Mapping and Modelling the Ocean’s Unseen Biodiversity

Martin Ostrowski
Deepa Varkey, Amaranta Focardi & Ian Paulsen
Mark Brown
Jodie Van de Kamp, Andrew Bissett, Lev Bodrossy
Justin Seymour
1. Indicators of processes (species, symbioses, functional genes, biogeochemical provinces)

2. Sensitive stress indicators

3. Metabolic models => C and energy flow

4. Resolution of ‘living’ provinces in time and space

5. Computational approaches and capacity are available

6. Remote instrumentation, molecular technology…

**Why should microbes be an important component of models?**

---

**Prochlorococcus**

0.6 µm dia.

$10^6$ cells.ml

$10^{29}$ cells

1,700 genes
<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>N</th>
<th>P</th>
<th>S</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Synechococcus</strong></td>
<td>↓ CO₂, aa</td>
<td>NO₃ ⇄ NH₄</td>
<td>PO₄⁻</td>
<td>Sulfolipids</td>
<td>Toxins, hypoxia</td>
</tr>
<tr>
<td><strong>Prochlorococcus</strong></td>
<td>↓ CO₂, aa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Alphaproteobacteria</strong></td>
<td>↑ CO₂, aa</td>
<td>↓ NO₂</td>
<td>C-P</td>
<td>DMSP ⇄ DMS</td>
<td>SO₄²⁻, H₂S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pathogens</td>
</tr>
<tr>
<td><strong>Diazotrophs</strong></td>
<td>↓ CO₂, aa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Archaea</strong></td>
<td>↑ ↓ CH₄</td>
<td></td>
<td></td>
<td></td>
<td>Anamox</td>
</tr>
<tr>
<td><strong>Phytoplankton</strong> (plastid)</td>
<td>↓ ↓ CO₂</td>
<td></td>
<td></td>
<td>↑ DMSP</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HABs</td>
</tr>
<tr>
<td><strong>Viruses and Phage</strong></td>
<td>↓ ↑ ↑ CO₂</td>
<td></td>
<td>PO₄⁻</td>
<td></td>
<td>Pathogens</td>
</tr>
</tbody>
</table>

**Diagram:**
- **Synechococcus** processes CO₂ and amino acids (aa) and uses NO₃ to convert to NH₄.
- **Prochlorococcus** uses CO₂ and amino acids.
- **Alphaproteobacteria** uses CO₂ and amino acids, converting NO₂ to C-P and processes DMSP to DMS.
- **Diazotrophs** fix N using CO₂ and amino acids.
- **Archaea** uses CH₄ for processes.
- **Phytoplankton** (plastid) uses CO₂.
- **Viruses and Phage** use CO₂.
The Australian Microbiome Initiative

**Tags**
- amplicons (11104)
- Pelagic_Public (9019)
- 16S (5068)
- A16S (3153)
- 18S (2962)
- Sediment (818)
- raw (684)
- Coral (649)
- metagenomics (570)
- Coastal water (420)

**Show More Tags**

**Formats**
- FASTQ (11787)

---

11,788 datasets found for "Marine"

**Order by:** Relevance

---

**Marine Microbes Metatranscriptome 102.100.100.58176**

Google Earth reference BP18, seagrass leaves, sample sent for transcriptomics (NS samples)

[FASTQ](#)

---

**Marine Microbes Metatranscriptome 102.100.100.58175**

Google Earth reference BP18, seagrass leaves, sample sent for transcriptomics (NS samples)

[FASTQ](#)

---

**Marine Microbes Metatranscriptome 102.100.100.58174**

Google Earth reference BP18, seagrass leaves, sample sent for transcriptomics (NS samples)

[FASTQ](#)

---

**Marine Microbes Metatranscriptome 102.100.100.58149**

Google Earth reference BP16, seagrass leaves, sample sent for transcriptomics (NS samples)

[FASTQ](#)
Data is immediately available
(as soon as it is sequenced, QAQC’d and processed)

Enhanced access

The Australian Environmental Microbiome Research Data Cloud

We're aiming to make large and complex continental scale environmental metagenomics data more accessible to researchers through the establishment of a cloud-based data analysis system.

ARDC funded Bio–RDC project

- Researcher’s Requirements
- Gathered at BPA coordinated Marine Microbes Meeting March 2018

Functionality to analyse OTU data across sites/populations:

- Ordination
- Clustering
- Rarefaction
- Search by Sequence Similarity
- Plots (Box, Bar, Venn etc)
- Maps (bubbleplots, heatmaps)
- Browsing Taxonomic abundance data
- BYO OTU data
Microbial communities in time and space

7 NRS
4 New Coastal programs
1 billion observations
Actual Sequence Variants (zOTUs)
~175,000 bacterial-
~350,000 eukaryote-
56,000,000 unique functional genes

570 metagenomes
3.2 TB data
Hierarchical clustering

>0.83 correlation between clustering methods

19 clusters, \( n = 1,341 \)

Cyanobacteria, phytoplankton (chloroplasts), SAR11, Roseobacter…
1. Latitudinal and seasonal patterns in key functional groups
   • Microbial primary producers prominent at all NRS
   • *Synechococcus* prominent => dominant in subtropical NRS
   • Late winter blooms in microbial eukaryotes at MAI
   • Defined seasonal cycles in N-fixation, Roseobacter (S-cycling)
   • ROT significantly different than East Coast NRS

2. Microbial communities are highly dynamic in time
Modelling 2: Hindcasting and the trajectory of SE microbial communities

Boosted Regression Trees
10x holdout cross validation
8-12 predictor variables
~3,500 zOTUs (>79% of abundance data)
3,382 samples → 119,282 community profiles
PH, MAI, NSI → ROT > 0.75 model.cv.mean
Model correlation mean > 0.85

Hindcast models at NRS 1946-> nMDS of reconstructed communities

Brown & Ostrowski, in prep
Australian Microbiome
2018-2021

Sustain ° Enhance ° Engage

New NRS-style coastal reference sites (water, sediment)

Qld, NSW, Tas, Vic, SA
• Partnership between IMOS, MNF, Bioplatforms Australia, Research Universities and Institutes has accelerated progress

• a comprehensive baseline of microbial diversity linked to a rich set of phys-chem-biol contextual data

• Genomic data provides detailed new dimensions on the function and future of marine systems

• Research will play a key role in identifying relevant molecular indicators for ‘real-time’ and remote monitoring
Ian Paulsen, Amaranta Focardi, Deepa Varkey

Mark Brown, Martin Ostrowski, Lev Bodrossy

Justin Seymour, Martina Doblin

The Marine Microbes consortium
Jodie van de Kamp, Andrew Bissett, Eric Raes, Kirianne Goossen, Swan Li-San Sow

The Australian Microbiome
Modelling 1: Boosted Regression Trees, Machine Learning

**Boosted Regression Trees**

Elith 2008

3,300 zOTU

- Day length
- Temp (°C)
- Salinity
- Depth
- Nitrate
- Stratification
- N:P
- Phosphate
- Silicate

Model uses
- Identify gaps
- Observe trends (hindcasting)
- Speculative reconstruction

Work towards quantitative models (Carbon, Energy, Ecosystem stability)

**Time series**

Reconstruct community

Modelled cyanobacterial community at TS MAI PHB 1946–

10x holdout cross validation
Independent validation
Predictions for each zOTU
Models vs Independent evaluation dataset
CARS data at 0.5° Resolution
Community predictions based on top 50 zOTU (~40% of sequences in the total dataset)
Community clusters identified by simprof|mclust|k-means
Follow-up: detailed characterisation of ‘microbial-biogeochemical provinces’

1. Models identify a number of community transitions across this latitudinal gradient
2. Tasman sea off East Tasmania is characterised by a seasonal regime switch from May-Jun and Dec-Jan
Total Samples: 5,068
Bacterial
Archaea
Eukaryotes

<table>
<thead>
<tr>
<th></th>
<th>2010</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
<th># samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centaur Reef</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65</td>
</tr>
<tr>
<td>Channel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Coral</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>433</td>
</tr>
<tr>
<td>Darwin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>72</td>
</tr>
<tr>
<td>Geoffrey Bay</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>IN16v03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>53</td>
</tr>
<tr>
<td>KAI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>365</td>
</tr>
<tr>
<td>MAI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>318</td>
</tr>
<tr>
<td>NSI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>324</td>
</tr>
<tr>
<td>PHB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Pioneer Bay</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>138</td>
</tr>
<tr>
<td>ROT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>177</td>
</tr>
<tr>
<td>Seagrass</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>216</td>
</tr>
<tr>
<td>Seaweed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>419</td>
</tr>
<tr>
<td>Sediment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>218</td>
</tr>
<tr>
<td>Sponge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3,327</td>
</tr>
<tr>
<td>Voyage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
</tr>
<tr>
<td>Wreck Rock</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>97</td>
</tr>
</tbody>
</table>

Total Samples: 5,068
100,000
Cells per millilitre of seawater

$10^{29}$
Cells in the ocean

4,000 X
Total biomass is 4,000 X total whale biomass

1,751,080
Base pairs in genome

0.6 x 0.8 µm in diameter

80%
up to 80% of the Prochlorococcus in the East Australian Current are lysed by phage everyday

600-2,500
Ribosomes per cell

Waldbauer et al., 2012
Dufresne, et al., 2003
$\sim 10^6$ Bacteria per ml.
$\sim 10^7$ Virus per ml.

The 1977 revolution
Regulatory networks
Phenotypic adaptations
Biophysical characterisation
of nutrient transport systems
cellular models

microbial C-stocks
and flux
N,S-Cycle dynamics
Viral turnover of
phototrophic communities
Functional genomics
=> ecosystem models

Functional Genomics

Biological Carbon Pump
Viral shunt
DOM, POC
Aggragate - sinking particle
Microbial loop
CO₂
Beyond Genomics:
molecules to global models

CO₂ → DOM, POC

Aggregates → sinking particle

Viral shunt

Biological Carbon Pump

DOM, POC

Microbial loop