Survey of metaproteomics software tools for functional microbiome analysis

R. Sajulga[1], C. Easterly[1], M. Riffle[2], B. Mesuere[3], T. Muth[4], S. Mehta[1], P. Kumar[1], J. Johnson[1], B. Gruening[5], H. Schiebenhoefer[6], C. A. Kolmeder[7], S. Fuchs[8], B. L. Nunn[2], J. Rudney[1], T. J. Griffin[1], P. D. Jagtap[1]

[1]University of Minnesota, Minneapolis, MN; [2]University of Washington, Seattle, Washington; [3]Ghent University, Ghent, Belgium; [4]Federal Institute for Materials Research and Testing, Berlin, Germany; [5]University of Freiburg, Freiburg, Germany; [6]Robert Koch Institute, Berlin, Germany; [7]University of Helsinki, Helsinki, Finland; [8]Robert Koch Institute, Wernigerode, Germany.

Introduction

- Several software tools exist for analyzing microbiomes at the functional level by measuring their combined proteome-level response to environmental perturbations.
- In this survey, we explore the performance of six available tools, so that researchers can make informed decisions regarding software choice based on their research goals.
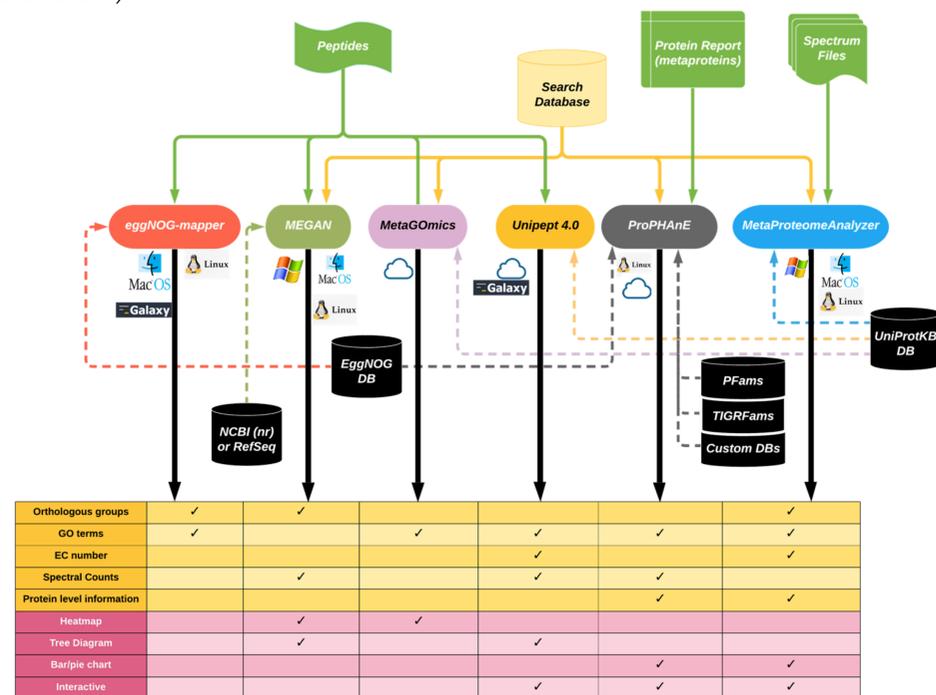
Functional Tools

- Eggnog-mapper
- MEGAN6
- MetaGOmics
- MetaProteomeAnalyzer (MPA)
- ProPHAnE
- Unipept 4.0

Method

Used each software tool to analyze the same dataset.

- **Dataset:** Tandem mass spectrometry-based proteomic data obtained from dental caries plaque samples grown with and without sucrose in paired biofilm reactors.
- **Pre-Processing:** Microbial peptides from one sample pair were identified by the X! Tandem search algorithm via SearchGUI and subjected to functional analysis using the six software tools.
- **Post-Processing:** The outputs from each tool are compared and converted to **Gene Ontology (GO)** terms for further comparison.

These six tools are unique in their inputs (input files, databases) and outputs (tabular and visualizations):
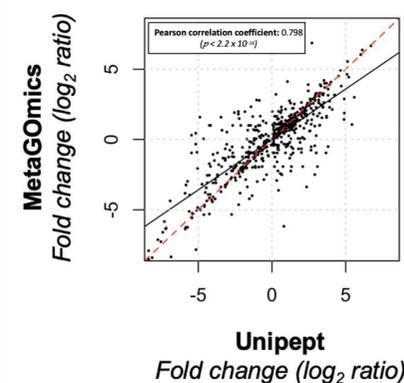

Results

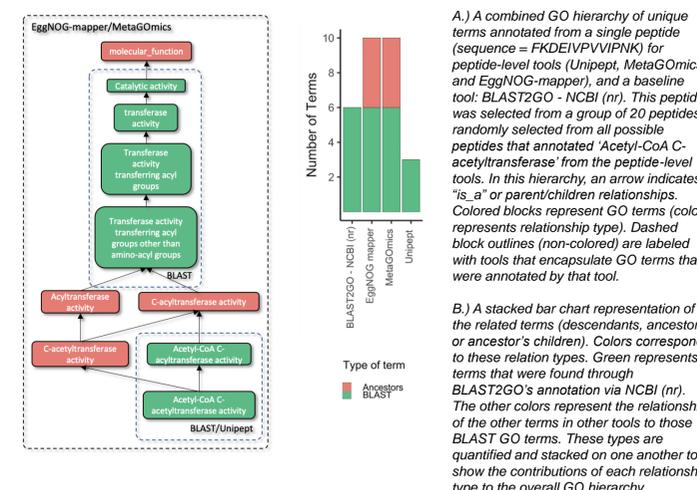*Functional analysis of the oral dysbiosis dataset using molecular function GO terms.*

Tool	EggNOG mapper	MEGAN	MetaGOmics	MPA	ProPHAnE	Unipept
Type of functional annotation	Proteins	eggNOG orthologous groups	GO terms	Proteins	Protein families, eggNOG orthologous groups	GO terms
Total number of annotation groups	18,440	1,665	2,829	23,169	3,999	3,471
Total (and unique) number of translated GO terms for all ontologies	533,066 (6,155)	76,529 (4,155)	2,829 (2,829)	77,204 (1,056)	189,054 (2,598)	3,471 (3,471)
Total (and unique) number of translated molecular function GO terms	88,582 (1,411)	21,212 (1,613)	900 (900)	42,084 (634)	57,208 (1,057)	1,726 (1,726)
Total number of molecular function GO terms exclusive to the tool	265	168	113	1	16	394
Total number of molecular function expanded GO terms	1,466	1,693	1,002	974	1,135	2,249
Total number of molecular function expanded GO terms exclusive to the tool	204	118	80	1	11	447

- Each software tool has differing levels of inputs in terms of type, amount, and exclusivity.

A.) A combined GO hierarchy of unique terms annotated from a single peptide (sequence = FKDEIVPVVIPNK) for peptide-level tools (Unipept, MetaGOmics, and EggNOG-mapper), and a baseline tool: BLAST2GO - NCBI (nr). This peptide was selected from a group of 20 peptides randomly selected from all possible peptides that annotated 'Acetyl-CoA C-acetyltransferase' from the peptide-level tools. In this hierarchy, an arrow indicates "is_a" or parent/children relationships. Colored blocks represent GO terms (color represents relationship type). Dashed block outlines (non-colored) are labeled with tools that encapsulate GO terms that were annotated by that tool.

B.) A stacked bar chart representation of the related terms (descendants, ancestors, or ancestor's children). Colors correspond to these relation types. Green represents terms that were found through BLAST2GO's annotation via NCBI (nr). The other colors represent the relationship of the other terms in other tools to those BLAST GO terms. These types are quantified and stacked on one another to show the contributions of each relationship type to the overall GO hierarchy.

- For the four peptide-level functional tools, there are differences in GO term annotation when compared with a BLAST2GO NCBI (nr) tool for benchmarking.
- The sets of unique GO terms annotated from these four tools would be the same if all the ancestors for each GO term were included.
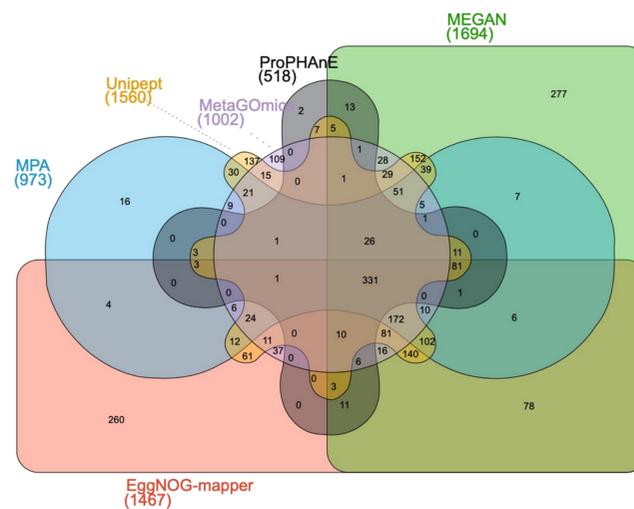

Comparison of quantitative expression for molecular function GO terms from Unipept and MetaGOmics. Log2ratio of spectral counts 'with sugar sample' (WS) against 'no sugar sample' (NS) was calculated for MetaGOmics- and Unipept-generated molecular function GO terms. Unipept identified 1,109 molecular function GO terms, while MetaGOmics identified 900 molecular function GO terms. The data points in figure represent quantitative values for 460 molecular function GO terms that overlapped between Unipept and MetaGOmics.

- Correlation between two tools (MetaGOmics and Unipept) that use the same database have a Pearson correlation coefficient of 0.798.

An Edwards-Venn diagram that illustrates all intersection combinations between the unique GO term sets (molecular function) of all six functional tools.

- Despite using the same dataset, the sets of gene ontology (GO) terms from each functional tool do not have uniform overlap.

Conclusion

- Functional software have notable differences in their required inputs and outputs.
- Access to different tools can help generate the right balance of annotation and results processing.
- Therefore, two functional analysis tools with distinct approaches related to database searching and filtering of results for usage were integrated into Galaxy and are available in the ToolShed.

Unipept

eggNOG-mapper

Acknowledgements