Developing MetQy: an R package to query metabolic functions of genes and genomes

Andrea Martinez-Vernon & Orkun Soyer

School of Life Sciences, University of Warwick

BACKGROUND

- **KEGG**: extensive gene and metabolic knowledge contained in a collection of **databases**

KEGG DATABASE MAPPING

GENES ➡ KEGG ORTHOLOGS ➡ KEGG GENOMES ➡ KEGG MODULES

PROBLEM

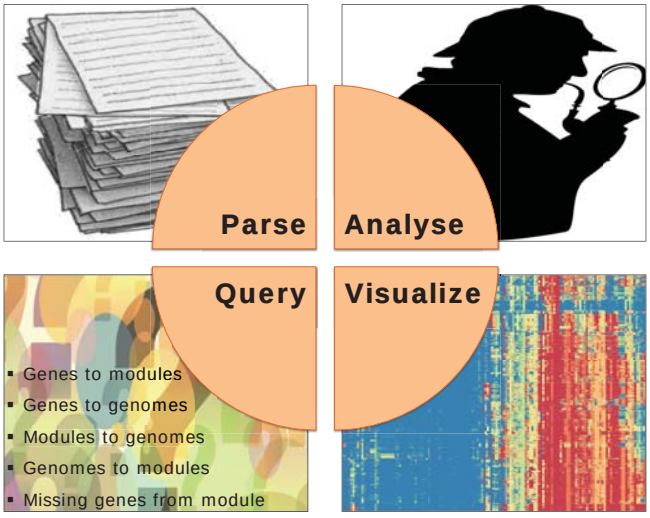
- Limited usability of KEGG web interface
- Scaling-up of analyses
- Using FTP is time-intensive & needs in-house expertise

AIM

Develop a computational tool that:

1. is easy-to-use, open source and accessible
2. allows easy interfacing between KEGG databases
3. enables large-scale automated cross-analyses on these

MetQy FUNCTION FAMILIES

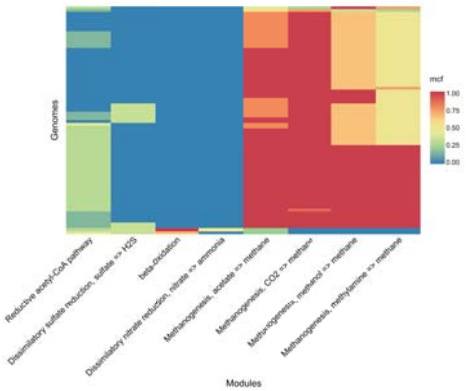


R package

BIOLOGICAL EXAMPLE TO ILLUSTRATE MetQy FUNCTIONALITY

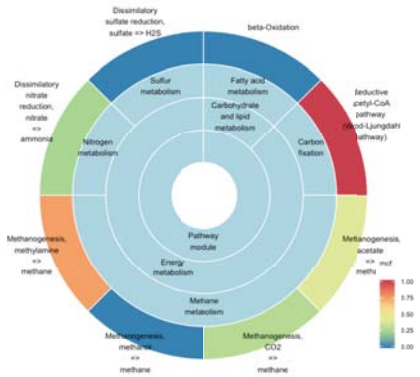
- Analysed modules that are loosely related to anaerobic digestion (AD) across organisms whose name contain "methano"
- Calculate KEGG **module completeness fraction (*mcf*)*** using *query_genomes_to_modules()*

A) Module fraction completeness (*mcf*)



plot_heatmap()

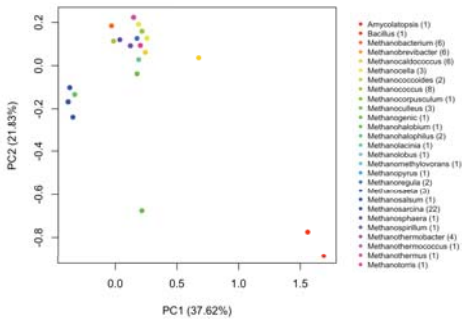
B) Zoom in - *mcf* of T04272



Methanogenic archaeon ISO4-H5 (T04272)

plot_sunburst()

C) Genus-level PCA of *mcf*



analysis_genomes_module_output()

*mfc = number of module reactions present / total number of reactions

FIND OUT MORE!

github.com/OSS-Lab/MetQy

Martinez-Vernon et al. *Under revision*
doi: https://doi.org/10.1101/215525

ACKNOWLEDGEMENTS

Funding from the University of Warwick and the EPSRC & BBSRC Centre for Doctoral Training in Synthetic Biology (grant EP/L016494/1).