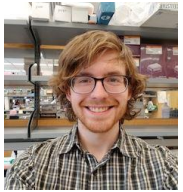
Survey of metaproteomics software tools for functional microbiome analysis



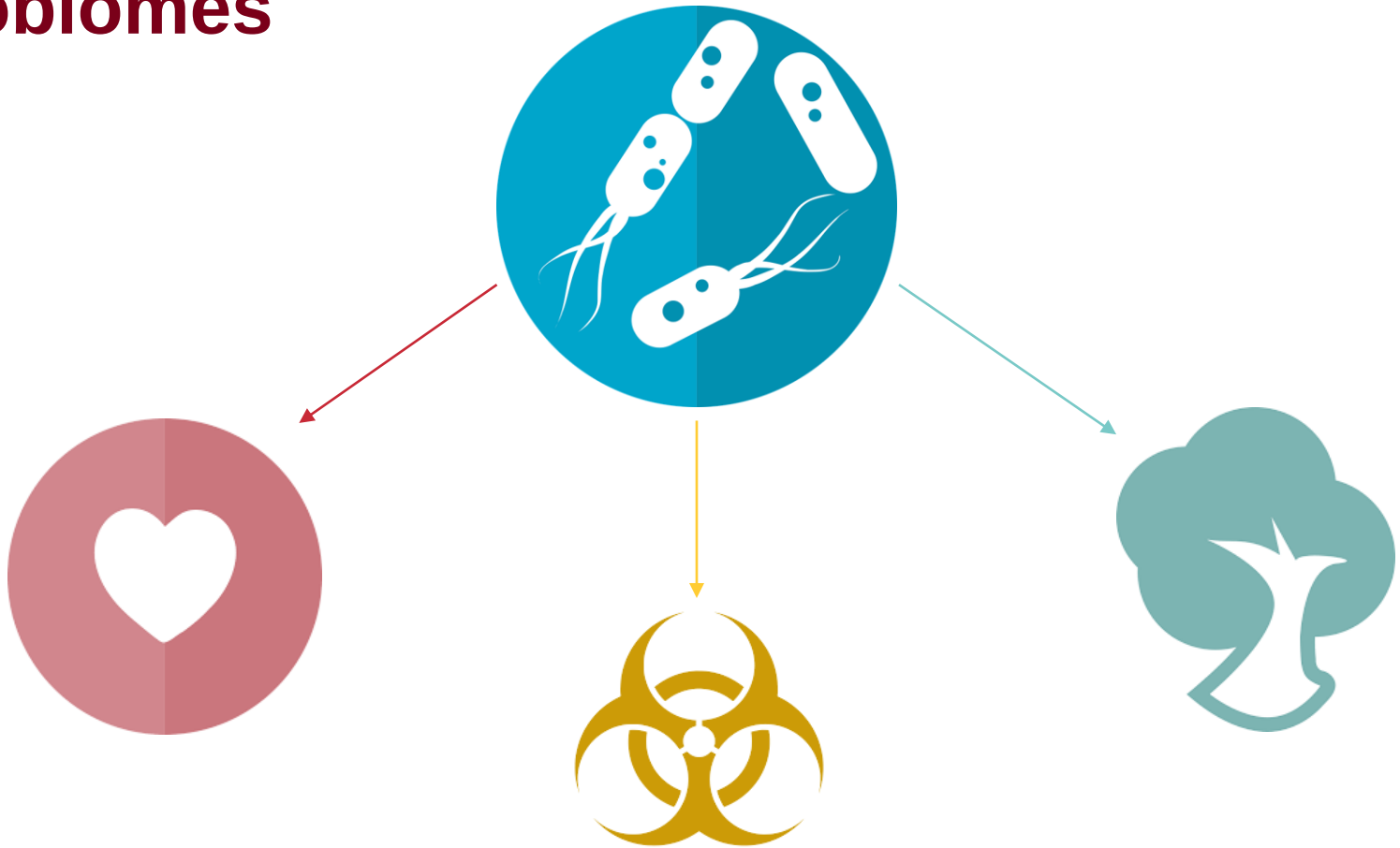
***Ray Sajulga**, **Caleb Easterly**,*

M. Riffle, B. Mesuere, R.G. Singh, T. Muth, S. Mehta, P. Kumar, J. Johnson, B. Gruening, D. May, W. J. Hervey, C. Kolmeder, A. Tanca, N. Held, M. A. Saito, B. Nunn, J. Rudney, T. J. Griffin, Pratik D. Jagtap

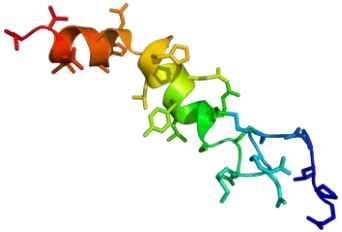
Wednesday, December 5th, 2018

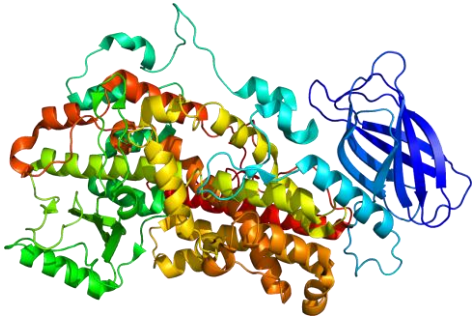
UNIVERSITY OF MINNESOTA
Driven to Discover®

Microbiomes



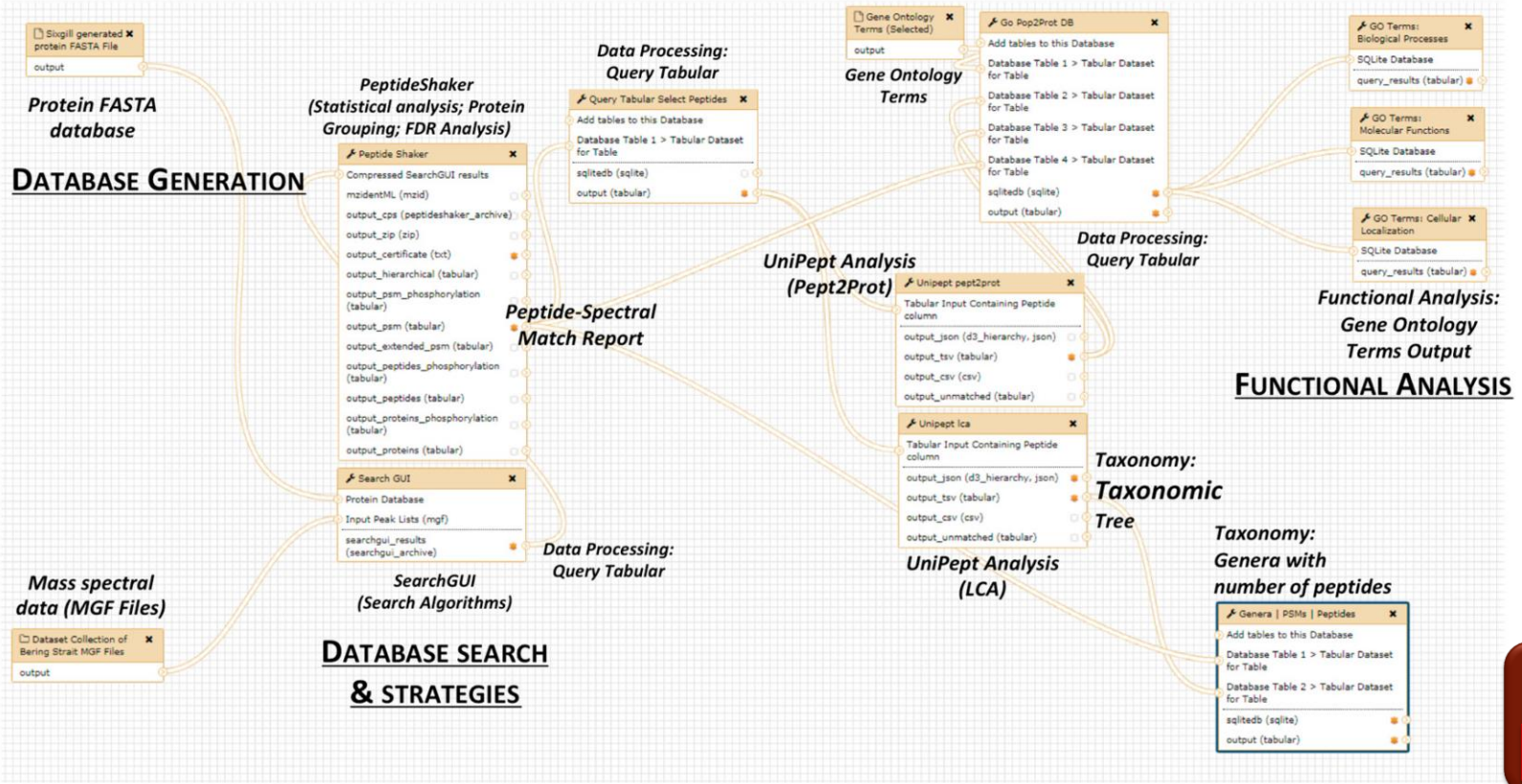
Metaproteomics and the functional microbiome





$f(x)$

Metaproteomic functional tools within Galaxy

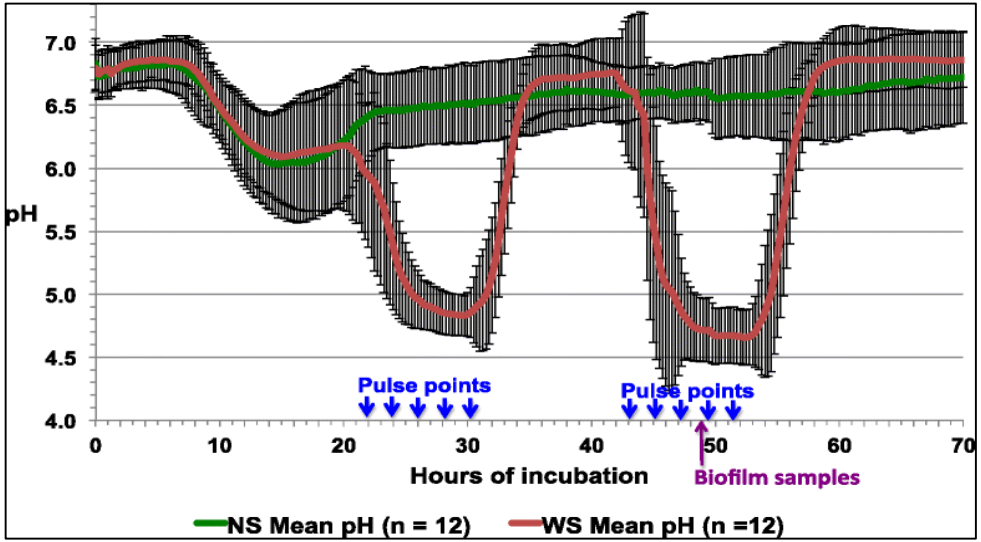


Metaproteomics Software Tools

	EggNOG	MEGAN6	metaGOmics	Meta Proteome Analyzer	Unipept
LEVEL OF ANALYSIS	Peptide (terms from protein orthologs)	Protein	Peptide	Metaproteins; Protein Groups	Peptide
ANNOTATION DATABASE	EggNOG	NCBI nr	UniProtKB	UniProtKB	UniProtKB
FUNCTIONAL ONTOLOGIES	COG categories, GO terms, BiGG reactions, KO groups	InterPro2GO, eggnog orthologous group, SEED and KEGG	GO terms	EC numbers, protein-level information from UniProt (ontology terms), KEGG, KO groups	GO Terms, EC numbers

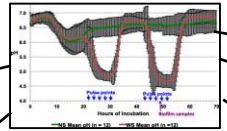
Test Study:
Sucrose-induced oral dysbiosis

- Mass spectral data was acquired from plaque samples from **one of the twelve subjects at high risk for dental caries grown in biofilm reactor** in the presence (With Sucrose, or **737WS**) and absence of sucrose (No Sucrose, or **737NS**) (16 fractions)
- Mass spectra were searched against the Human Oral Microbiome database (HOMD) to identify microbial peptides.
- Appropriate input files were provided to expert researchers / developers of the software.



Rudney et al., BMC Microbiome [DOI: 10.1186/s40168-015-0136-z]

Analysis of same dataset using multiple software



INPUTS

EggNOG	MEGAN6	metaGOmics	Meta Proteome Analyzer	Unipept
Peptide List	Peptides in FASTA format Search database BLAST-P Results	Peptides with spectral counts Search database	Spectrum files Search database	Peptide List
Huerta-Cepas *et al* Mol Biol Evol. (2017);34(8):2115-2122. doi: 10.1093/molbev/msx148	Huson *et al* Methods Enzymol. (2013);531:465-85. doi: 10.1016/B978-0-12-407863-5.00021-6.	Riffle *et al* Proteomes. (2017);6(1). pii: E2. doi: 10.3390/proteomes6010002.	Muth *et al* Anal Chem. (2018);90(1):685-689. doi: 10.1021/acs.analchem.7b03544	Gurdeep Singh et al J Proteome Res. (2018) doi: 10.1021/acs.jproteome.8b00716

OUTPUTS

EggNOG	MEGAN6	metaGOmics	Meta Proteome Analyzer	Unipept
Peptides annotated with protein hits and GO terms	Many options – used eggnog orthologous groups with spectral counts	GO terms with fold changes and associated statistical significance	UniProt protein IDs	GO terms with spectral counts

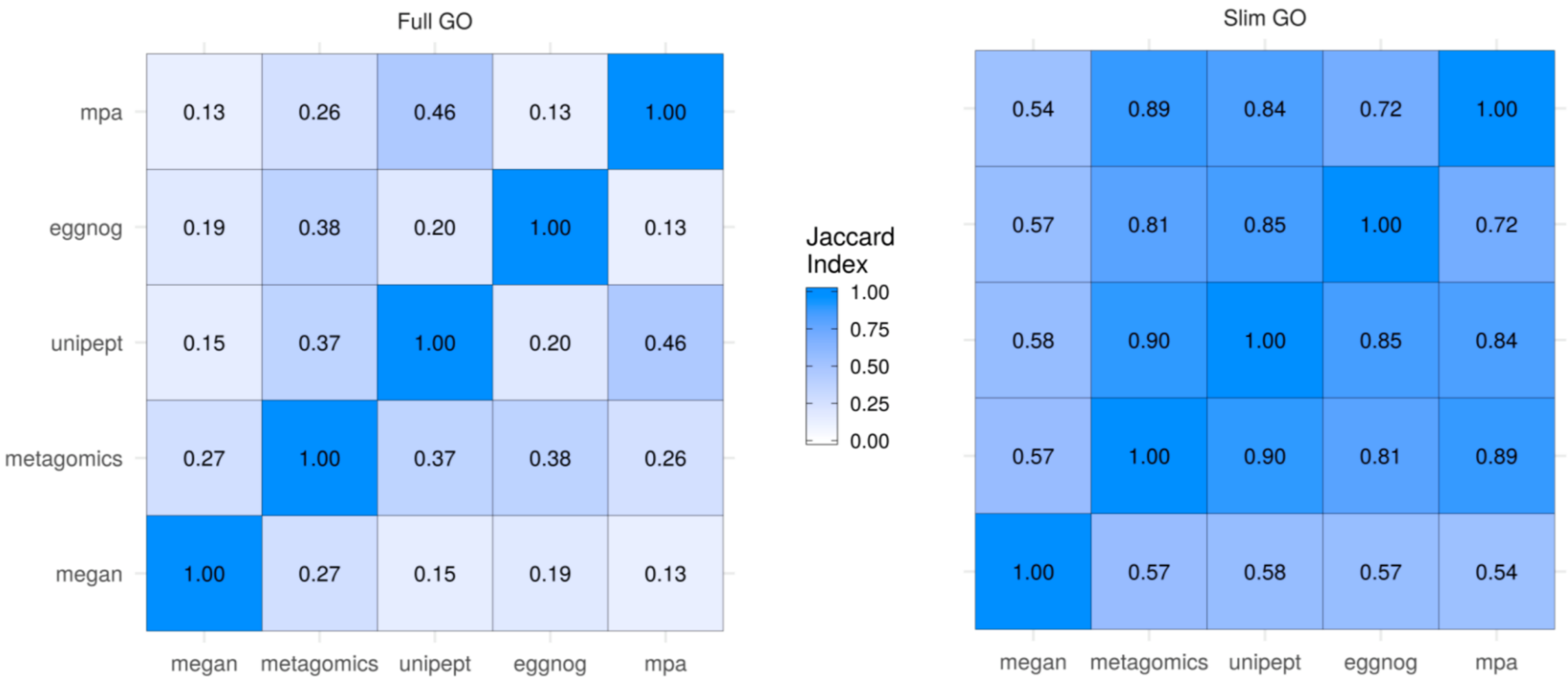
Identification Summary:

Sucrose-induced dysbiosis dataset

Tool	Native Output			GO term translation			
	Type	# Total	# Significant at FDR < 5%	# Total	# Exclusive to Tool	# Slim Total	# Slim Exclusive
eggNOG mapper	Proteins	18,440	NA	6,155	3070	144	13
MEGAN	eggNOG orthologous groups	1,665	Query eggnog API with orthologous group IDs	1,450	103	82	0
MetaGOmics	GO terms	3,944	1,958	3,944	645	116	0
MPA	Proteins	23,169	Query UniProt API with protein IDs	1,102	15	103	0
Unipept	GO terms	2,036	NA	2,036	217	123	0

GO Term Similarity:

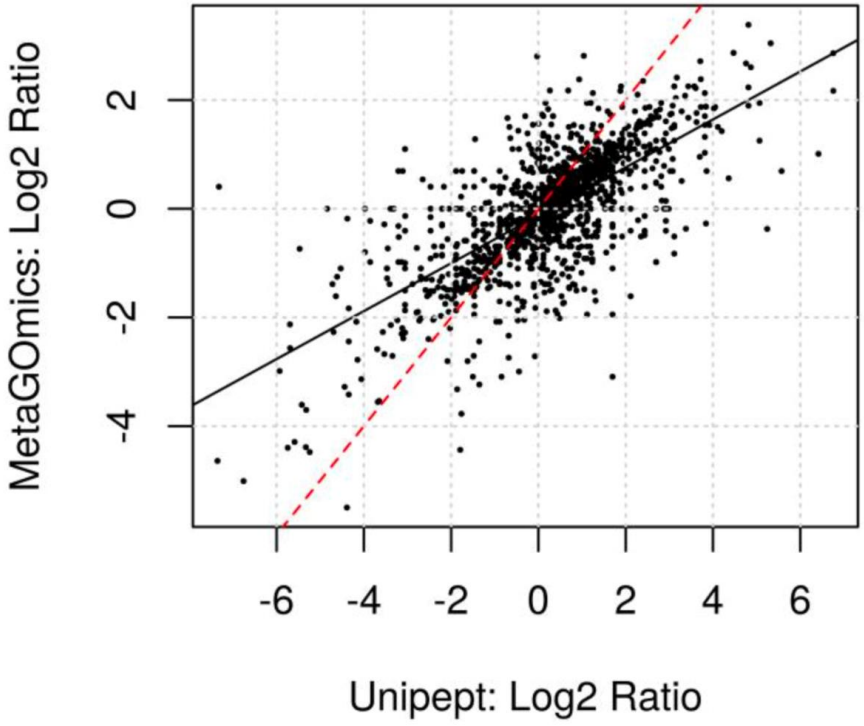
Sucrose-induced dysbiosis dataset



Fold-change Correlation:

Sucrose-induced dysbiosis dataset

1625 GO terms
80% of Unipept
41% of MetaGOmics

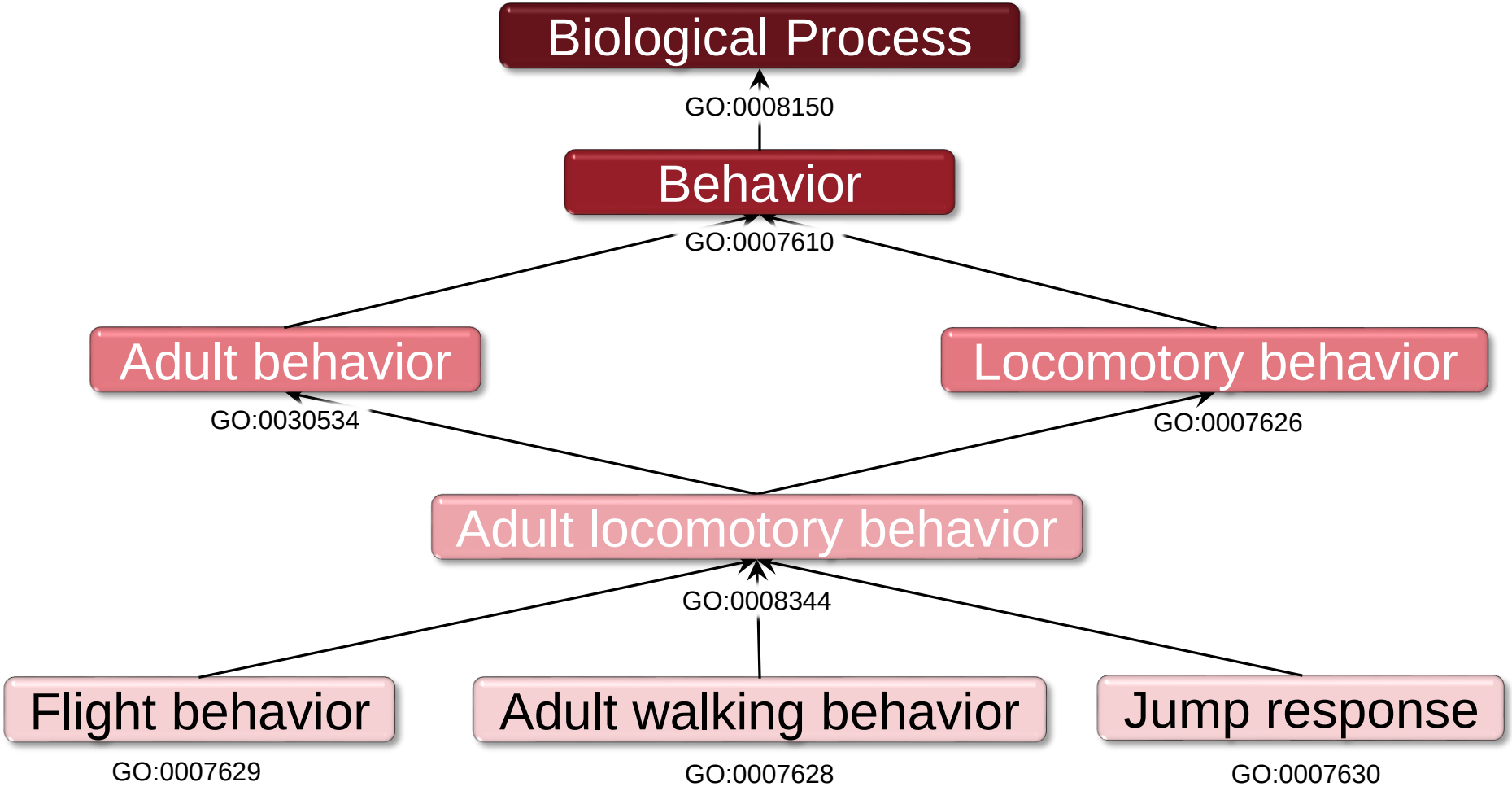


Pearson correlation coefficient
0.693 ($p < 10^{-16}$)

Differentially Expressed Terms

Sucrose-induced dysbiosis dataset

Rank	**eggNOG mapper** *(eggNOG proteins)*	**MEGAN** *(eggNOG orthologous groups)*	**MetaGOmics** *(GO terms)*	**MetaProteomeAnalyser** *(UniProtKB proteins)*	**Unipept** *(GO terms)*
1	Peptidase propeptide and YPED domain	Streptococcal surface antigen repeat	Pyruvate oxidase activity	Ferritin	Peptide deformylase activity
2	Orn lys arg decarboxylase	Acetolactate synthase	Oxidoreductase activity, acting on the aldehyde or oxo group of donors, oxygen as acceptor	Non-heme iron-containing ferritin	PFK-1 activity
3	Glycosyl hydrolase family 70	Catalyzes the condensation of the acetyl group of acetyl-CoA with 3-methyl-2-oxobutanoate to form 3-carboxy-3- hydroxy-4-methylpentanoate	Serine-type endopeptidase inhibitor activity	Clp protease ClpX	D-tagatose 6-phosphate catabolic process
4	DNA protection during starvation protein	Beta-hexosaminidase	Response to wounding	Chaperone protein DnaK	Tagatose-6-phosphate kinase activity
5	Oxidoreductase required for the transfer of electrons from pyruvate to flavodoxin	Cell wall	Poly(ribitol phosphate) teichoic acid metabolic process	ATP-dependent protease ATP- binding subunit	Response to wounding


Biological Process
GO:0008150
Behavior
GO:0007610
Adult behavior
GO:0030534
Locomotory behavior
GO:0007626
Adult locomotory behavior
GO:0008344
Flight behavior
GO:0007629
Adult walking behavior
GO:0007628
Jump response
GO:0007630

Single Peptide Analysis

Sucrose-induced dysbiosis dataset

AFLPGSLVDTRPVR

Single Peptide Analysis

BLASTP against UniProtKB

Parameters
- Target database: UniProtKB
- E-Threshold: 10
- Matrix: Auto
- Filtering: None
- Gapped: Yes
- Hits: 50

EggNOG Mapper in Galaxy

Parameters:
- Annotation type: DIAMOND
- Scoring matrix and gap costs: PAM30, 9 and 1
- Taxonomic Scope: Bacteria
- Orthologs: use all orthologs
- Gene Ontology evidence: use non-electronic terms
- Seed Orthology Search Options
  - Min e-value: 200000
  - Min bit score: 20

MEGAN

No native GO term output
(In progress)

AFLPGSLVDTRPVR

MetaGOmics

Parameters:
- Parameters
- Target database: UniProtKB
- E-Threshold: 10
- Matrix: Auto
- Filtering: None
- Gapped: Yes
- Hits: 50

MetaProteomeAnalyzer

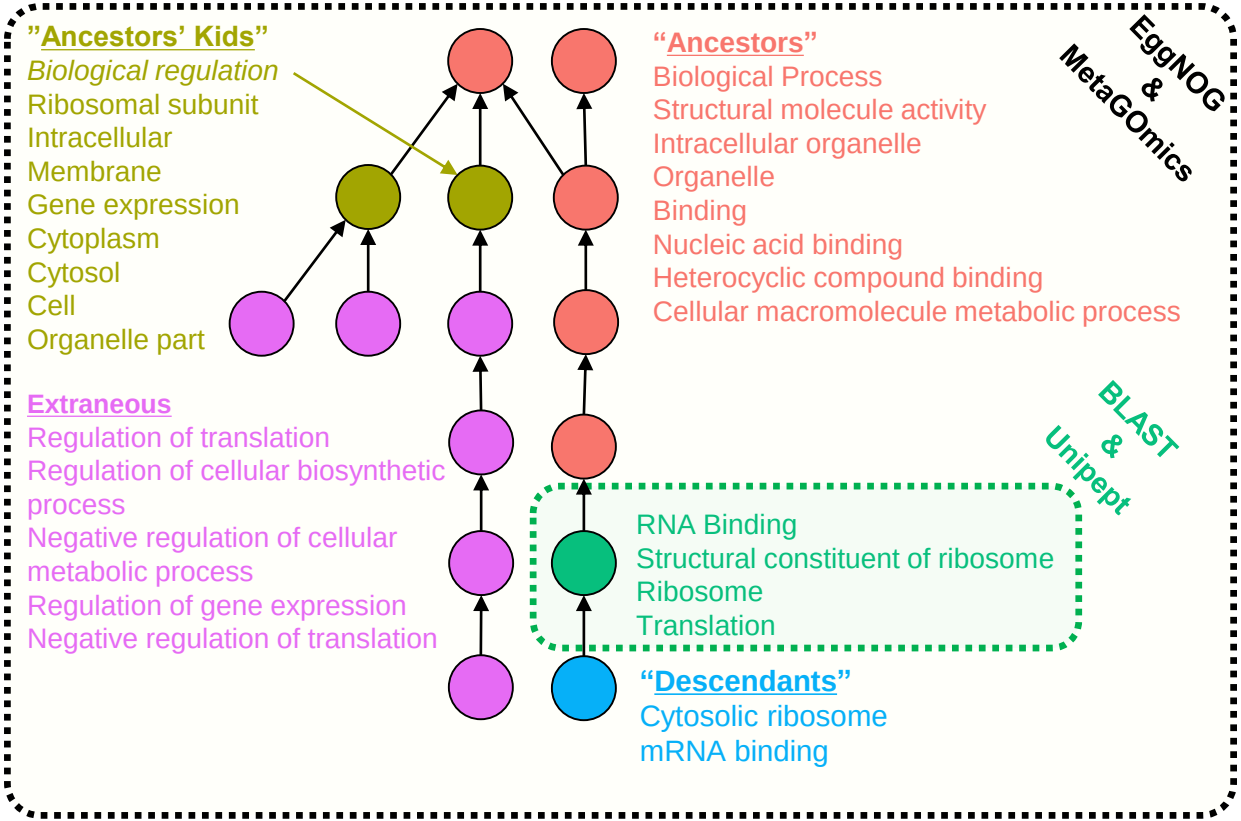
Matches to UniProt proteins
No native GO term output

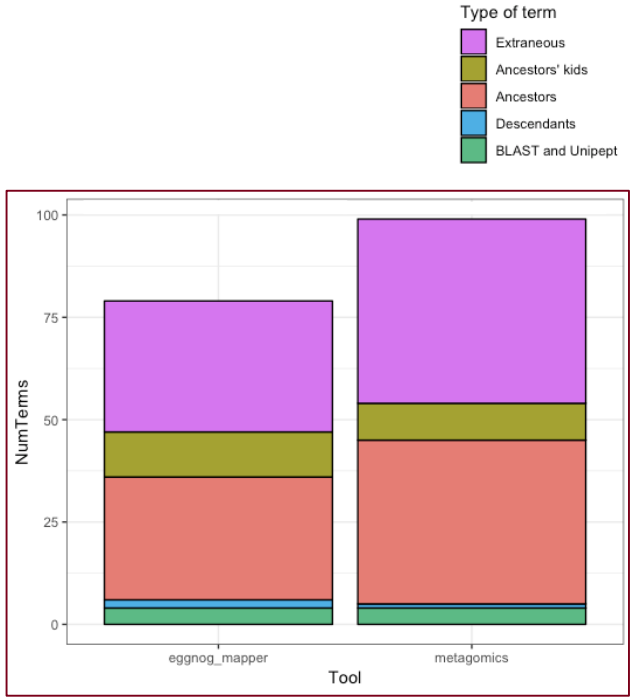
Unipept

Parameters:
- Equate I and L: FALSE
- Filter duplicate peptides: FALSE
- Advanced missed cleavage handling: TRUE

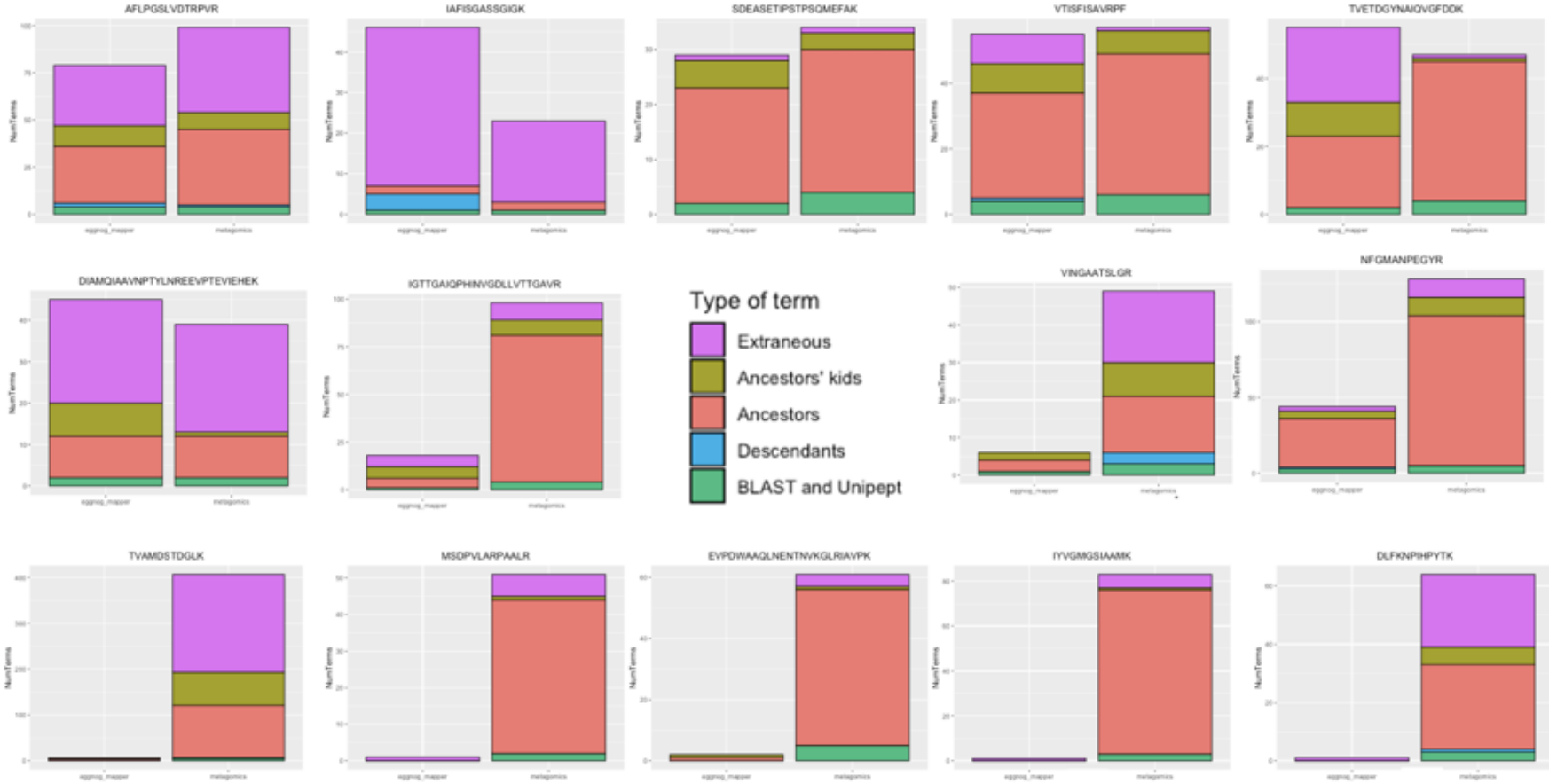
Single Peptide Analysis

AFLPGSLVDTRPVR (737 dataset)





Peptide Analysis Panel Sucrose-induced dysbiosis dataset



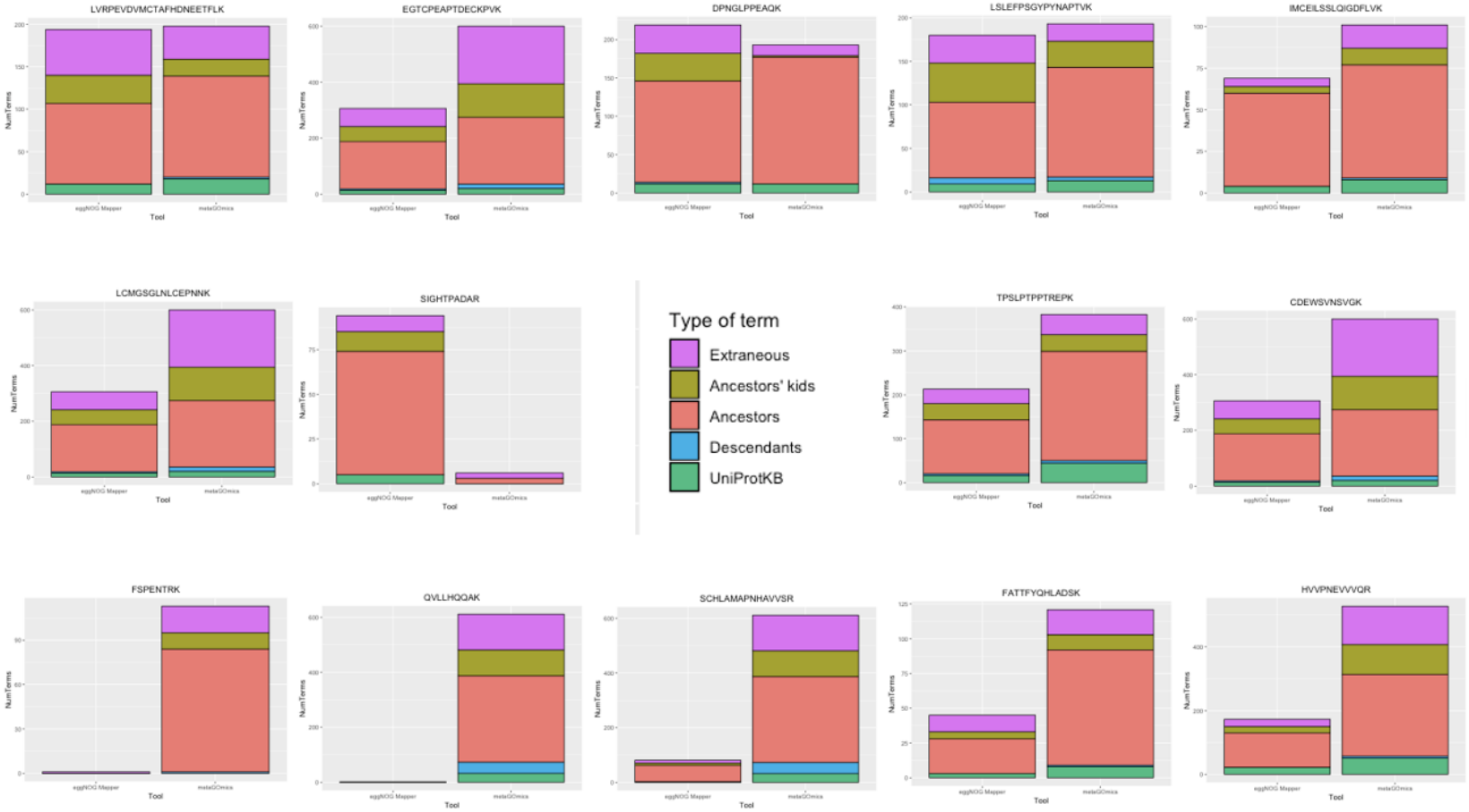
Universal Proteomic Standards (UPS)

Gamma-Synuclein (O76070)

- Adult locomotory behavior
- Chemical synaptic transmission
- Protein secretion
- Regulation of dopamine secretion
- Regulation of neurotransmitter secretion
- Synapse organization

UniProt Accession Number	UPS1 Amount (fmol)	UPS2 Amount (fmol)	UniProt Protein Name
P68871	5,000	50,000	Hemoglobin beta chain
P00167	5,000	5,000	Cytochrome b_5
P00709	5,000	500	Alpha-lactalbumin
O76070	5,000	50	Gamma-synuclein
P99999	5,000	5	Cytochrome c [Apocytochrome c]
P05413	5,000	0.5	Fatty acid-binding protein

Peptide Analysis Panel Universal Proteomics Standard



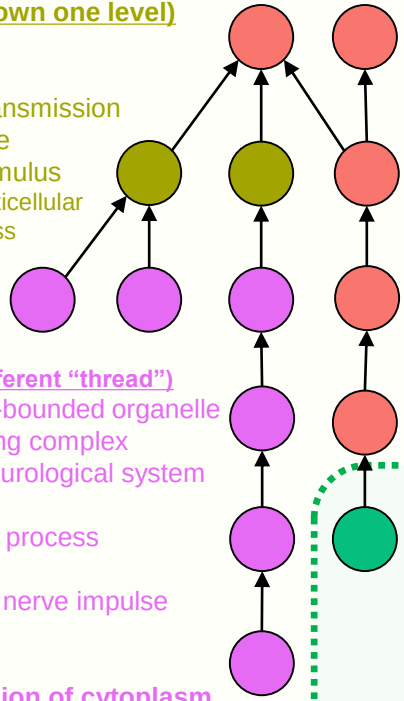
Single Peptide Analysis

EEVAEEAQSGGD (**UPS**)

Higher-level (down one level)
Organelle
Cell
Regulation of transmission of nerve impulse
Response to stimulus
Regulation of multicellular organismal process

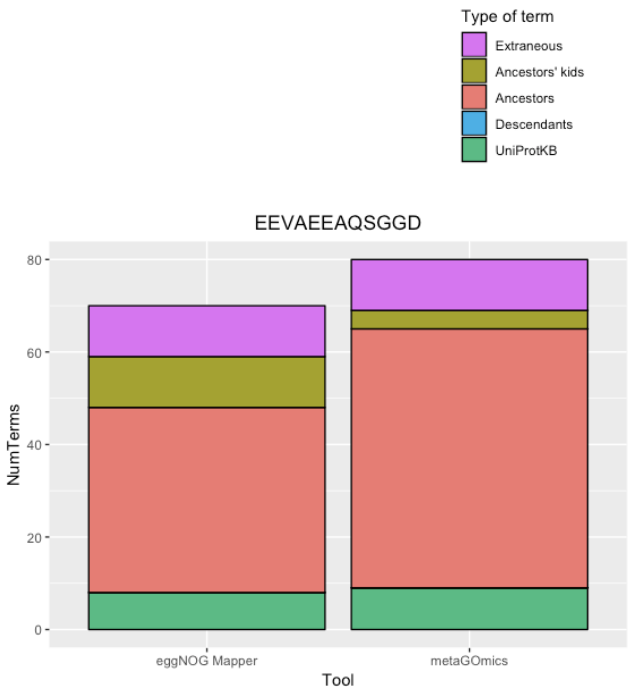
Higher-level terms
Biological Process
Reg
Intracellular organelle
Organelle
Binding
Nucleic acid binding
Heterocyclic compound binding
Cellular macromolecule metabolic process

EggNOG & MetaGOmics



Similar level (Different "thread")
Non-membrane-bounded organelle
Protein-containing complex
Regulation of neurological system process
Nervous system process
Behavior
Transmission of nerve impulse
Signaling
Spindle
Perinuclear region of cytoplasm (https://doi.org/10.1002/cm.1035)

UniProtKB
Protein secretion
Chemical synaptic transmission
Regulation of neurotransmitter secretion
Regulation of dopamine secretion
Axon
Neuronal cell body
Adult locomotory behavior
Cytoplasm
Synapse organization



Metaproteomics Software Tools

	EggNOG	MEGAN6	metaGOmics	MetaProteomeAnalyzer	Unipept
Operation System	macOS Linux	macOS Linux Windows	Web	MacOS Linux Windows	Web
Functional Visualization	Downstream processing required	Heatmaps PCA plots Hierarchical cluster analysis Tree diagrams Rarefaction curves	Static GO hierarchy colored by up/downregulation	Interactive Bar and Pie charts Tabular views for different functional Levels (e.g. EC, KO etc.)	Interactive treeview of E.C. numbers
Level of annotation	High-volume High-level Database specific	Low-volume	High-volume High-level	Low-volume	Medium-volume Similar to BLAST

Current Status

- **Metaproteomics studies lay an emphasis on measurement of functional response of a microbiome.**
- **Functional analysis tools have their distinct analytical methods; methods of annotation and quality – which is reflected in their outputs. Each tool also has its distinct functional visualization output. Developers are constantly improving these tools.**
- **It is recommended that functional tool be selected depending on the researcher's metaproteomics analytical needs.**
- **Current status of tools within Galaxy (December 2018):**
  - EggNOG Mapper already available in Galaxy
  - We intend to add Unipept 4.0 to Galaxy workflows to support taxonomy and functional analysis

ACKNOWLEDGMENTS

University of Minnesota

Timothy Griffin
Praveen Kumar
Candace Guerrero
Subina Mehta
Adrian Hegeman (Co-I)
Art Eschenlauer
Shane Hubler
Ray Sajulga
Caleb Easterly
Andrew Rajczewski

Minnesota Supercomputing Institute
James Johnson
Thomas McGowan
Michael Milligan

Brook Nunn
U of Washington

Biologists / collaborators
Laurie Parker
Joel Rudney
Maneesh Bhargava
Amy Skubitz
Chris Wendt
Brian Crooker
Steven Friedenberg
Kevin Viken
Kristin Boylan
Marnie Peterson
Somiah Afiuni
Brian Sandri
Alexa Pragman
Wanda Weber
Amy Treeful

Josh Elias
Stanford University

Lloyd Smith (Co-I)
Michael Shortreed
UW-Madison

Thomas Doak
Jeremy Fisher
Indiana University

Matt Chambers
Nashville, TN

Stephan Kang
Intero Life Sciences

Judson Hervey
Naval Research Institute
Washington, D.C.

Karen Reddy
Mo Heydarian
Johns Hopkins University

Lennart Martens (Co-I)
Bart Mesuere
Robbert G Singh
VIB, UGhent, Belgium

Harald Barsnes
Marc Vaudel
University of Bergen, Norway

Magnus Øverlie Arntzen (Co-I)
Francesco Delogu
NMBU, Oslo, Norway

Carolin Kolmeder
University of Helsinki, Finland

Thilo Muth
Bernhard Renard
Robert Koch Institut

Alessandro Tanca
Porto Conte Ricerche, Italy

Bjoern Gruening
Bérénice Batut
University of Freiburg, Freiburg, Germany

Anamika Krishanpal
Priyabrata Panigrahi
Persistent Systems Limited

Ira Cooke
Melbourne , Australia

Funding

NSF

NIH NATIONAL CANCER INSTITUTE
Informatics Technology for Cancer Research

GalaxyP
galaxyp.org