FUNCTIONAL PREDICTION AT THE (META-)GENOMIC SCALE BY EVOLUTIONARY ANALYSIS

JAIME HUERTA CEPAS

BORK GROUP, EMBL HEIDELBERG

1. PHYLOGENOMIC APPROACHES FOR ORTHOLOGY DETECTION AND FUNCTIONAL PREDICTION

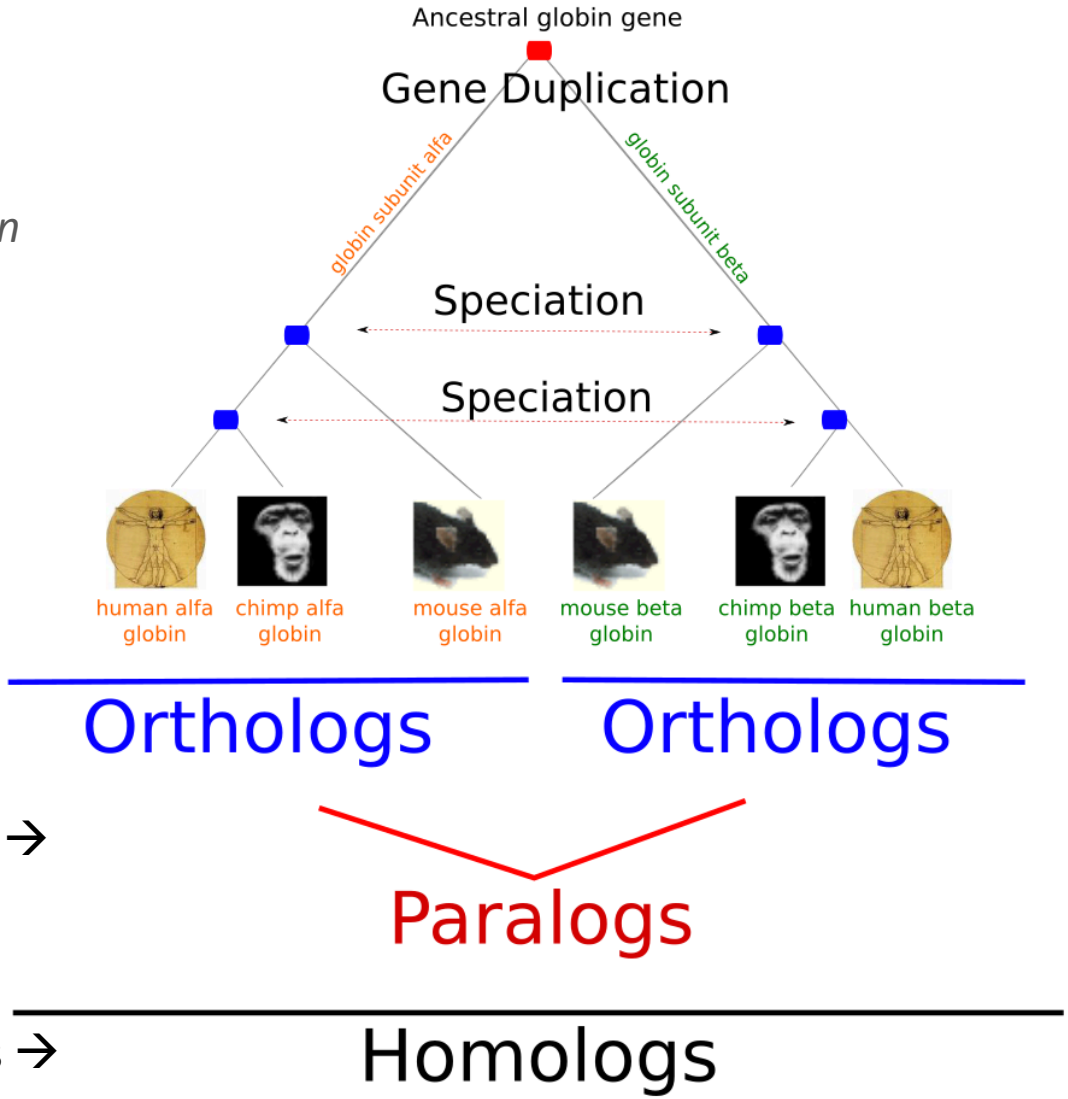
ORTHOLOGY AND FUNCTIONAL PREDICTION

Ortholog conjecture:

Function is more conserved between orthologs than between paralogs

↓

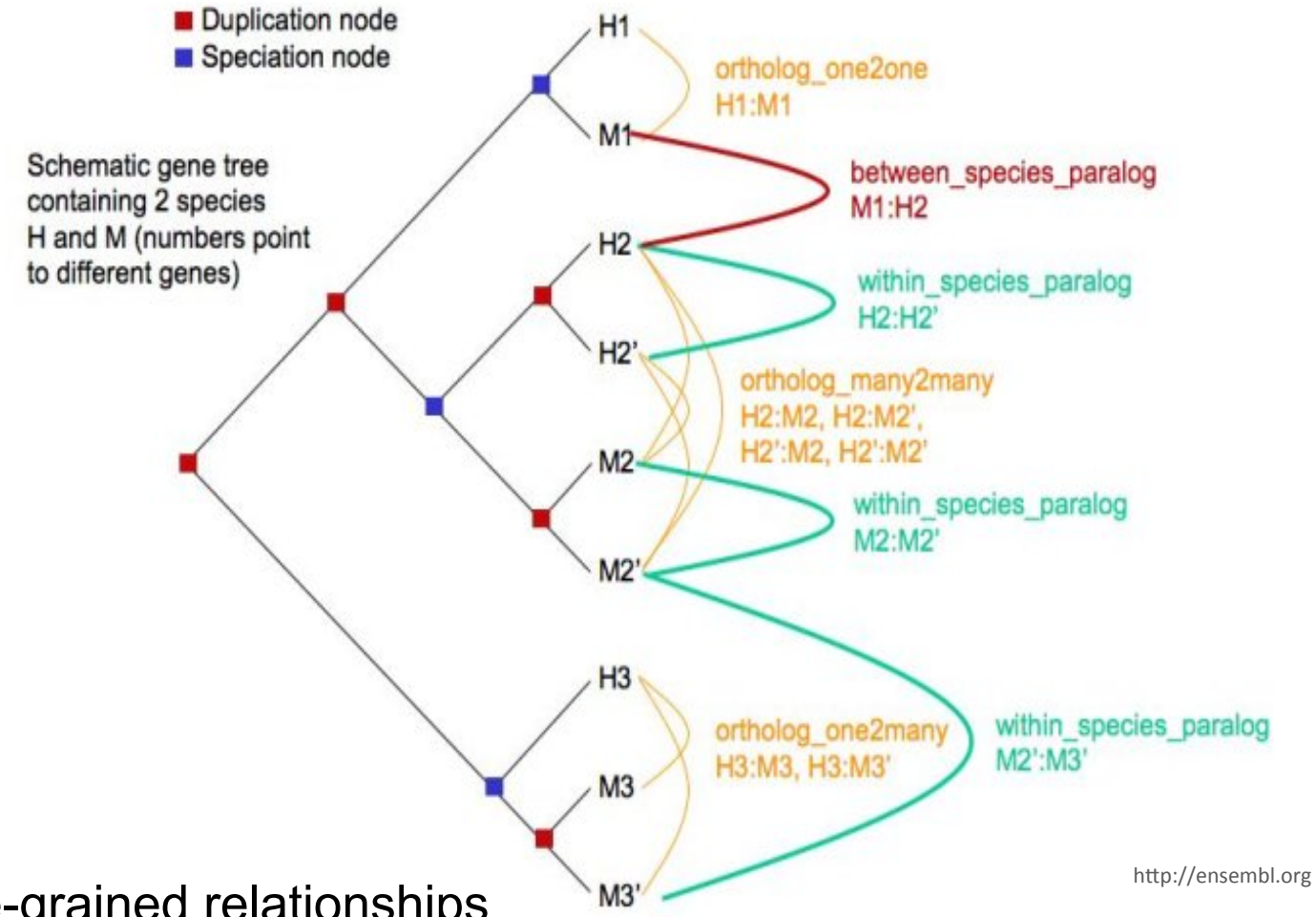
Predicting gene function turns into an orthology prediction problem



Deeper Evolutionary → analysis

BLAST results →

ORTHOLOGY DETECTION METHODS (PHYLOGENY)

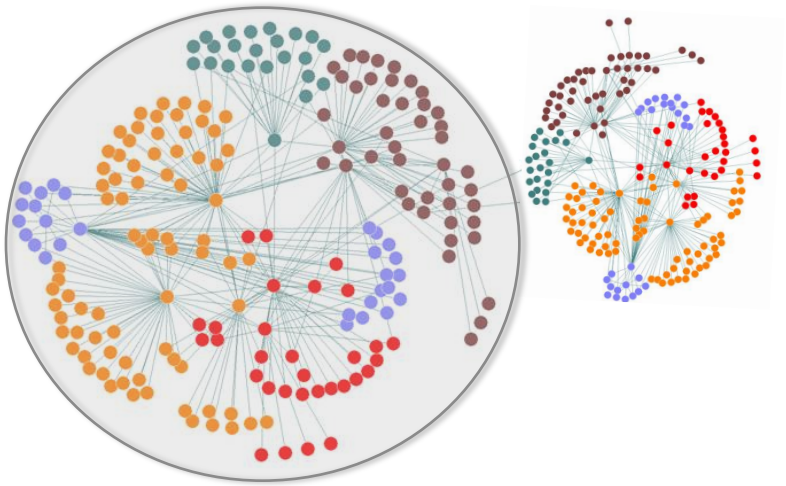


http://ensembl.org

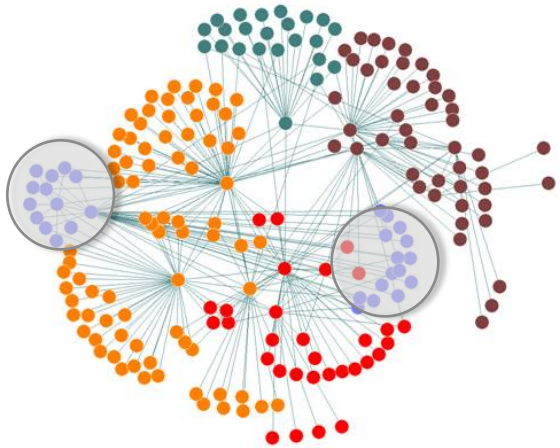
- Infer fine-grained relationships
- Allows tracing the origin of duplication and speciation events

ORTHOLOGY DETECTION METHODS (CLUSTERING)

- Build Clusters of Orthologous Groups (COGs) based on sequence similarity network
- Tend to be faster and cope for larger amounts of species/ sequences
- No co-orthology disambiguation and requires taxonomic constraints



1 COG (all species)

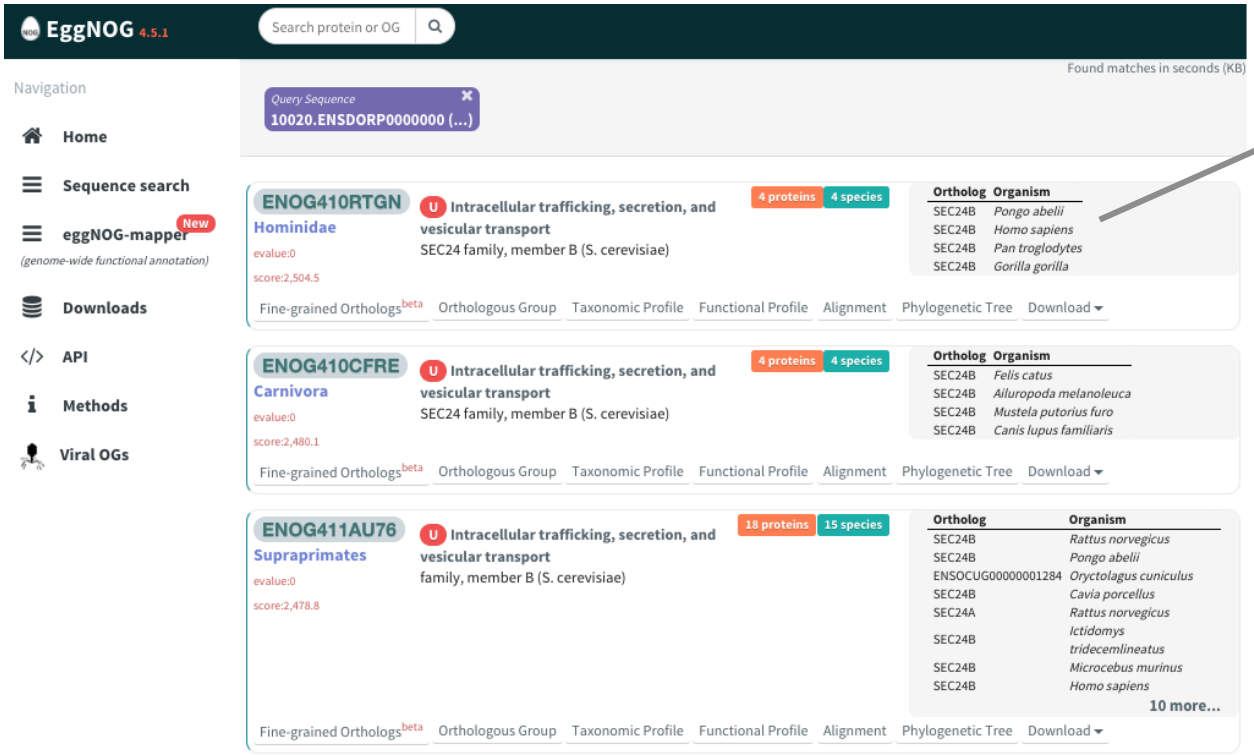


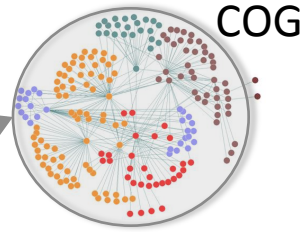
2 COGs (mammal species)

EggNOG 4.5.1

A database of orthologous groups and functional annotation

Organisms	Viruses	Orthologous Groups	Trees & Algs.
2,031	352	190k	1.9M



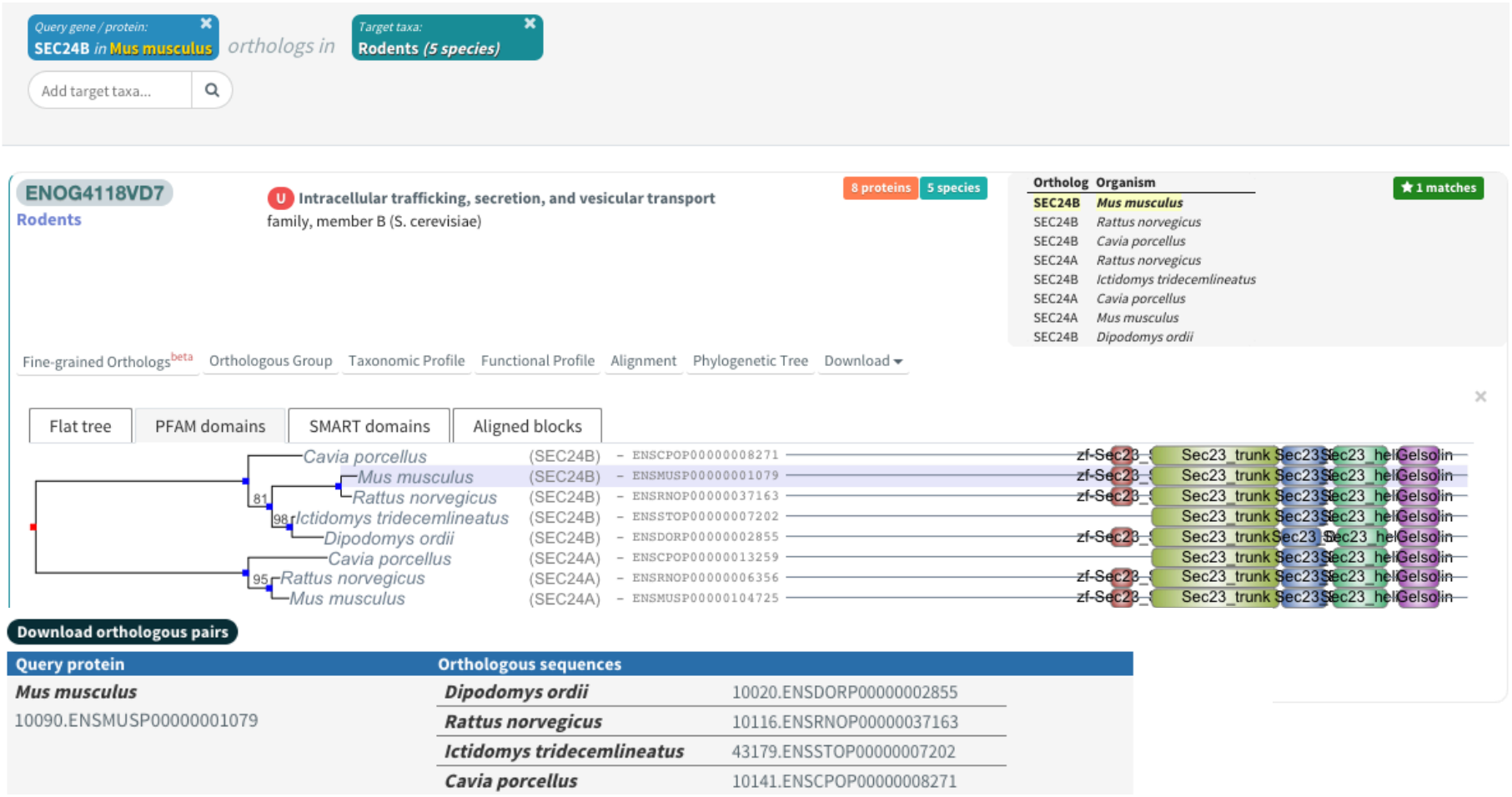


- Rapid sequence search
- Find orthologous groups
- Functional annotations
  - Domains
  - Gene Ontology
  - KEGG
  - Taxonomic distribution

http://eggnog.embl.de

Huerta-Cepas, et al. 2016 (Nucleic Acid Research)

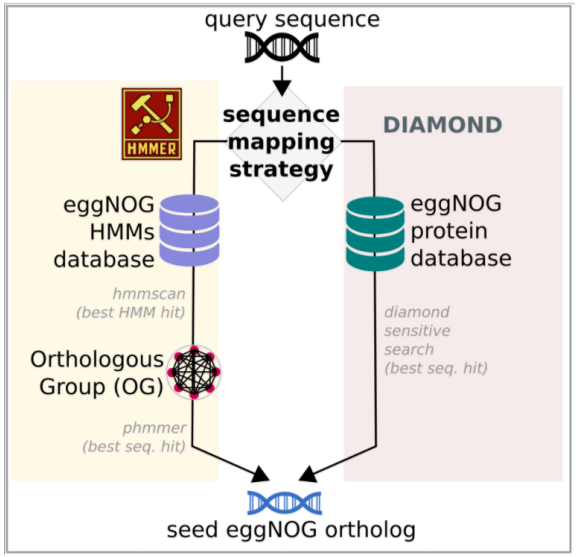
FINE GRAINED ORTHOLOGY PREDICTIONS

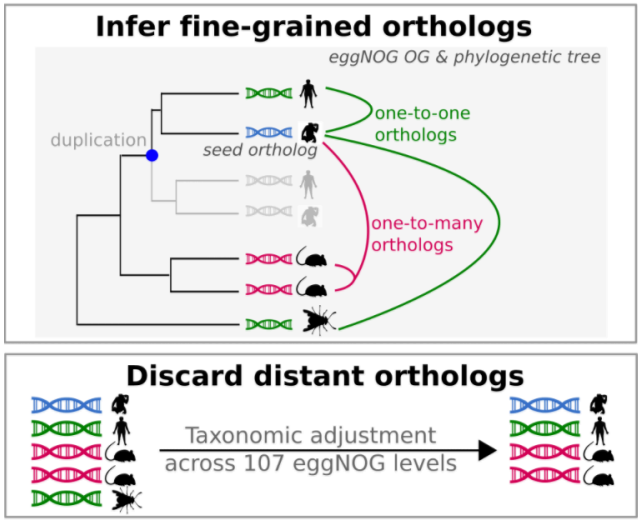


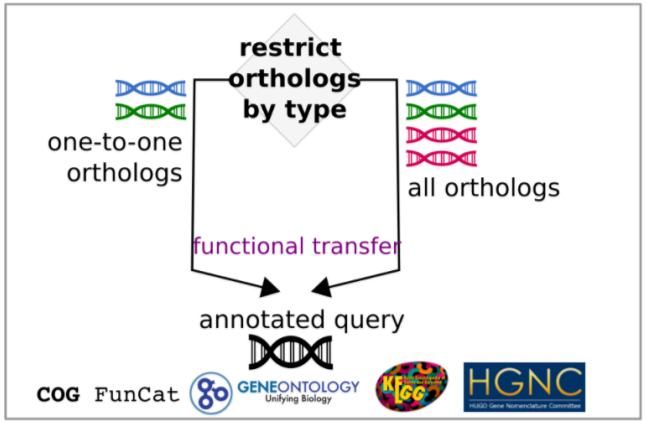
http://eggnog.embl.de

Huerta-Cepas, et al. 2016 (Nucleic Acid Research)

EFFICIENT ANNOTATION BY ORTHOLOGY MAPPING







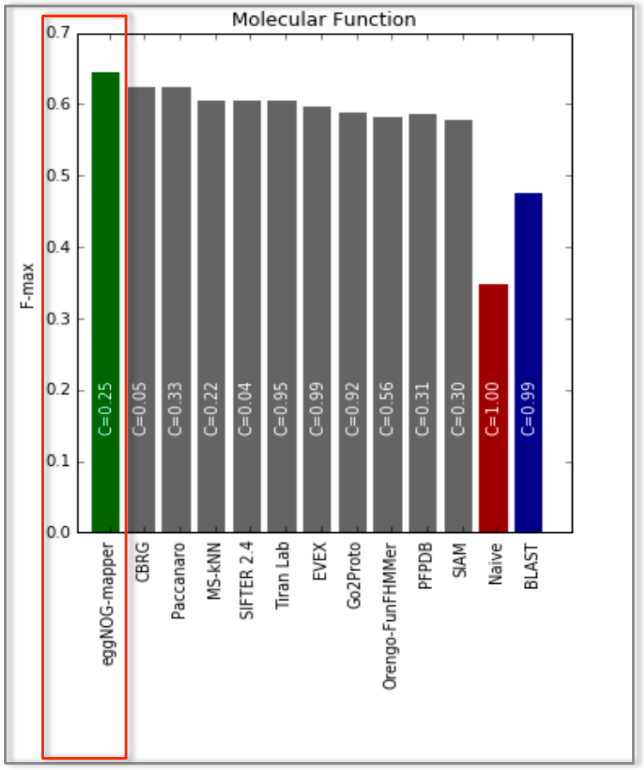
Find query's best match *Place it in tree* *Transfer terms from orthologs*

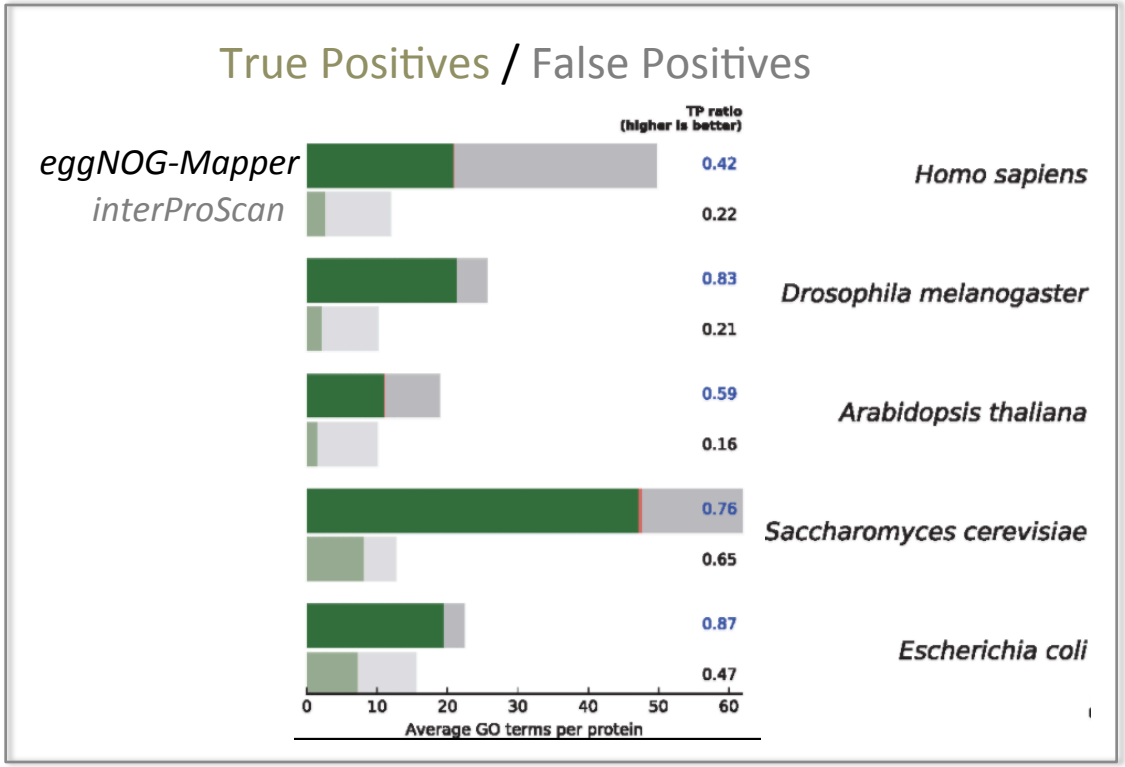
http://eggnog-mapper.embl.de

Huerta-Cepas et al. 2016 (submitted)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 686070

EGGNOG-MAPPER PERFORMANCE





CAFA2: Jiang et al. Genome Biology (2016)

CAFA2 Challenge:

- *eggNOG-mapper ranked **top 5 out of 126 methods** in the 3 GO categories*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 686070

3. FUNCTIONAL EXPLORATION OF METAGENOMIC DATASETS

FUNCTIONAL ANNOTATION OF METAGENOMIC DATA

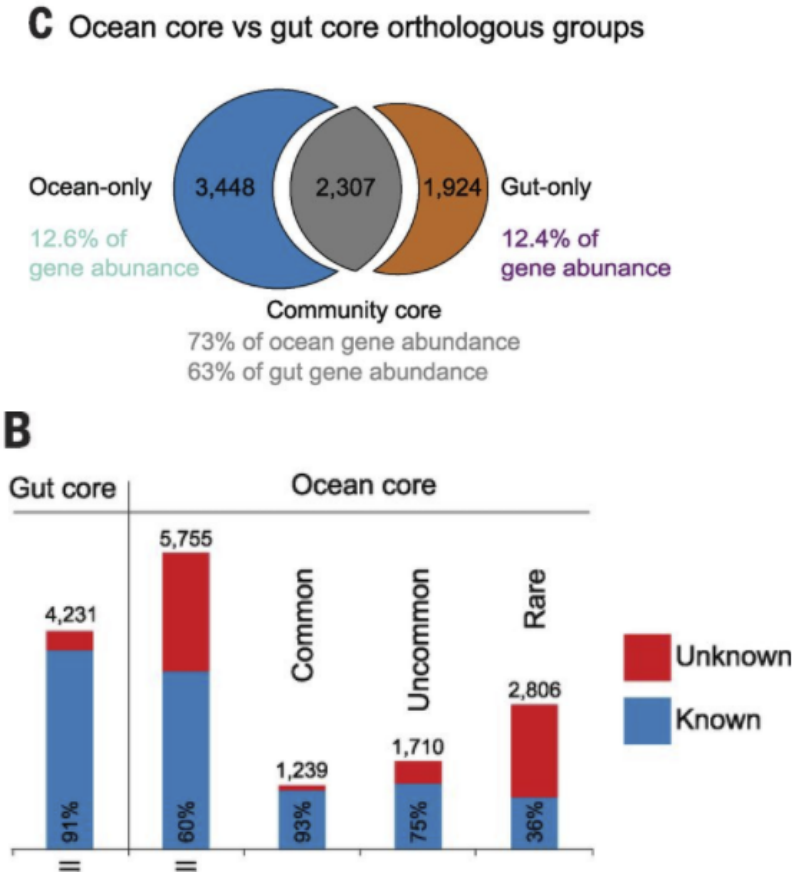
Objective: "develop cutting edge methods for using large scale data to design cell factories and bacterial communities for biotechnological applications" http://dd-decaf.eu

@EMBL

- Mining new enzymes out from genomic and metagenomic data
- Building metabolic models based on functional annotation of bacterial communities

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 686070

EXPLORING METAGENOMICS UNKNOWN FRACTION

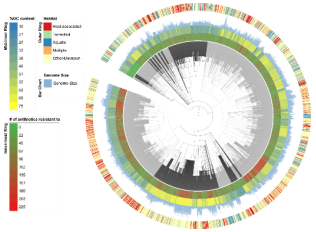


Structure and function of the global ocean microbiome
Sunagawa el al. Science 348 (6237), 1261359

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 686070

3. TOOLS

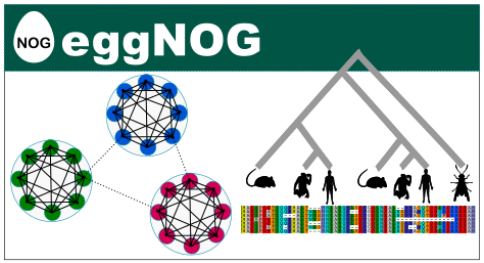
PROGENOMES



Consistently annotated prokaryotic genomes

http://progenomes.embl.de

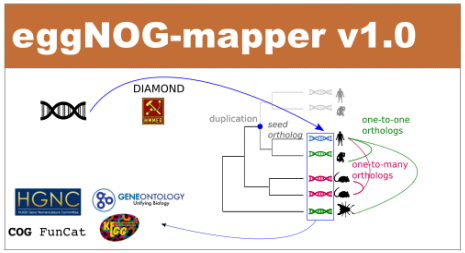
Mende et al. NAR 45 (2017) D286



Orthologous groups and functional information

http://eggnog.embl.de

Huerta-Cepas et al. NAR (2016)



Fast accurate functional annotation

http://eggnog-mapper.embl.de

Huerta-Cepas et al. MBE (2017)

ETE Toolkit

A Python framework to work with trees

```
from ete3 import Tree
tree = Tree('((A,B), D);')

print tree
#       /-A
#    /-|
# --|   \-B
#    \-D

A = tree & "A"
A.up.show()
```

Phylogenetic reconstruction and analysis

http://etetoolkit.org

Huerta-Cepas et al. MBE (2016)