KEGGaNOG: A Streamlined Solution for KEGG Pathway Annotation
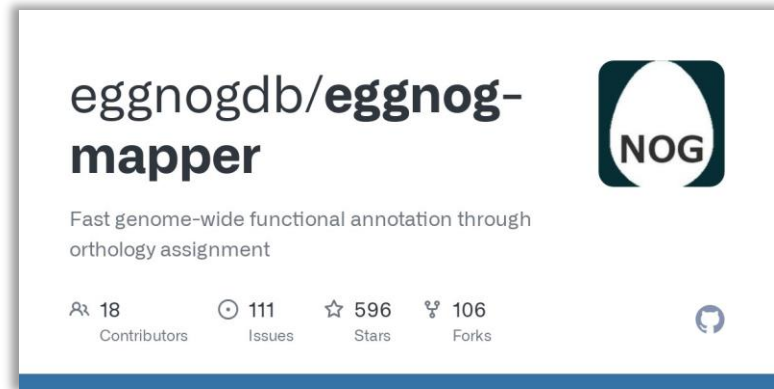
Ilia Popov[1], Igor Popov[1], Alexey Ermakov[1]

[1] «Bioengineering and Veterinary Medicine» Faculty, Don State Technical University, Rostov-on-Don, Russia

Introduction & Purposes of use

eggNOG-mapper is a tool for fast functional annotation of novel sequences

- It uses **precomputed orthologous groups** and phylogenies from the eggNOG database
- Its **annotations are** well **detailed**
- However, its **hard to interpret** them
- For instance, it is **hard to understand** which KEGG Orthology **(KO) terms** are coded in 'KEGG_ko' column of eggNOG-mapper annotations!

???

KEGG_ko
K00002
K13577
K01078
K22390
K19356
K22032
K08139
K08141
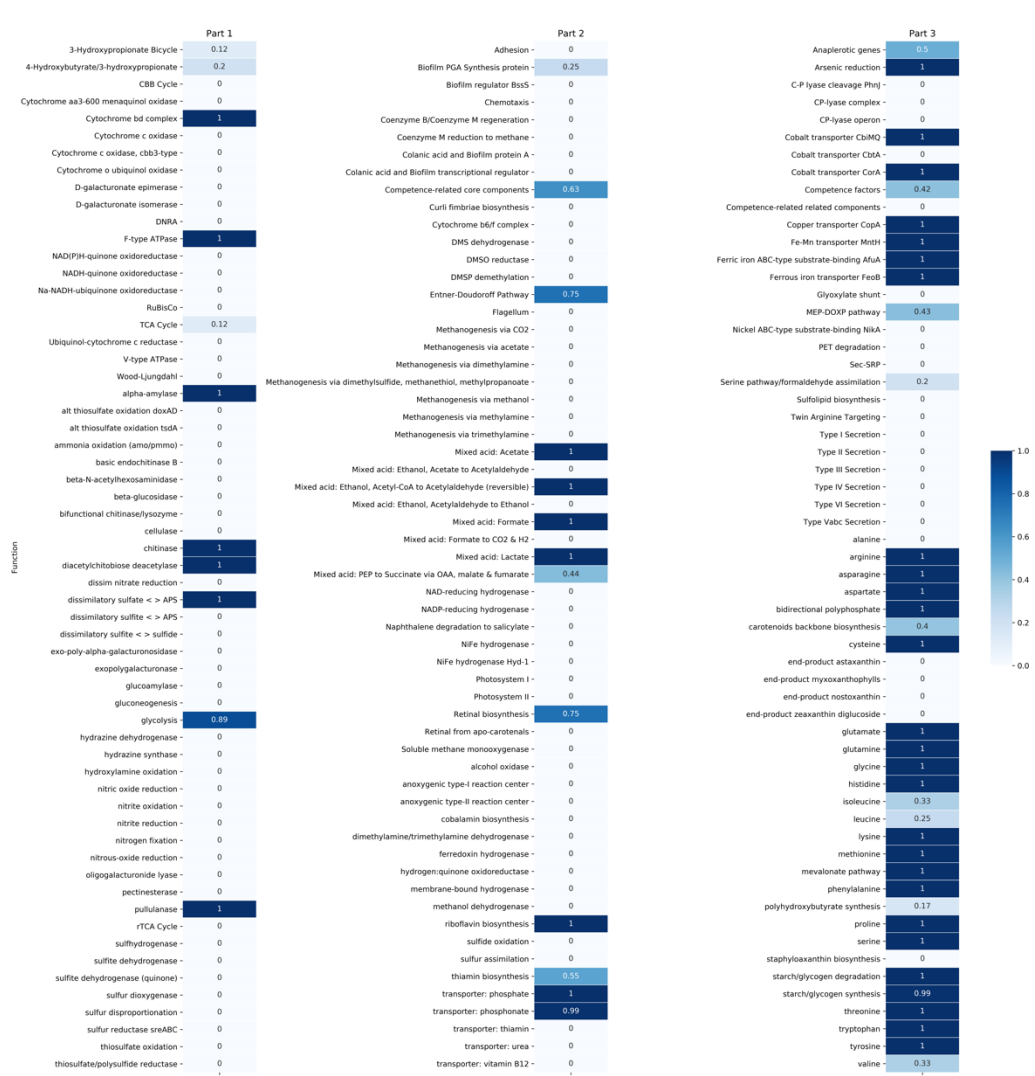K20523
K08504
K01876
K22503

KEGGaNOG solves this problem:
- It takes eggNOG-mapper annotations
- Extracts 'KEGG_ko' column
- Decodes KO terms with KEGG-Decoder and creates *.tsv* file with pathways completeness
- Provides several visualization options

KEGGaNOG is perfect for
- Annotating bacterial metabolic profiles (e.g. probiotics)
- Annotating metagenomes metabolic profiles

Sample outputs

Complete metabolic profile of probiotic *Lpb. plantarum* IS-12506. KEGGaNOG's default visualization with single sample layout. Heatmap depicts pathways completeness (0 to 1).

Complete metabolic profiles of 10 metagenome samples. KEGGaNOG's default visualization with multi sample layout.
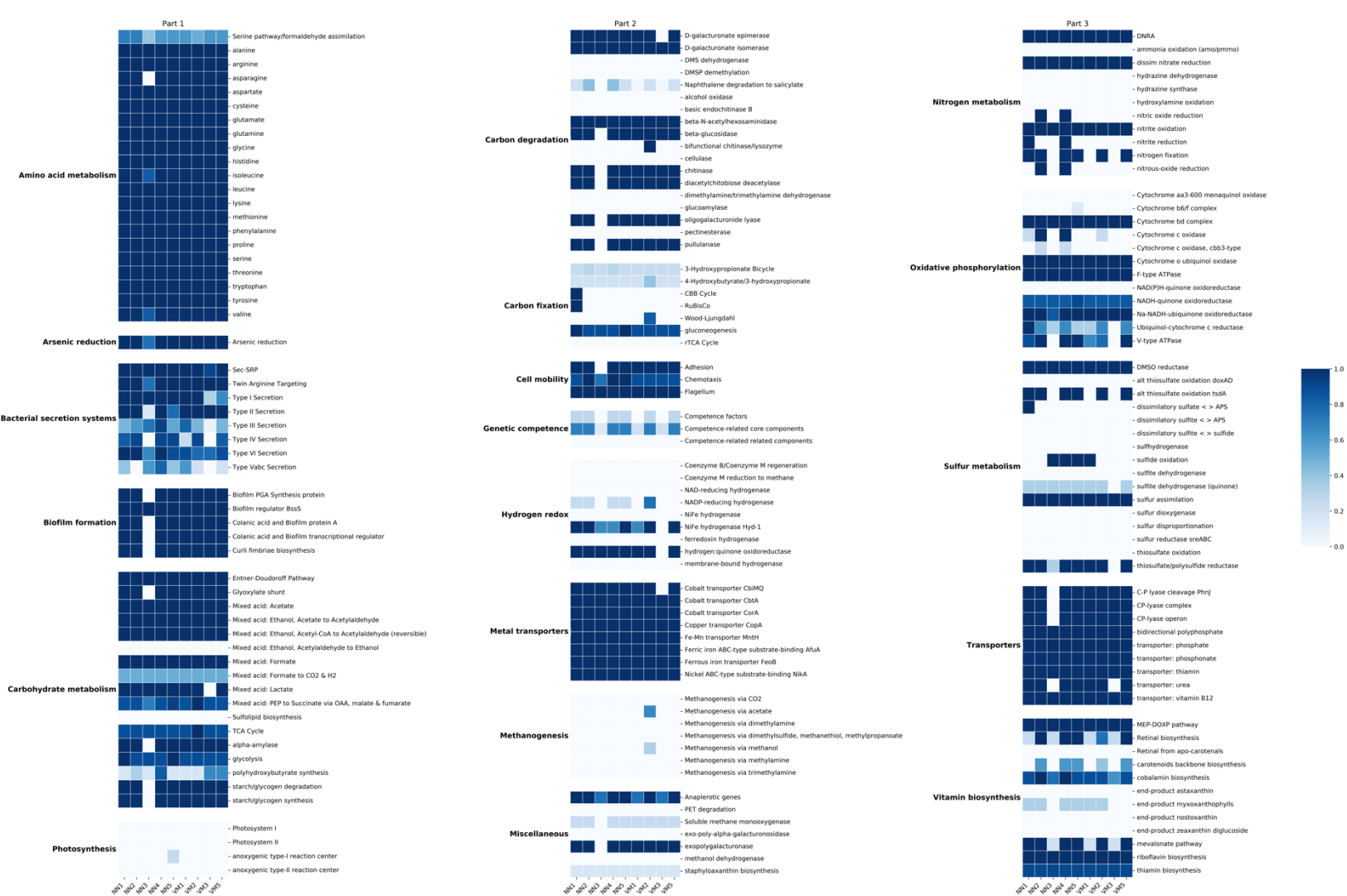
User APIs

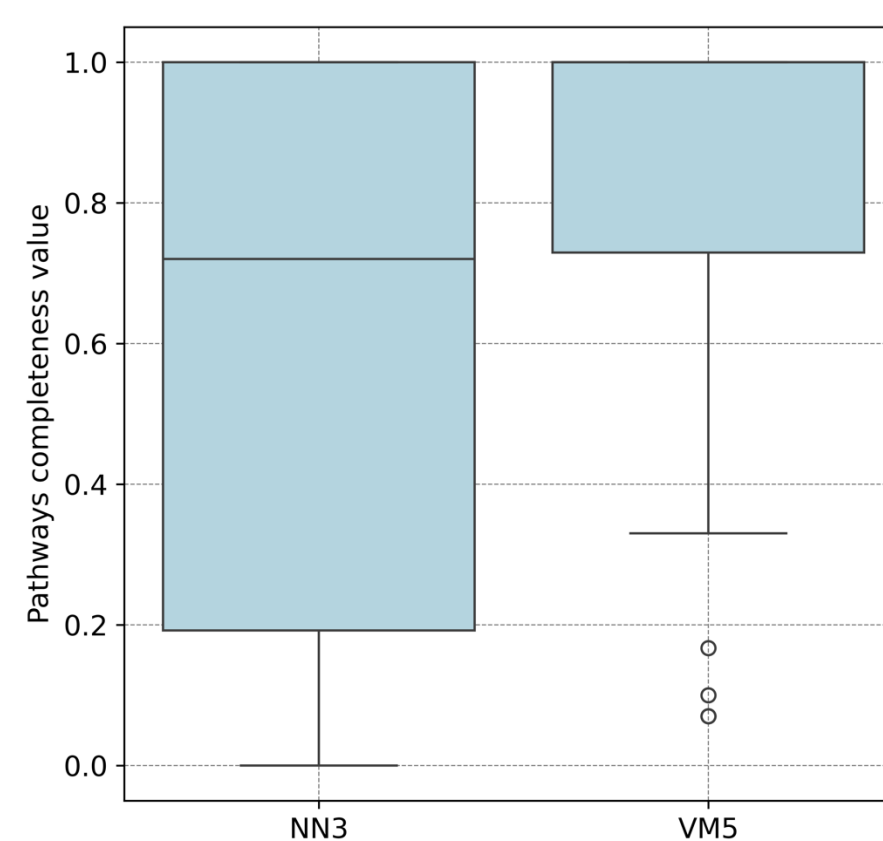User is also provided with several other options of visualization
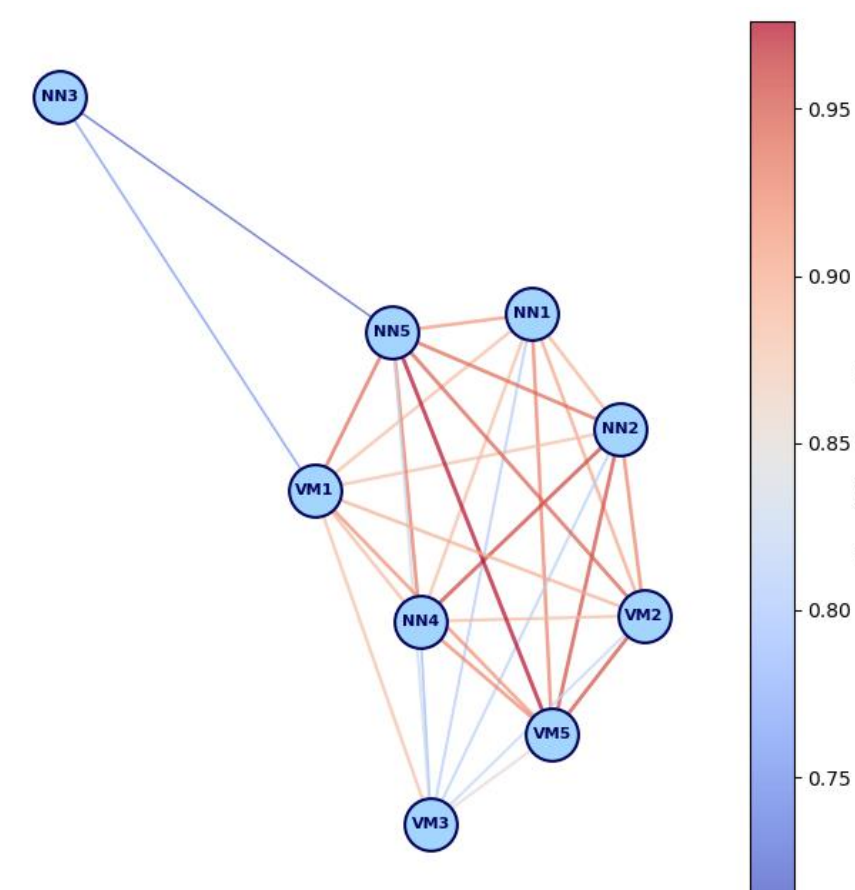
Option 1: Correlation network

KEGGaNOG builds correlation matrix and plots it as the network. User is provided with options to adjust the plot (cmap, minimal correlation to plot etc.)
This function allows to compare samples between each other and see the pattern – which sample differs from others

Option 2: Boxplots

KEGGaNOG uses the pathways completeness data and compares the completeness tendency among several samples
This function allows to compare samples in terms of which sample has core complex metabolic profile and which – not.

Several other APIs are under development at the moment:
1. Spider-plot API – to compare the completeness of one specific pathway between samples
2. PCoA API – to visualize metabolic profile similarity and divergence, highlighting clustering patterns and functional shifts

Maintenance updates are also under active development

Acknowledgements

Instructions

GitHub and README

Python Package Index