

Metagenomics Summer School 2023

Day 4

Genome annotation (continued)
Presentation of data
Present and discuss findings



Day overview QC **Assembly** Goals: **Binning** Gene annotations (continued) Visualizing metagenomic data Bin dereplication √ **Group task Bin refinement Bin taxonomy Viruses Gene prediction Contigs and QC Annotation and coverage Coverage and Analysis** taxonomy

Presentation of data





DRAM results overview

DRAM provides a glance of a genome's content:

- Metabolism-based heatmaps (completeness of pathways, presence absence of function)
- Combined annotations for identified functions
- Genome assembly statistics (completeness, contamination, quality)
- rRNA presence and tRNA counts





Task: Presentation of data

Go to Github MGSS webpage

Tasks:

- Explore DRAM results
- Coverage heatmap and nMDS ordination
- KEGG metabolic pathways
 - Build KEGG pathway map of genes related to nitrogen metabolism
- Gene synteny
 - Build a sulfur assimilation gene alignment plot
- CAZy heatmaps
 - Build a heatmap of CAZy annotations



Presentation of data

Finding the answer to your question is only half of the issue

How do we report/present our data?

- 1. Heatmaps of bin and viral contig coverage across samples
- 2. Ordinations to investigate relatedness of samples
- 3. KEGG pathway maps
- 4. Gene synteny analysis
- 5. Heatmaps of genomic features
- 6. Inference of gene trees
- 7. Creating metabolic schematics





Visualisation: Genome distributions

How similar/different are the communities in these samples? How are the organisms distributed across the environment?

Coverage heatmaps

- Shows relative abundance of MAGs across samples
- Reveals
 dominant/rare/potentially
 co-occurring taxa/MAGs
 which may be interesting for
 further analyses

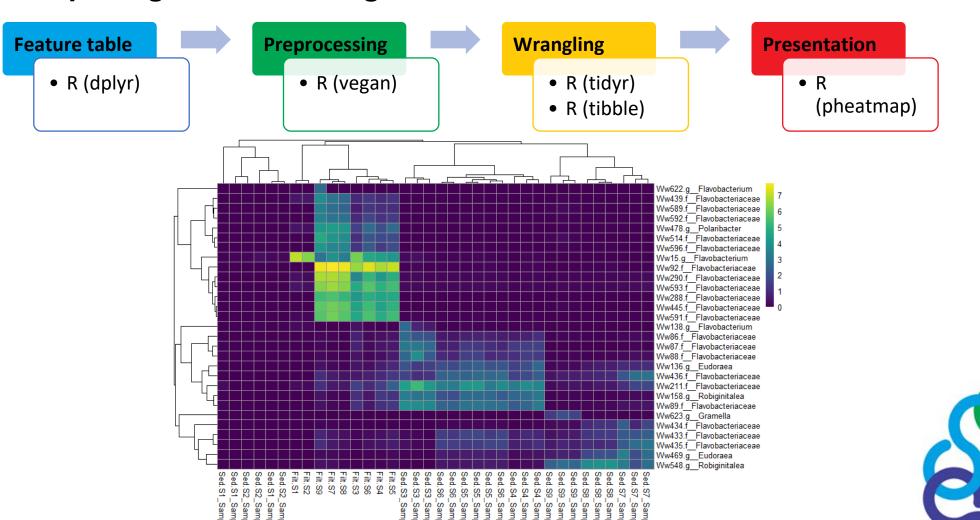
Ordinations

- Shows compositional similarities between samples
- Coverage + Environmental metadata can reveal potential drivers of community assembly and composition
- Data transformations and choice of dissimilarity metric can reveal different facets in compositional changes

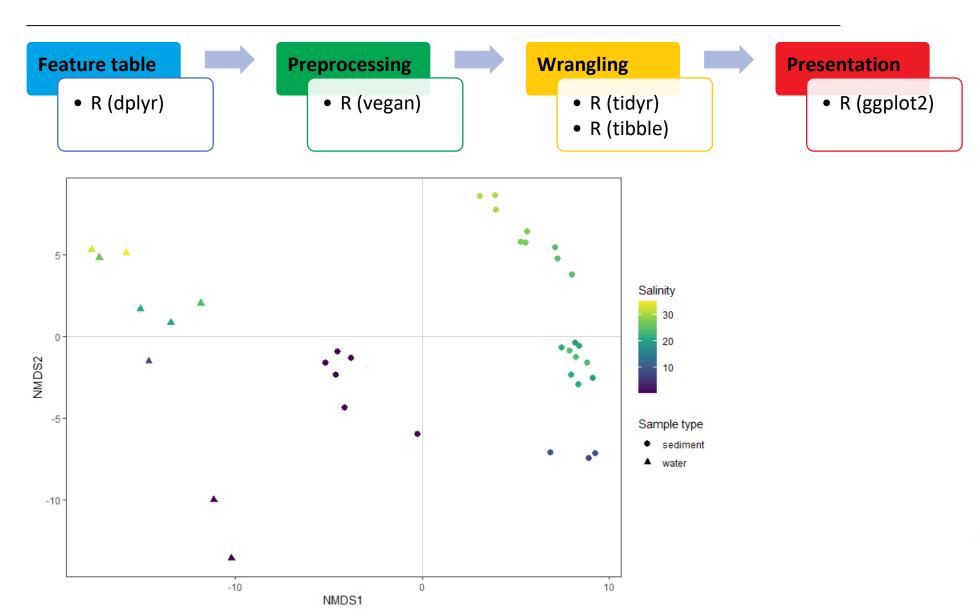


Visualisation: coverage heatmaps

Heatmaps of genome coverage



Visualisation: nMDS ordinations







Task: Presentation of data

Go to Github MGSS webpage

Tasks:

- **/**
- Explore DRAM results
- Coverage heatmap and nMDS ordination
- KEGG metabolic pathways
 - Build KEGG pathway map of genes related to nitrogen metabolism
- Gene synteny
 - Build a sulfur assimilation gene alignment plot
- CAZy heatmaps
 - Build a heatmap of CAZy annotations





Task: Workshop survey



Visualisation: Metabolism

What can this organism do? How does it compare to other organisms?

Pathway maps

- Pathway completeness
- Maps + annotation information can help infer alternative genes/pathways

Gene synteny

- Shows local gene context
- More detailed than reporting gene table
- Sometimes absence of genes from operon is biologically informative

Annotation heatmap

- Presence/absence of function
- Semi-quantitative (number of hits)
- Compare between MAGs



Visualisation: Metabolism

What can this organism do? How does it compare to other organisms?

Gene tree

Phylogenetic relationship of functions

Metabolic schematic

- Shows how gene and their products relate to each other
- Can show multiple organisms to show potential relationships





Visualisation: Pathway maps

KEGG annotation

Wrangling

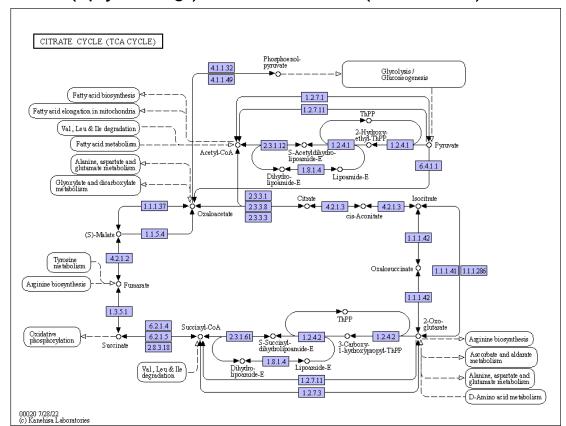
KEGG metadata

Visualisation

- BLAST (or BLAST-like)
- Hidden Markov Model (KoFamScan)

R (dplyr, stringr)

• R (KEGGREST)

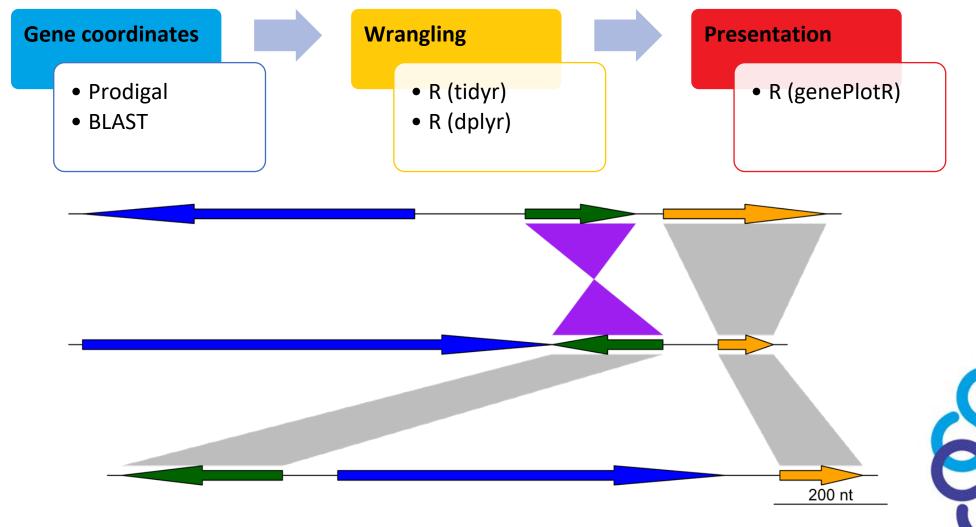


R (pathview)



Visualisation: gene synteny

Gene synteny analysis



Visualisation: functional heatmaps

Heatmaps of genomic features, e.g. carbohydrate metabolism

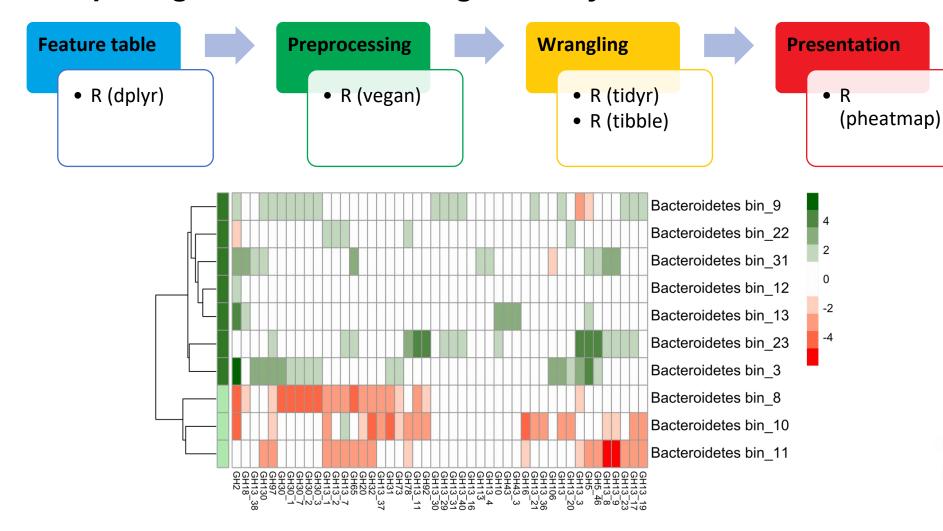
Simple figure to display complex data tables

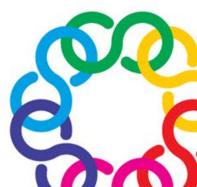
- M genomes x N features in one place
- Presence/absence or relative abundance (multi-copy)
- Fixed layout, or clustering by patterns



Visualisation: functional heatmaps

Heatmaps of genomic features, e.g. carbohydrate metabolism



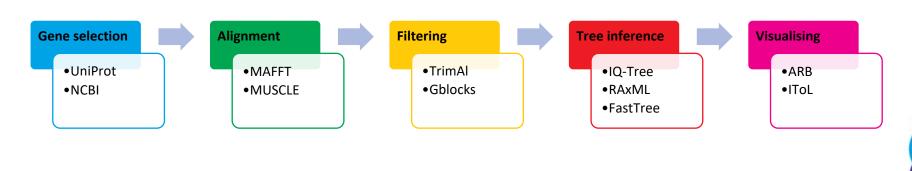


Visualisation: gene trees

Inference of gene trees

Gene trees are a great way to present

- Confirmation of annotation
- Novelty of detection / horizontal gene transfer
- Rate of evolution in the feature

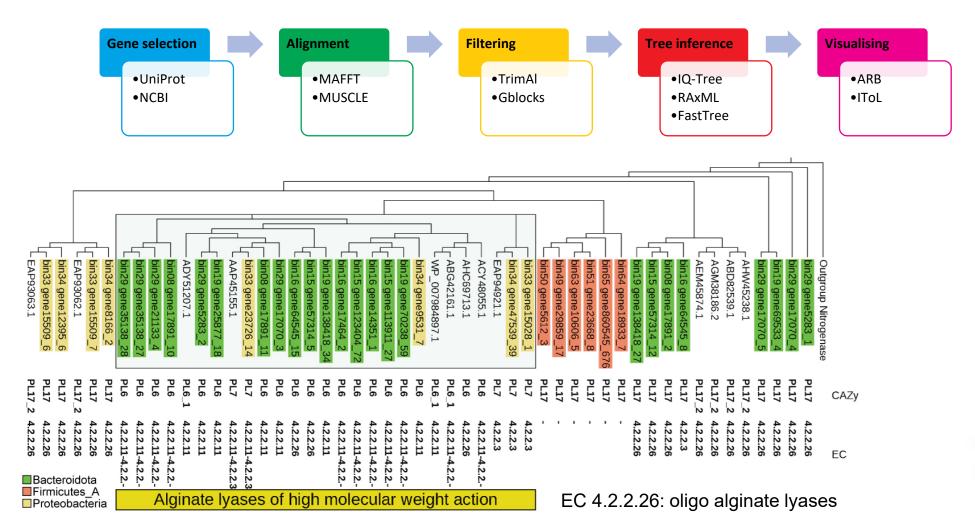






Visualisation: gene trees

Inference of gene trees

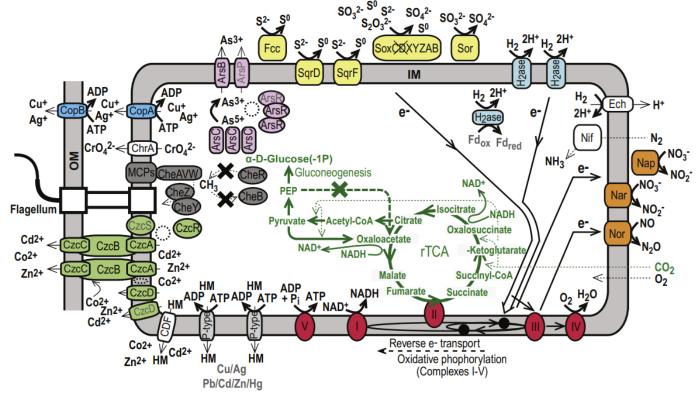




Visualisation: schematics

Creating metabolic schematics

Summarise the entire core metabolism of an organism into a single figure





Handley et al. (2014) doi:10.1111/1462-2920.12453

Visualisation: schematics

Creating metabolic schematics

Summarise the entire core metabolism of an organism into a single figure

What's the magic tool for producing these?

- Illustrator, Inkscape, GIMP (and a lot of time)
- Use tools for picking colour schemes
 - ColorBrewer2 (http://colorbrewer2.org)
 - IWantHue (https://medialab.github.io/iwanthue/)











Task: Presentation of data

Go to Github MGSS webpage

Tasks:

- Explore DRAM results
- Coverage heatmap and nMDS ordination
 - KEGG metabolic pathways
 - Build KEGG pathway map of genes related to nitrogen metabolism
 - Gene synteny
 - Build a sulfur assimilation gene alignment plot
 - CAZy heatmaps
 - Build a heatmap of CAZy annotations





Task: Project

Tasks:

- Analyse data for group work
- Prepare group presentation
 - Use the white board for illustrations
 - **■** Things to include:
 - The attribute you found
 - Details about the attribute
 - The organism(s) you found it in
 - A brief explanation of biological relevance
 - The tools and annotations you used
 - Anything else?



Present and discuss findings





Task: Present and discuss findings

Tasks:

- Each group to give an informal presentation of their data
 - Report your findings with regards to your objective
 - What did you find?
 - How did you make the discoveries?
 - Functional prediction only
 - Functional prediction and taxonomic context
 - Each group present for 5 mins each (max!)



Wrap up and Q/A

