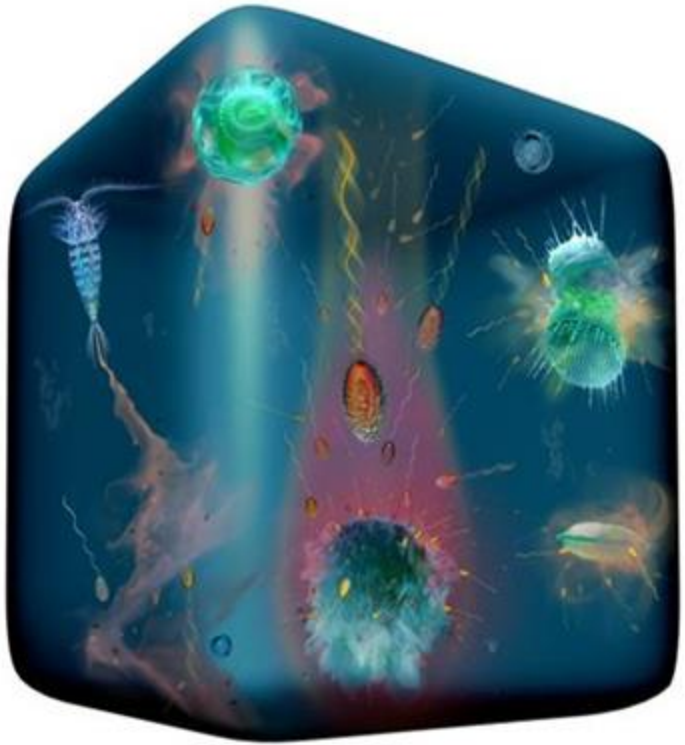


Microbial Cube showing marine microbial micro-environments: Stocker MIT

Exploring microbial activities with metatranscriptomics

Sarah J. Tucker
Postdoctoral Fellow
Marine Biological Laboratory

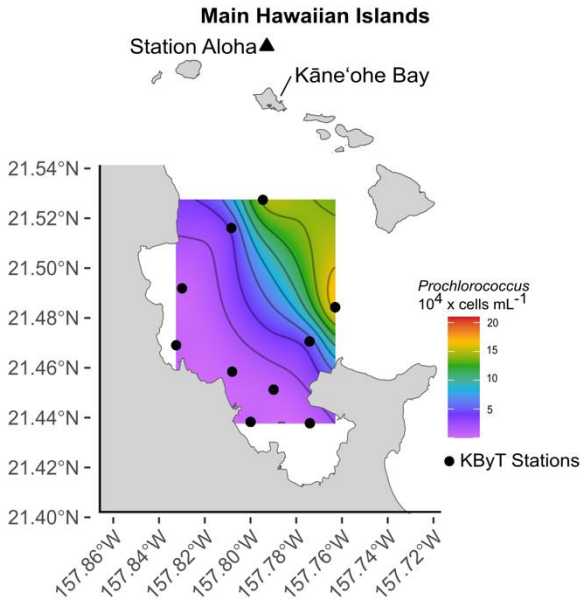
STAMPS 2025

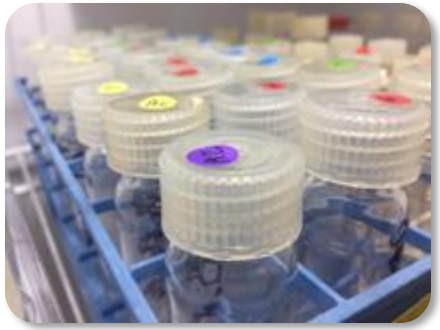
I am interested in the genomic, metabolic, and physiological diversity of marine microbes and how this diversity impacts ecological interactions and ocean processes.

time-series observations | cultivation | omics | community-based research | computation

PhD at the University of Hawai‘i at Mānoa in Dr. Mike Rappé’s Lab

Simons Foundation Postdoctoral Fellow in Marine Microbial Ecology in Dr. A. Murat Eren’s Lab





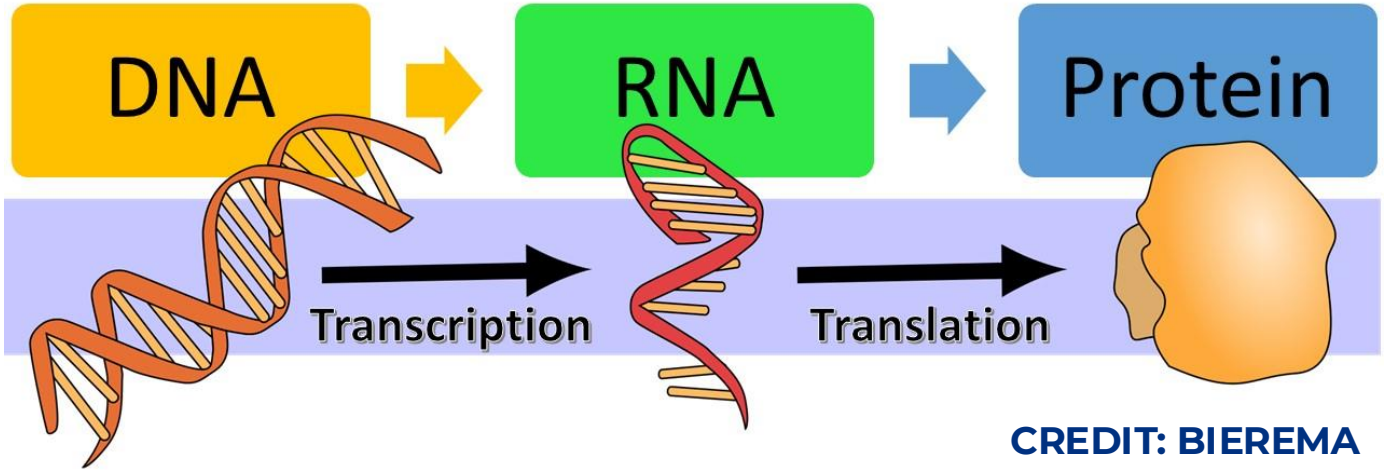
Line-up

- 20 min- Intro to metatranscriptomics
- 30 min- Metatranscriptomics of *Synechococcus*
- 20 min- Normalization strategies
- 20 min- Gene expression changes in *Synechococcus* over a diel cycle

Metatranscriptomics?!?!

Metagenomes: community-level gene potential

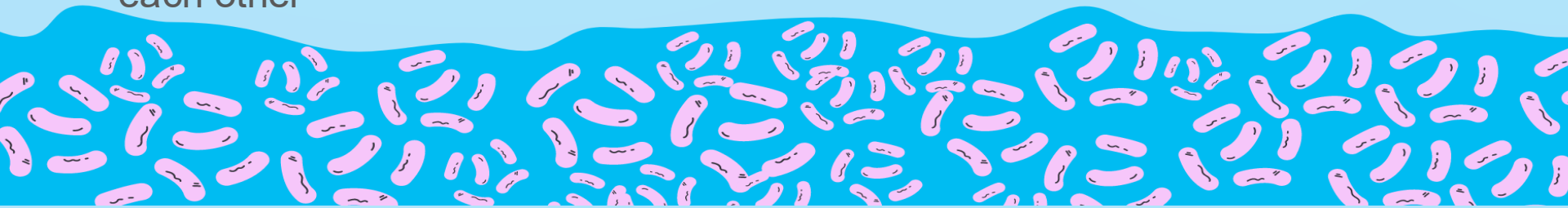
Metatranscriptomes: community-level gene expression
focused on messenger RNA (mRNA)
Portesky et al., 2005



CREDIT: BIEREMA

Why use metatrancsriptomics?

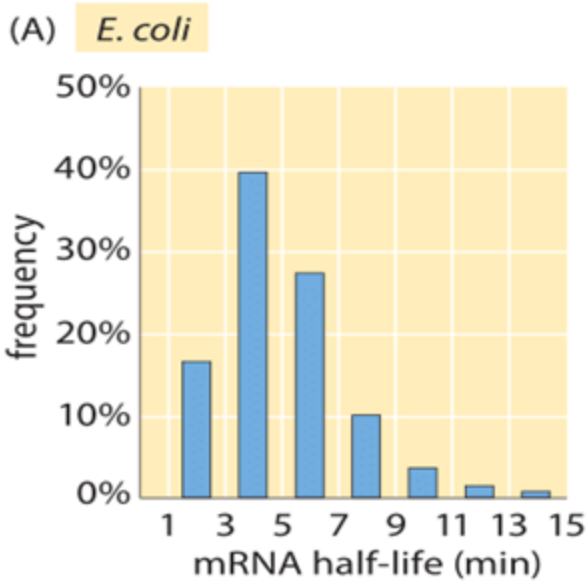
- Fluctuations in mRNAs pools provide a highly sensitive bioassay for environmental signals that are relevant to microbes
- Reveal which microbes perceive specific environmental cues/changes and the metabolic pathways they invoke to respond to it
- Explore the prevalence of particular functions under specific environmental conditions
- Examine ecologically relevant processes of microbes, including how microbes are shaped by and shape environments and how microbial activities influence each other



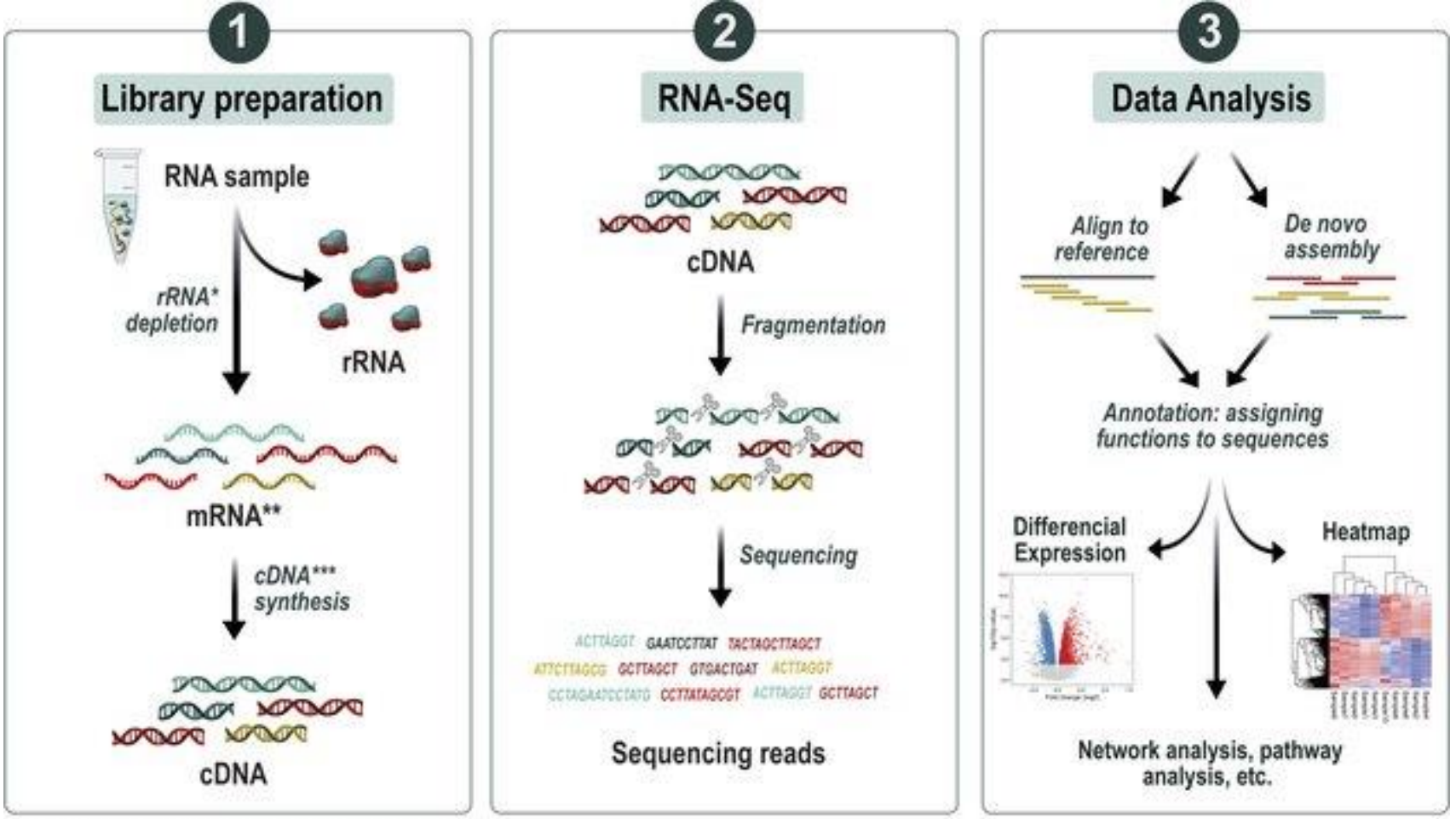
Metatranscriptomes are messy!

Challenges and caveats: collection & library prep

- mRNA degradation
  - mRNA has a short half-life within a few minutes
  - RNAses are everywhere
  - Low biomass samples are easily contaminated
- Most of the RNA in the cell is ribosomal RNA
    - Depletion mechanisms in lab and computationally



J. A. Bernstein et al., 2002



* ribosomal RNA (rRNA): RNA that forms the core of ribosomes.
** messenger RNA (mRNA): carries genetic information from DNA to the ribosome.
*** complementary DNA (cDNA): synthesized from an mRNA template by means of reverse transcriptase & DNA polymerase.

Essner et al., 2024

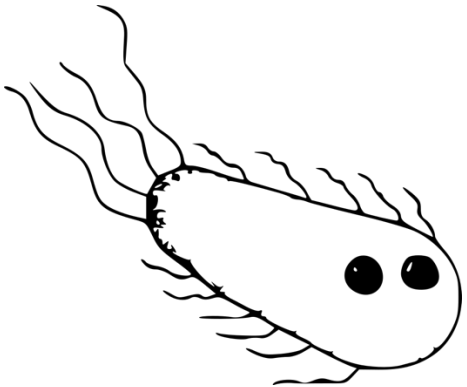
Challenges and caveats: interpretation & analysis

- Transcript abundance is not a reliable rate proxy for functions (e.g. denitrification) and the abundance of proteins.
- Quantification
  - Internal standards (Gifford et al., 2011; Satinsky et al., 2013)
    - >2.7 × 10^9 transcripts per liter of ammonia transporter *amt*
  - Differential gene expression
    - Normalization approaches

$$Copies_i\, L^{-1} = reads_i \times \frac{total\ reads \times \frac{spike\ copies}{spike\ reads}}{total\ reads} \times \frac{1}{volume\ filtered}$$

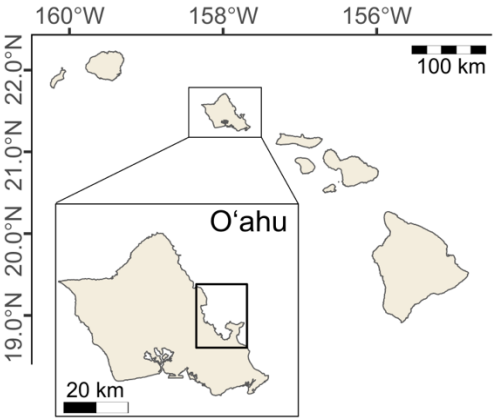
Challenges and caveats: interpretation & analysis

- Functional annotation of transcripts- many unassigned gene functions
  - 52 and 79 % of the average bacterial coding genes could be functionally annotated based on protein and domain-based homology searches, respectively (Lobb et al., 2020)
- Taxonomic annotation of transcripts
  - Mapping of transcripts to reference genomes or *de novo* assemblies

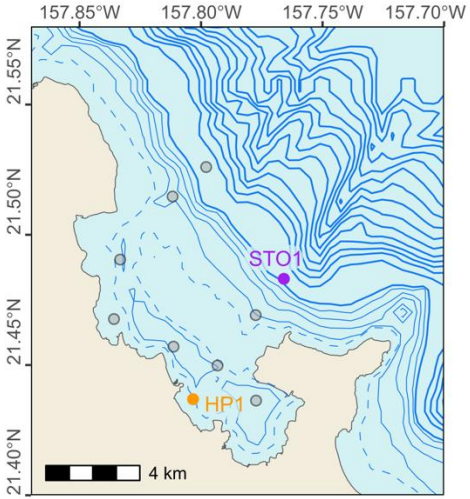


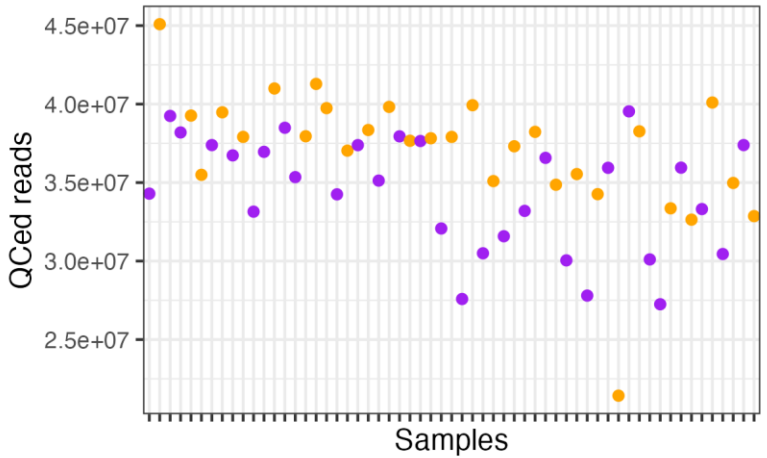
Exercise: Metatranscriptomics of *Synechococcus* in HaDS

- Hawai‘i Diel Sampling (HaDS)
  - multi-omics every 90 min for 48 hours: metagenomes, metatranscriptomes, tRNA-sequences
  - Metatranscriptomes were QCed- FastQC (Andrews, 2010), cutadapt (Martin 2011), SortMeRNA (Kopylova et al., 2012), iu-filter-quality-minoche (Eren et al., 2013)



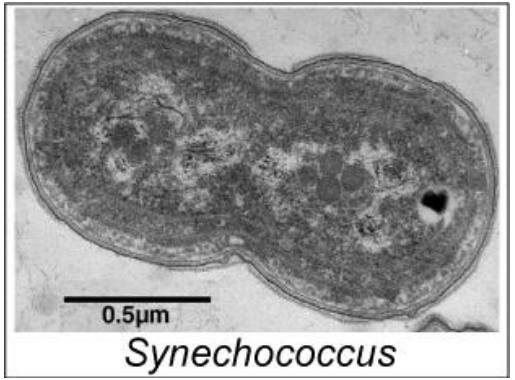
Tucker et al., in revision

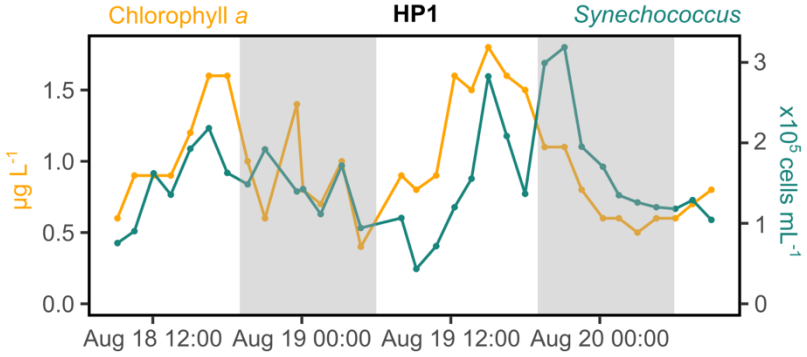




Exercise: Metatranscriptomics of *Synechococcus*

- Metatranscriptomic reads were recruited to a reference genome of *Synechococcus*
- *Synechococcus* is an abundant Cyanobacteria that shows highly diel patterns of abundance

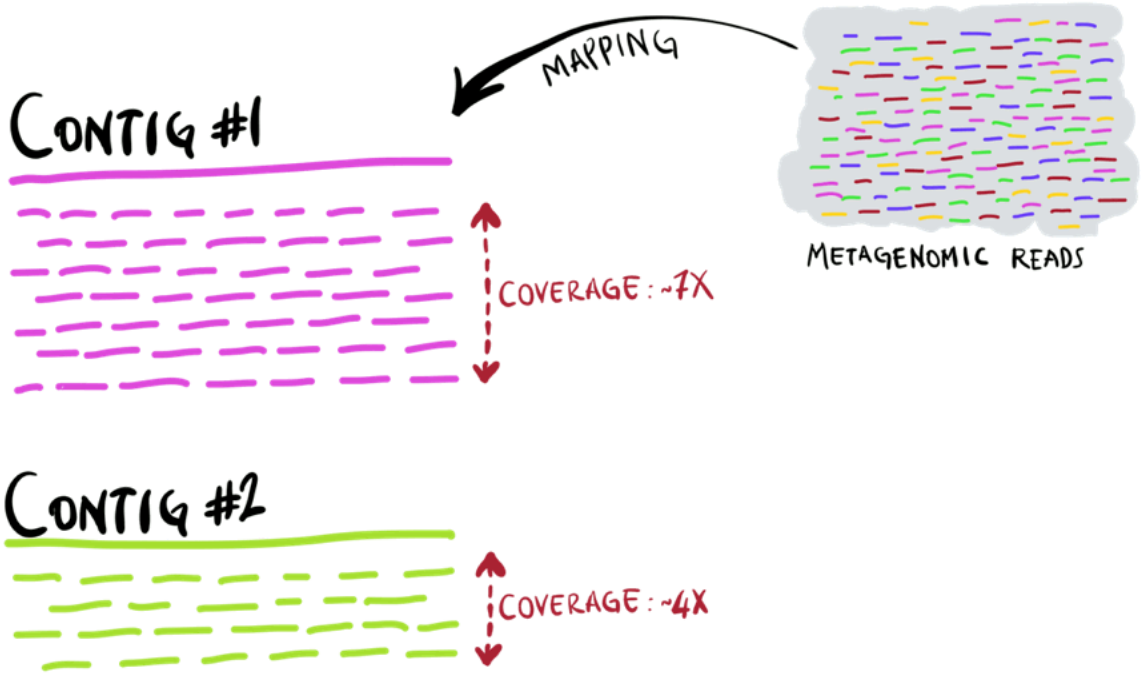




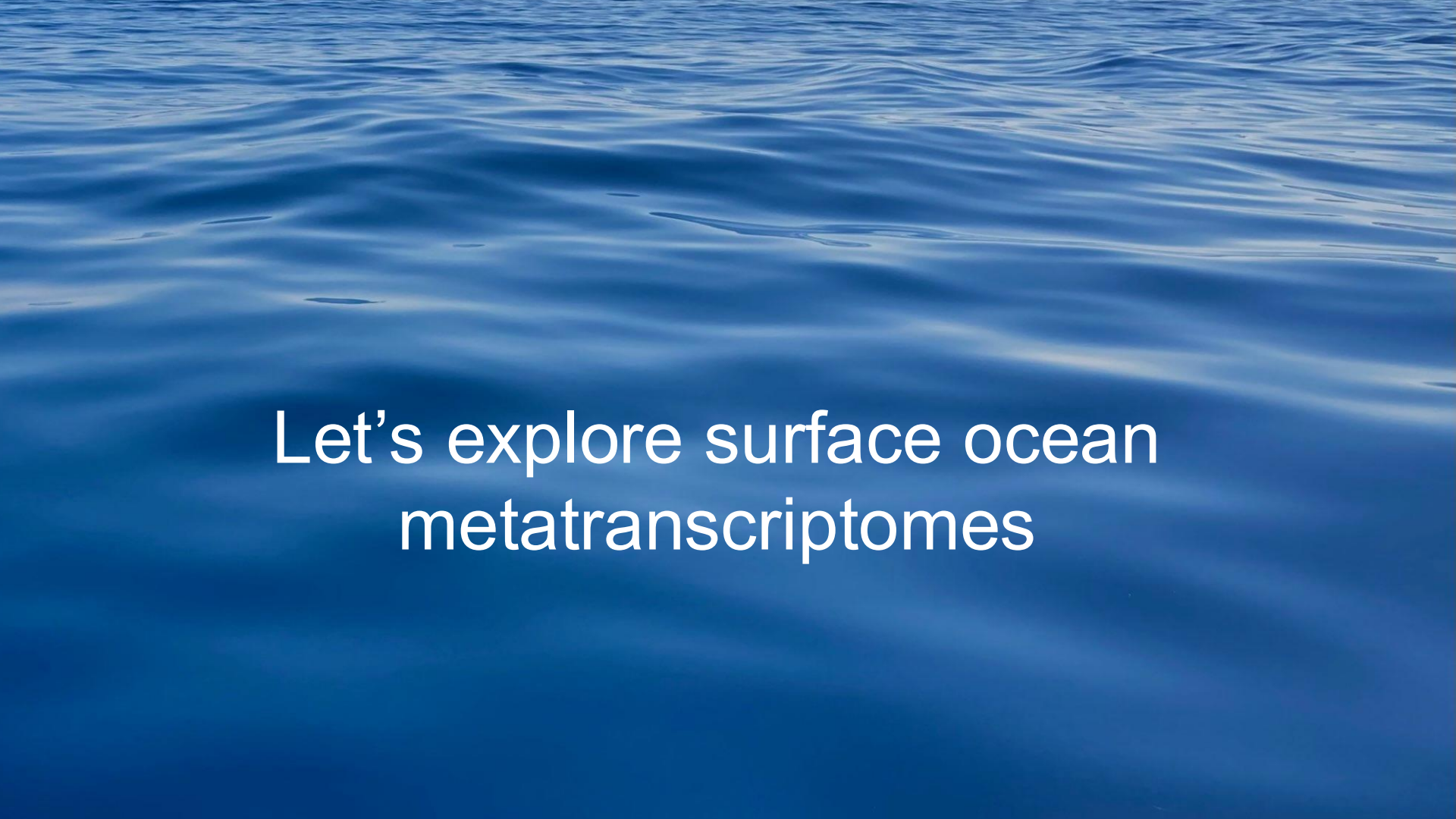
Quick primer on analyses done already

- Working w/ HaDS metatranscriptomes
  - 3 day and 3 night samples
- *Synechococcus* reference genome to "contig.db" in anvi'o v 8.0 (Eren et al., 2021)
  - Genes were called with Prodigal (Hyatt et al., 2010)
  - Genes annotation with NCBI's COGs (Tatusov et al., 2003) and a customized HMM database KOfams (Aramaki et al., 2020)
- Metatranscriptomic and metagenomic read recruitment to the reference genome with Bowtie 2 (Langmead & Salzberg, 2012)
- anvi-profile function stored coverage and detection statistics

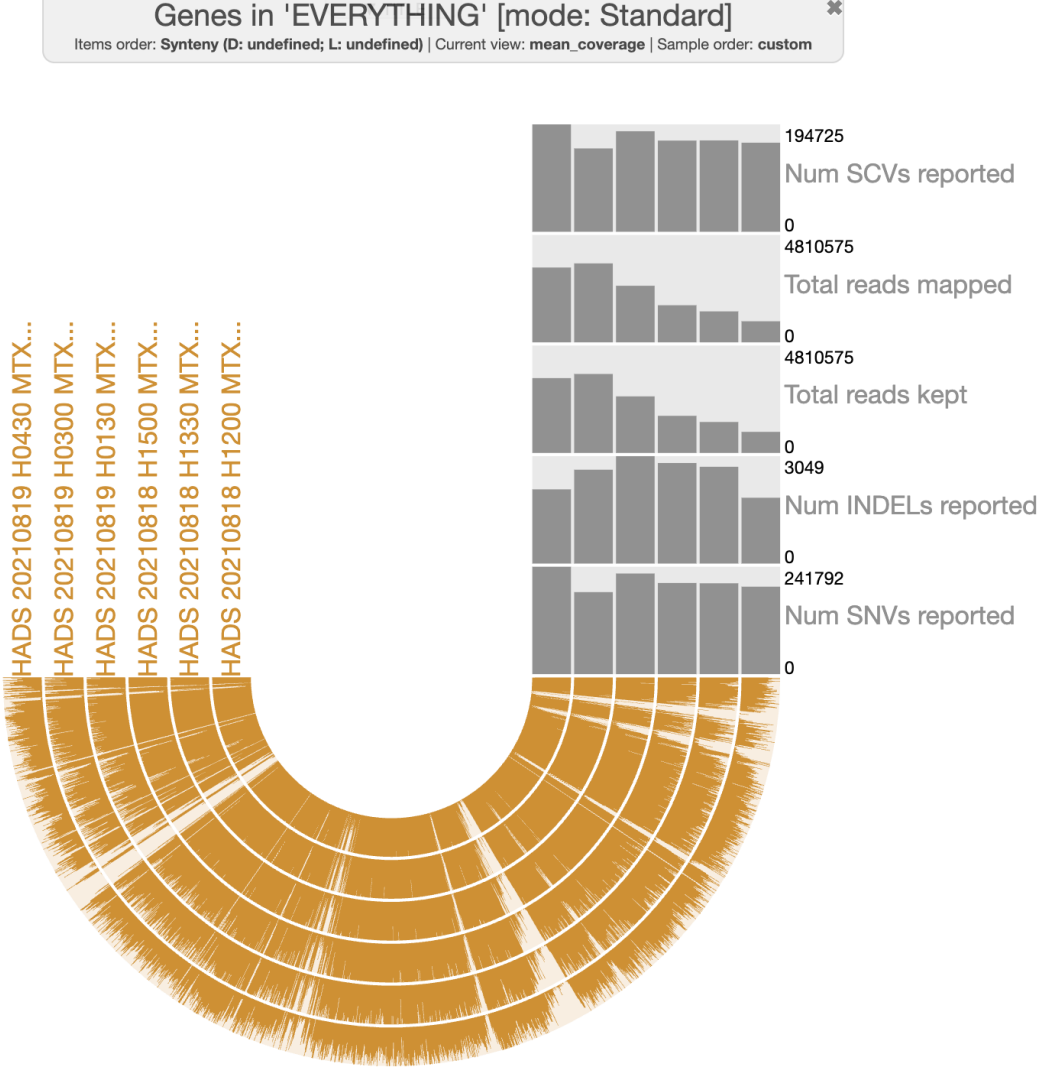
Coverage

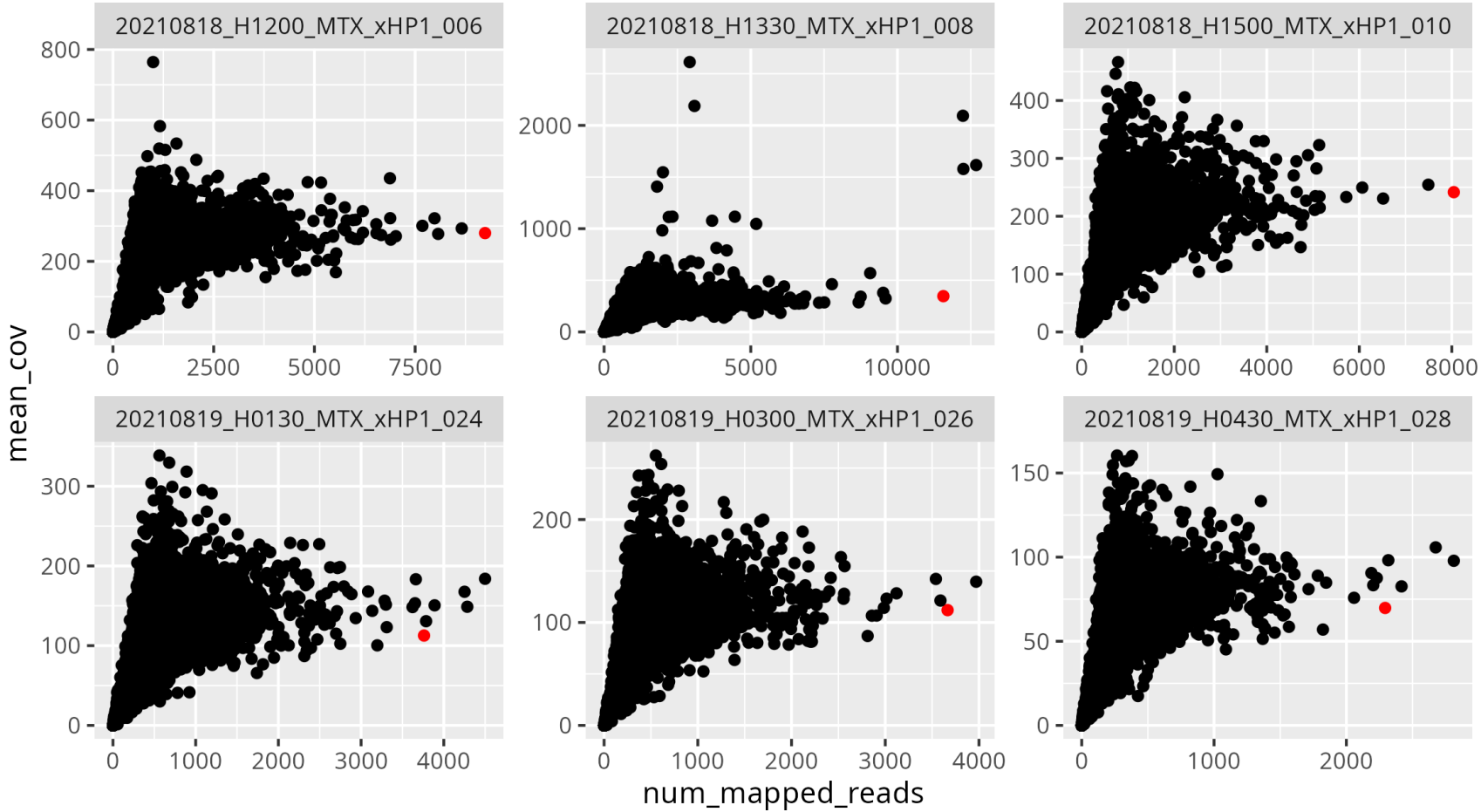


Forgaty & Moore
https://merenlab.org/2019/11/25/visualizing-coverages/


Let’s explore surface ocean
metatranscriptomes

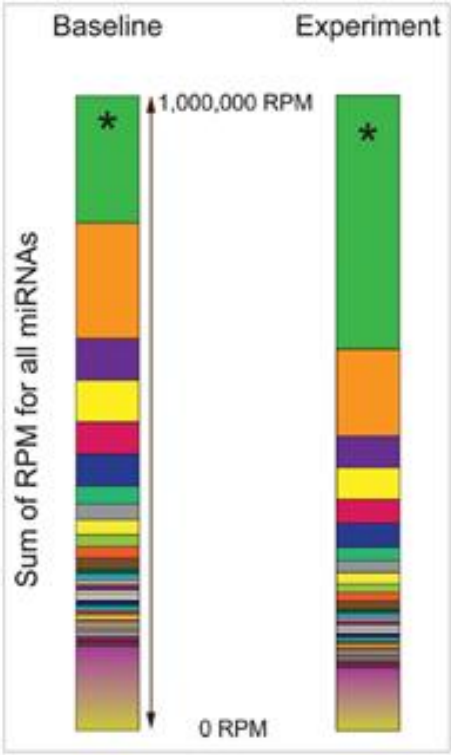
What are some considerations for analyzing metatranscriptomes?




20210818_H1200_MTX_xHP1_006
20210818_H1330_MTX_xHP1_008
20210818_H1500_MTX_xHP1_010
20210819_H0130_MTX_xHP1_024
20210819_H0300_MTX_xHP1_026
20210819_H0430_MTX_xHP1_028
mean_cov
num_mapped_reads

Normalization of metatranscriptomes

- Normalization approaches account for sequence library and differences in contig length, such as Transcripts Per Million (TPM)
  - counts per length of transcript (kb) per million reads mapped
  - **portions can be misleading**
  - RPKM & FPKM have similar concerns

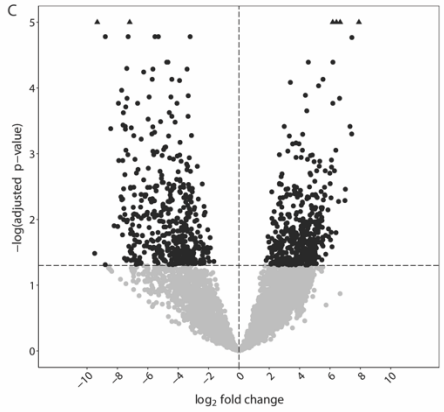


$$\text{TPM} = 10^6 * \frac{reads\ mapped\ to\ transcript\ /\ transcript\ length}{Sum\ (reads\ mapped\ to\ transcript\ /\ transcript\ length)}$$

Witwer & Halushka 2016

Normalization of metatranscriptomes

- RNA-Seq normalization and differential expression tool DESeq2
  - **Count-based- no correction for gene length**, use only for between sample comparison with **similar communities**
  - Median of ratios method normalization
  - Estimate differential expression by fitting generalized linear models and assuming negative binomial distributions
  - High type 1 error rates



Cohen et al., 2022

Using paired MGX + MTX

- Imagine a scenario where a lowly abundant taxa is highly active
  - You may undervalue the transcriptional activity of this taxa if you do not account for its abundance
- Normalizing gene RNA abundance by DNA abundance
  - Salazaar et al., 2019- normalize gene transcript counts by **the read counts** of house keeping genes in MGX, followed variance stabilization in DESeq2
  - Huber et al., 2019 & Zhang et al., 2021- read counts of MTX / read counts of MGX per gene and normalized to **total read depth**

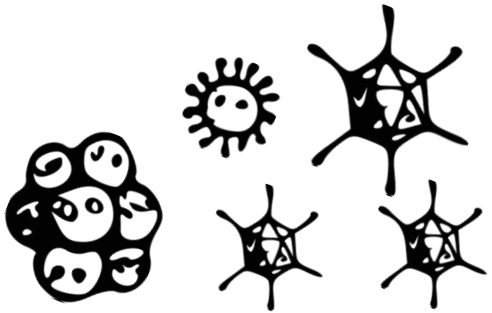
Normalization of metatranscriptomes

clr (MTX coverage per gene) - average clr (scg MGX coverage across samples)

- An estimate of how expression of a gene changes across samples while accounting for changes in abundance and community composition
- Advantages:
  - accounts for “methods specific” variation like sequencing depth
  - does not rely on proportions or counts
  - accounts for shifts in community composition by using a single copy core genes

Metagenome coverage of single-copy core genes (SCG) across samples

- Metagenomic co-assembly of our samples & assigned taxonomy using SCGs => gives us a reference database of taxa in metagenomes
- Read recruitment to the co-assembly and estimate abundance of SCGs across taxonomic groups => estimation of community composition per sample
- Transform the abundance information of the mean coverage per SCG => clr normalized community composition per sample



Normalization of metatranscriptomes

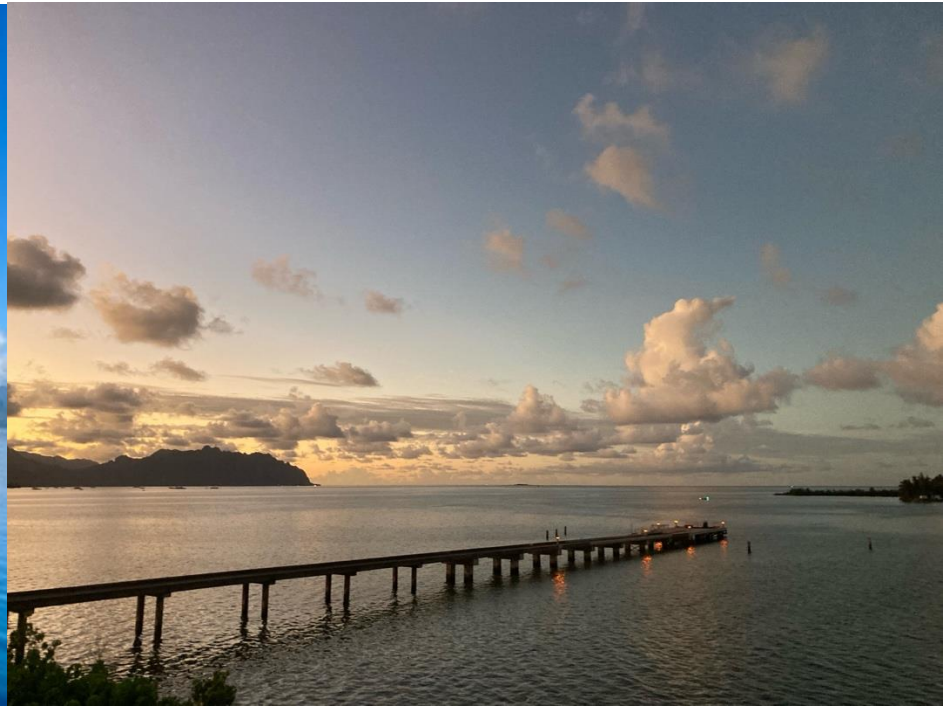
clr (MTX coverage per gene) - average clr (scg MGX coverage across samples)

- An estimate of how expression of a gene changes across samples while accounting for changes in abundance and community composition
- Advantages:
  - accounts for “methods specific” variation like sequencing depth
  - does not rely on proportions or counts
  - accounts for shifts in community composition by using a single copy core genes

Quick primer on analyses done already

- *Synechococcus* reference genome to "contig.db" in anvi'o v 8.0 (Eren et al., 2021)
- MTX and MGX read recruitment to the reference genome with Bowtie 2 and anvi-blitz exported coverage of genes in MTX & MGX samples
- anvi-estimate-scg provided community composition estimates from reads mapped to a co-assembly of the HaDS metagenomes
- clr (MTX coverage per gene) - average clr (scg MGX coverage across samples)
- *Synechococcus* gene expression relative to its abundance

How does the gene expression of *Synechococcus* change from day to night within coastal Kāneʻohe Bay?



Thank you!

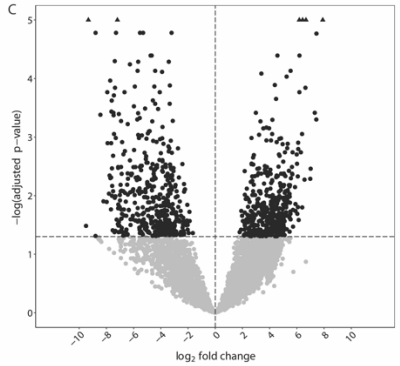
What sorts of questions can you address with metatranscriptomes?

- How similar is transcription across samples?
- Does taxa X experience stress Y across a transect?
- Which genes are differentially expressed between treatments?
- How does transcription of taxa X associate with transcription of taxa Y?

What sorts of questions can you address with metatranscriptomes?

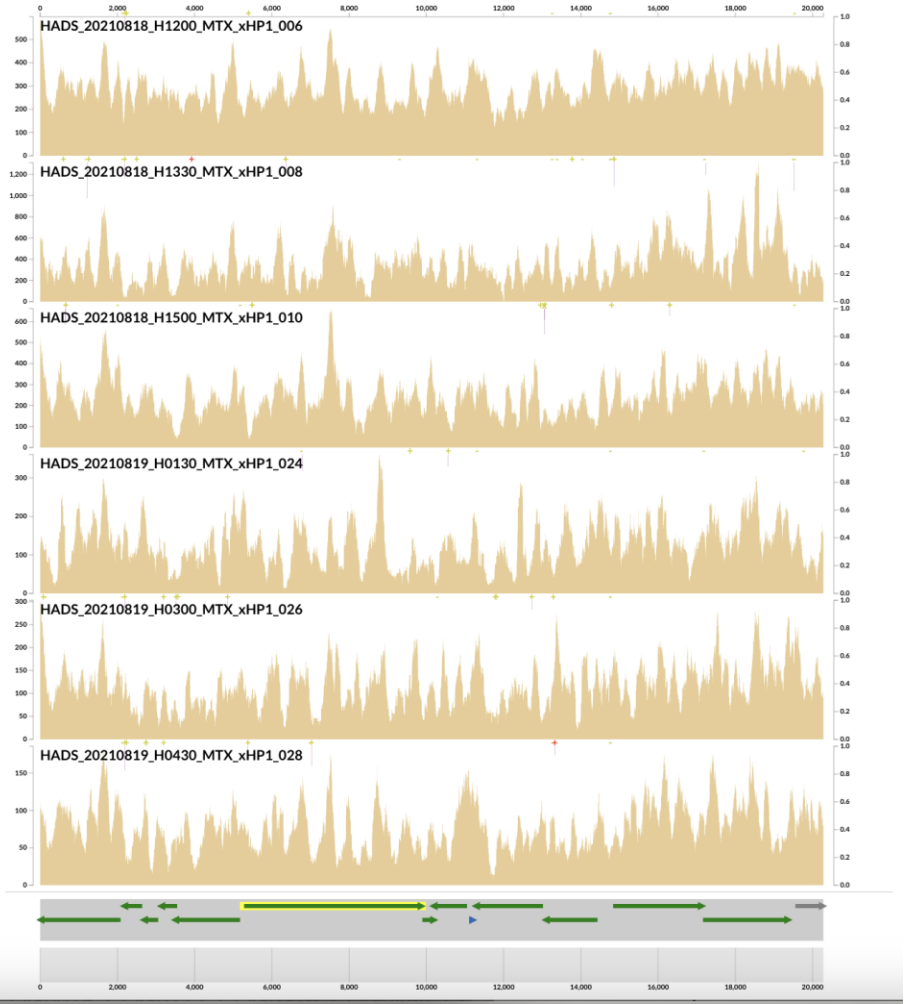
- How similar is transcription across samples? - **ordination**
- Does taxa X experience nutrient stress across a transect? - **heatmaps**
- Which genes are differentially expressed between treatments? – **volcano plots**
- How does transcription of taxa X associate with transcription of taxa Y? - **networks**

Cohen et al., 2022




HADS_20210819_H0000_MTX_xHP1_022
HADS_20210819_H0130_MTX_xHP1_024
HADS_20210819_H0300_MTX_xHP1_026
0
2,000
4,000
6,000
8,000
10,000
12,000
14,000
16,000
18,000
20,000


Questions?


0
2,000
4,000
6,000
8,000
10,000
12,000
14,000
16,000
18,000
20,000
HADS_20210818_H1200_MTX_xHP1_006
HADS_20210818_H1330_MTX_xHP1_008
HADS_20210818_H1500_MTX_xHP1_010
HADS_20210819_H0130_MTX_xHP1_024
HADS_20210819_H0300_MTX_xHP1_026
HADS_20210819_H0430_MTX_xHP1_028