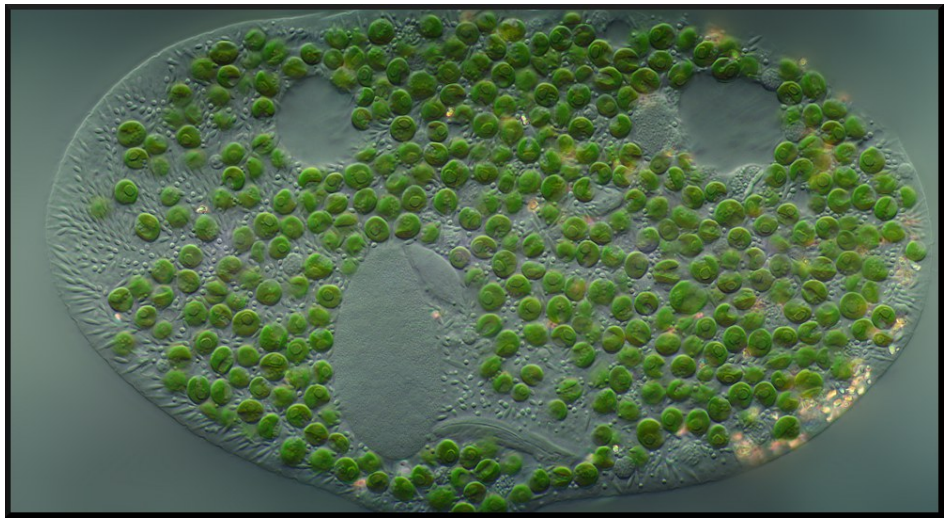
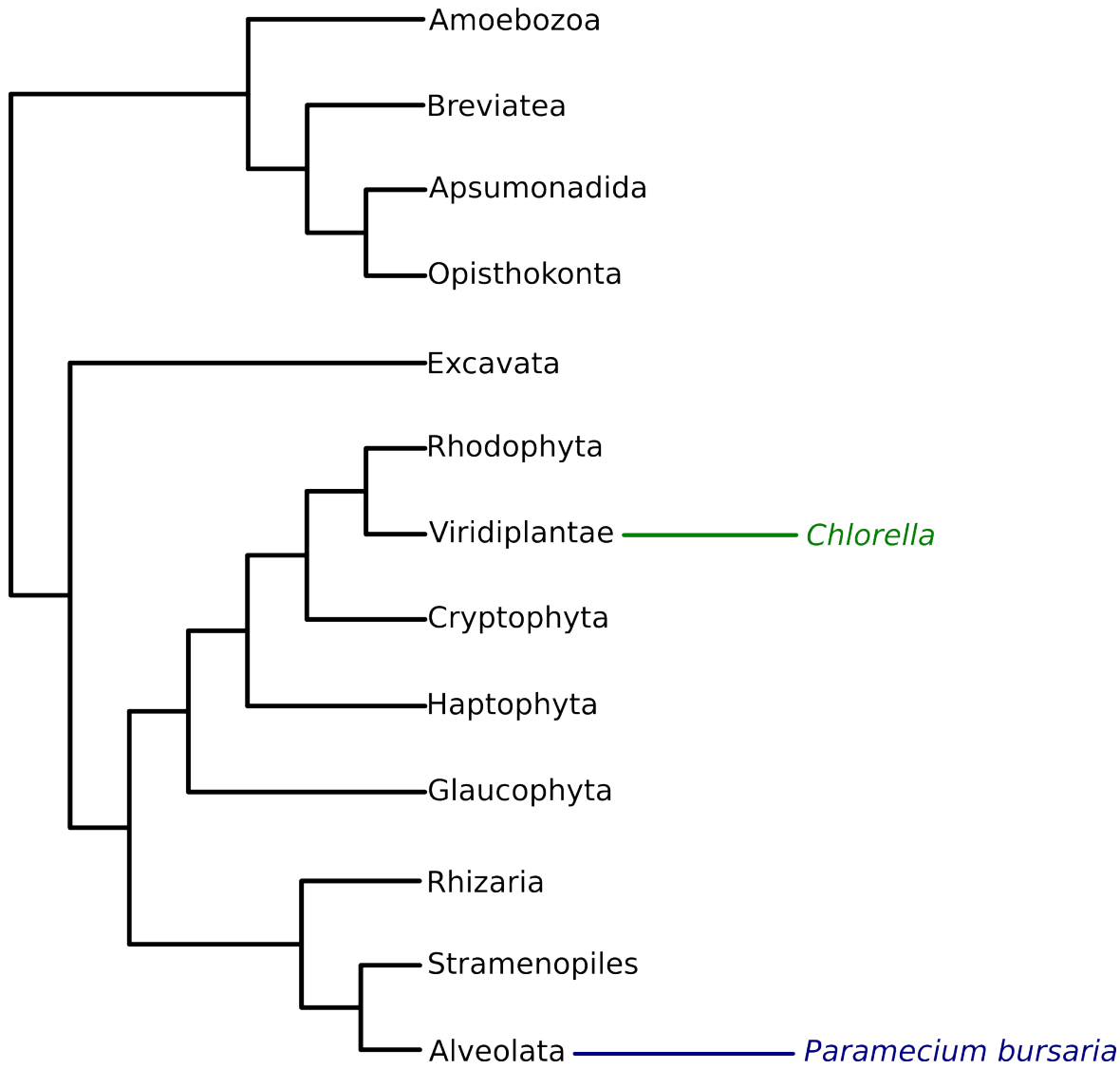
Key proteins implicated in the maintenance of photosynthetic endosymbiosis between *Paramecium bursaria* and *Chlorella*

Finlay Maguire
University College London & Natural History Museum



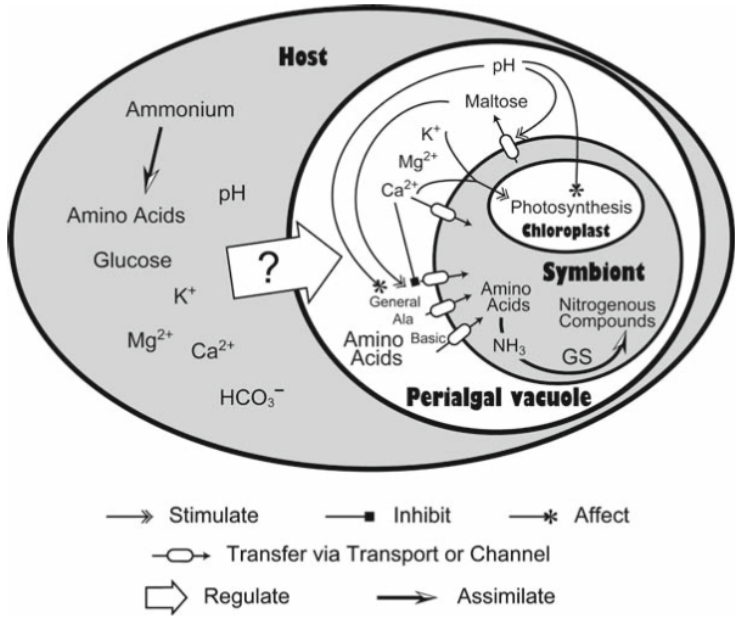
- Background
  - Phylogenetics & Molecular biology
  - Utility as a model organism for evolution of endosymbiosis
- Transcriptomics
  - Day/Night transcriptomes
  - Single cell transcriptomes
  - Assembly
  - Saturation analysis
  - Phylogenomic-SVM binning of transcripts
  - Differential expression analysis
- Metabolic reconstruction
  - Mapping
  - Comparative analysis
  - Predictions
- Further data to collect
  - Carbohydrate metabolism
  - Amino acid metabolism
  - Total metabolites
- Testing of predictions
  - RNAi
  - qtPCR

Paramecium - *Chlorella* Phylogenetics



A model for endosymbiosis

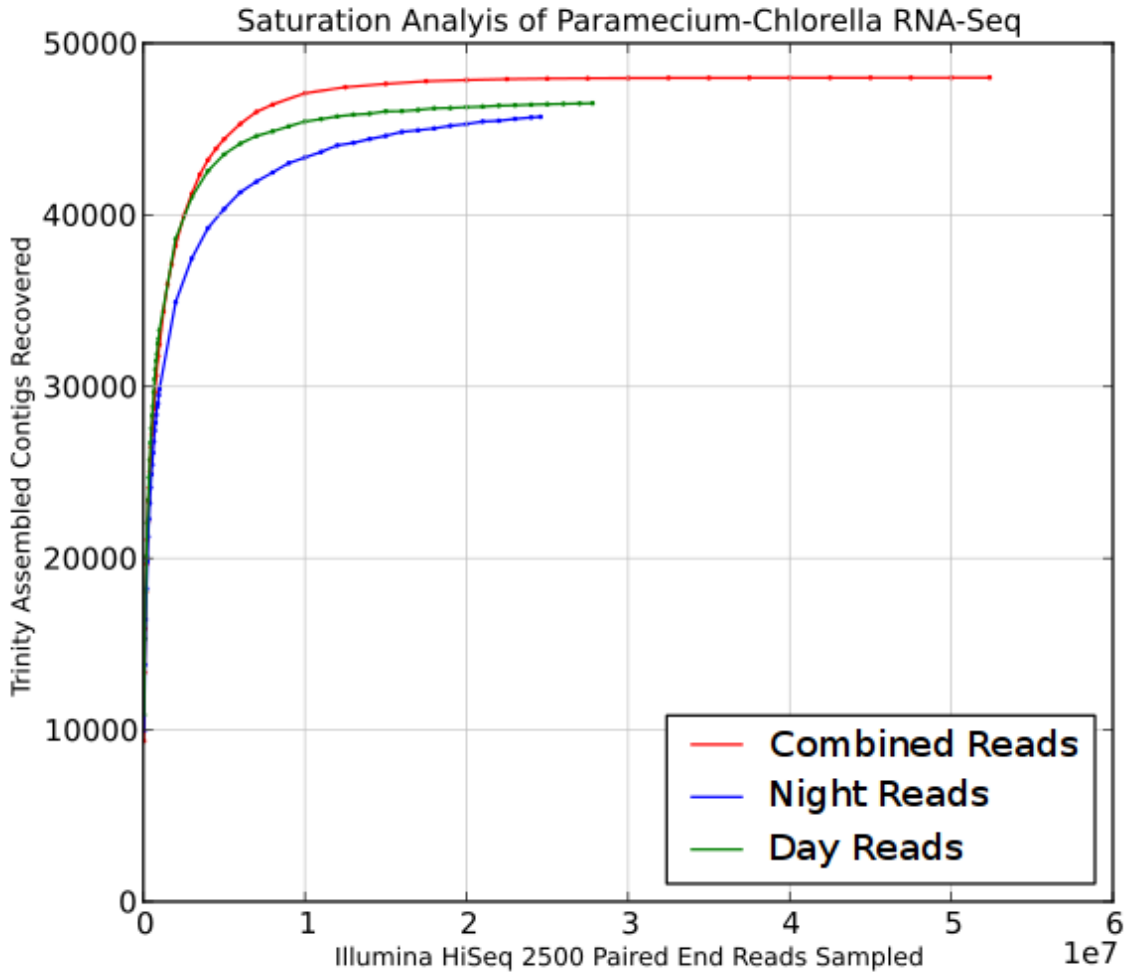
- Facultative endosymbiosis
- Evolution of secondary photosynthetic endosymbioses
- Established model organism
- RNAi available



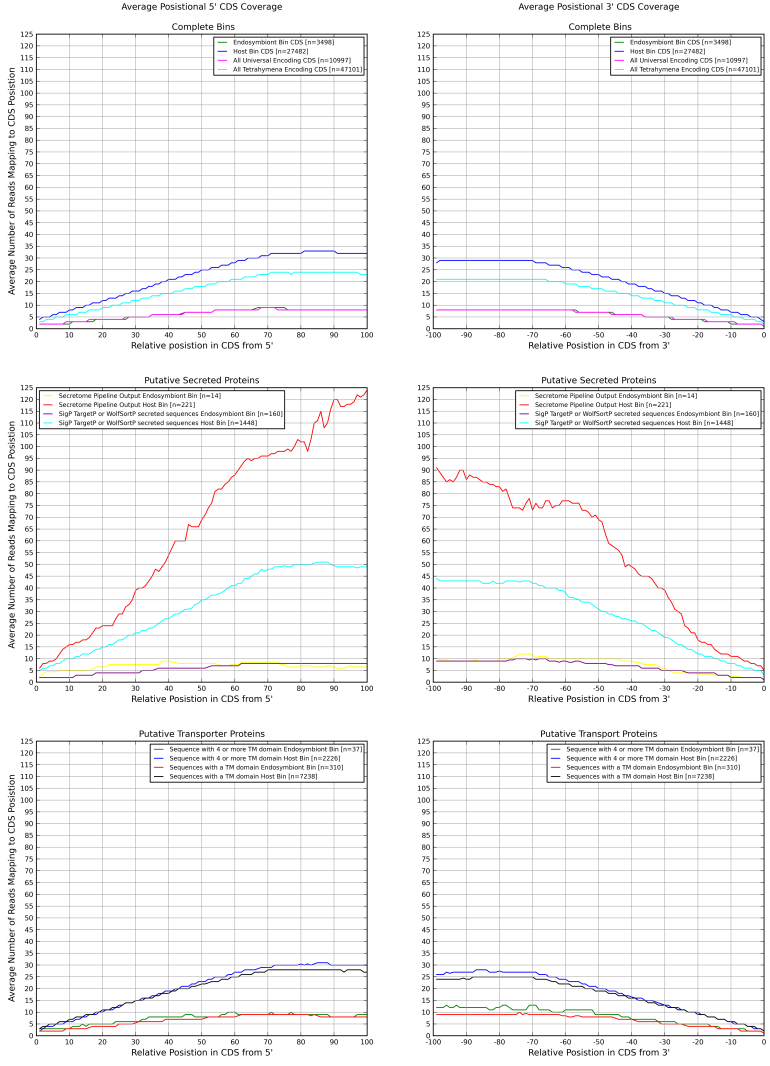
Day/Night Transcriptomics

- Day/Night Transcriptomics
  - Midday and Midnight
  - Removing bacterial foodstock
  - RNA Extraction
  - Sequencing
- Single cell transcriptomics
  - Paraquat treatment
  - Confocal microscopy
  - Single cell picking
  -
- Assembly De-Novo (Oasis vs Trinity)
- Referenced Assemblies

Saturation analysis

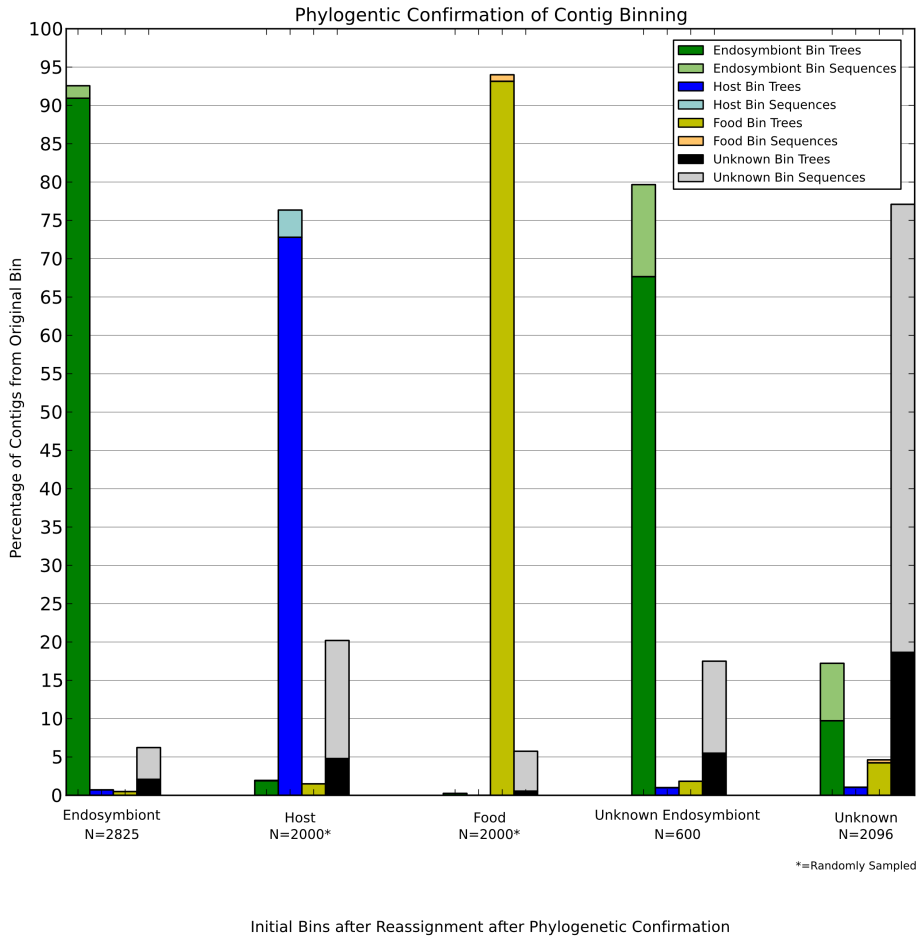


Terminal End Coverage



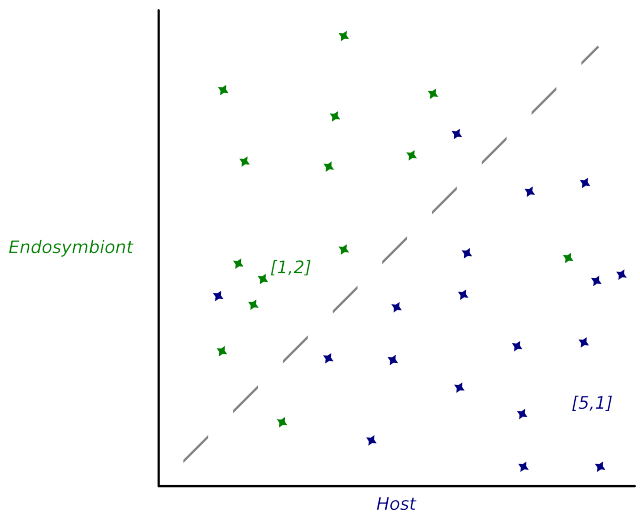
Phylogenomic binning

- Initial binning - BLAST hit and predicted CDS
- ML Phylogenomic pipeline

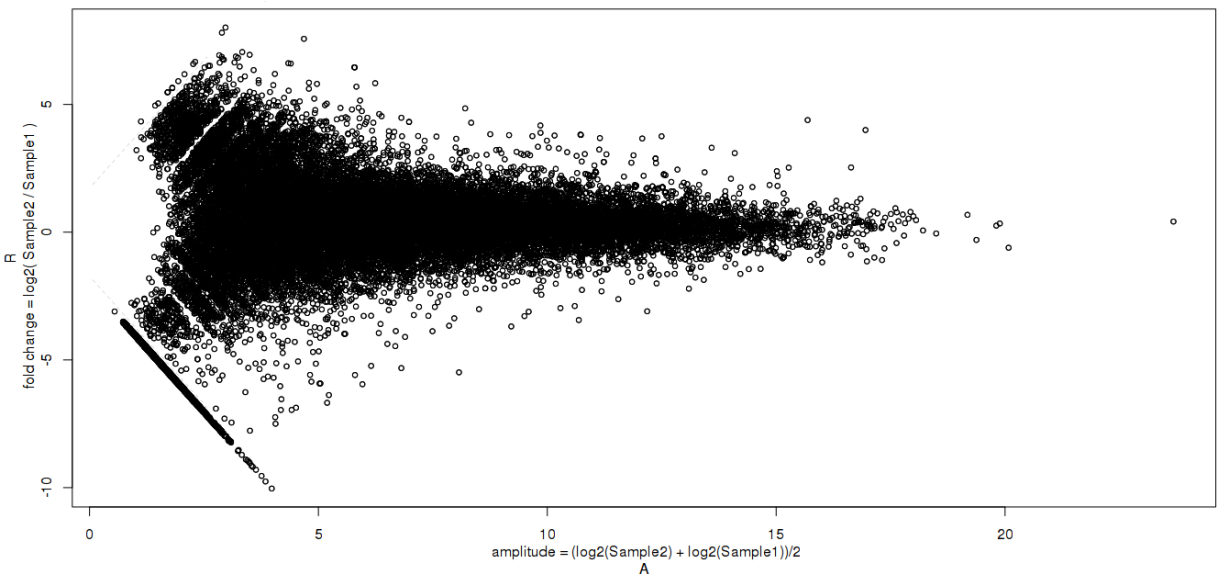


Machine learning automation

- 10,000 phylogeny manual labelling
- Taxonomic lookup via NCBI API
- Convert each phylogeny into a vector e.g. $V_i = [\frac{1}{Endosymbiont_i}, \frac{1}{Host_i}, \frac{1}{Food_i}, \frac{1}{Unknown_i}]$
- Train Multi-class Support Vector Machine on manually curated subset
- Cross-validation of SVM
- Optional unsupervised machine learning (e.g. clustering)

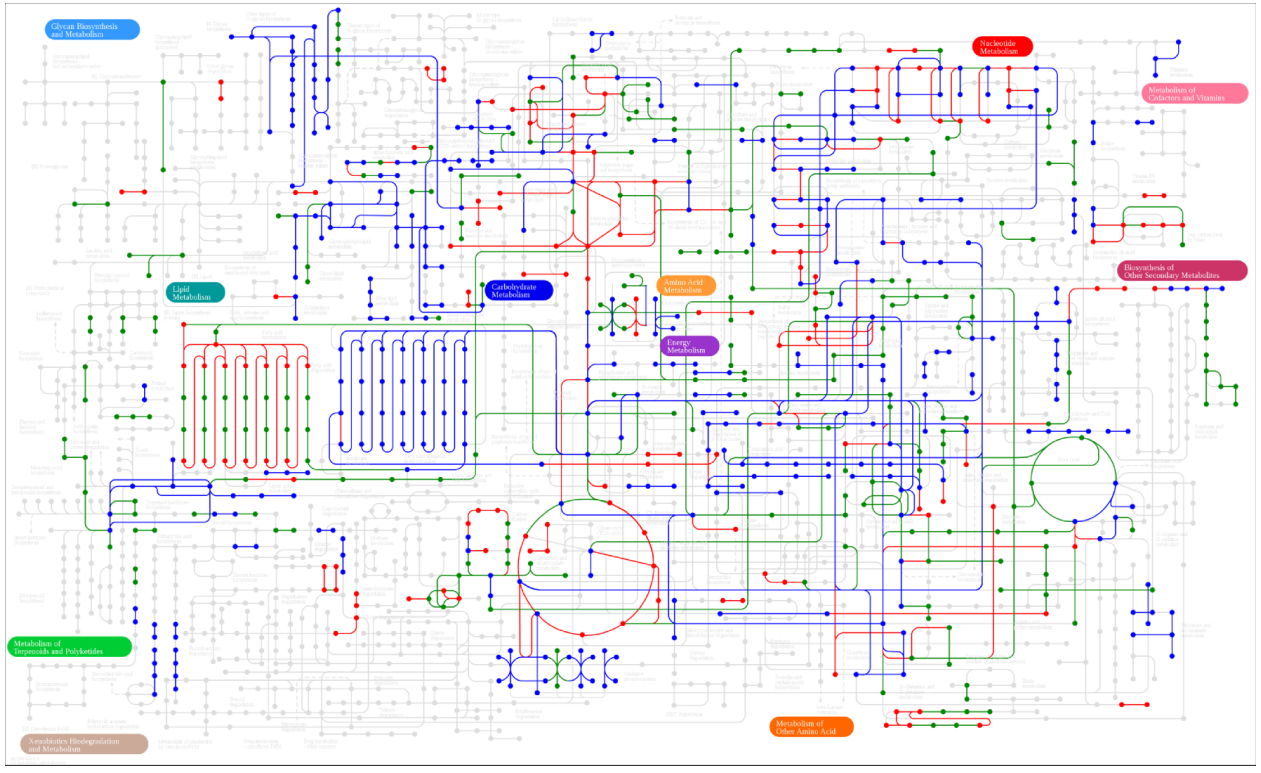


Differential Expression Analysis



Metabolic Reconstruction

- Annotation of partitioned datasets:
  - Trinotate
  - Blast2GO
  - KAAS
- Differential expression cut-offs:
  - day/night
  - With/without endosymbiont
- Pathway reconstruction:
  - Automated connection of units
  - Confidence based on prediction scoring
  - Manual handling of disjoints and inconsistency


Glycan Biosynthesis and Metabolism
Nucleotide Metabolism
Metabolism of Cofactors and Vitamins
Lipid Metabolism
Carbohydrate Metabolism
Amino Acid Metabolism
Biosynthesis of Other Secondary Metabolites
Energy Metabolism
Metabolism of Terpenoids and Polyketides
Xenobiotics Biodegradation and Metabolism
Metabolism of Other Amino Acid

- Candidate endosymbiont proteins:
  - 2 Amino Acid Permeases
  - Plasma Membrane Hydrogen ATPase
  - Ammonium Transporter
  - Plasma Membrane Iron Permease
  - Potassium Transporter
  - Sulfate Transporter

Work in progress

- Metabolomics data
  - Carbohydrate metabolism
  - Amino acid metabolism
  - Total metabolites
- RNAi
  - RNAi vector synthesised
  - *E. coli food-stock transformation*
  - Bug22 verification
  - Microinjection