Investigating Marine Microorganisms' Metagenomes and Single Cell Genomes

Ashley Bell[1], Joanna Warwick-Dugdale[1,2] and Ben Temperton[1]
[1] Plymouth Marine Laboratory, Plymouth, Devon, United Kingdom
[2] School of Biosciences, University of Exeter, Exeter, Devon, United Kingdom

Intro

- 90% of the ocean's biomass comprise of marine microorganisms; their impact on global systems remain largely understudied.
- Marine microorganisms are critical in the energy cycle and are the foundation for marine life.
- Studying these microorganisms remains challenging with a small fraction being culturable for in situ experimentation.
- Alternative study methods include obtaining genomes via metagenomics studies and Single-cell Amplified Genomes.
- Both of these methods have advantages and drawbacks with marine metagenomes being highly complex to analyse.
- SAGs suffering from both low coverage and bias.

Marine Metagenomics

- Metagenomes are all genetic material recovered from environmental samples
- Used to discover genomic content and population dynamics
- Do cellular marine lysogenic virus differ temporally and spatially?
- Samples from Bermuda Atlantic Time Series (BATS) by BIOS (Bermuda Institute for Oceanic Sciences)
- Separation of viral fraction from cellular content by filtration

Method

- Read quality control
- Metagenomic assembly
- Coverage vs other samples
- Visualisation, kmer counting and t-SNE calculation
- Automated binning software via clustering and coverage
- Extraction and annotation of cellular sequence
- Extraction of Viral signatures from contigs

t-SNE plots provide an alternative to separating metagenomes

SCAN ME

Results

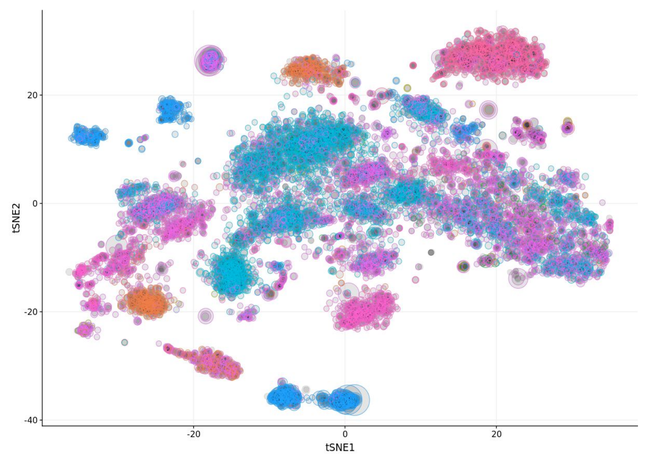

Fig 1. Scatterplot of a marine cellular sample overlayed with phylogenetics in colour

Discussion

- Metagenomes allow for population dynamic studies across space and time
- Recovery of all genomes is difficult, with completion at ~80%, and high amounts of contamination
- Manual binning is time consuming
- Automated binning software is less effective

Intro

- Genomes from isolated single cells are that whole genome amplified and sequenced
- Organisms usually only have 1 to a few copies of their genomic DNA. Not enough to sequence so need to amplify with current sequencing tech
- Are Single-cell Amplified Genomes (SAGs) a good way to study genome content?
- 451 SAG SAR11 genomes isolated and sequenced

Method

- SAG reads assembled using single cell assembler
- Phylogenetic tree constructed based on gene content
- Average nucleotide identity calculated
- Completeness and contamination calculated by gene content
- Mapping against metagenomes for presence absence and variable regions

Multiple SAGs allow for effective clade wide genetic studies

SCAN ME

Results

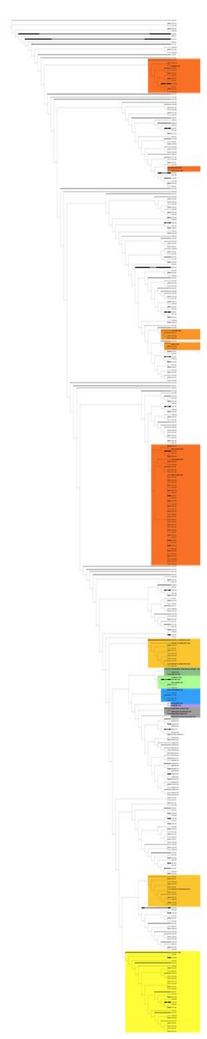

Fig 2. Phylogenetic tree of SAR11 SAGs. Coloured tabs show existing groups

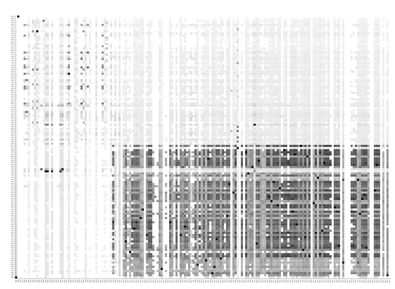

Fig 3. Average Nucleotide Identity heatmap of SAGs show clustering indicating differing populations based on their genomes

Discussion

- Wildly differing completeness due to PCR bias (4-97%)
- ANI and phylogenetic tree indicate possible new clades
- Mapping against metagenomes indicate differing populations and variable region