

# iMETAQUANTOME WORKFLOW: AN INTEGRATED METAPROTEOMICS WORKFLOW FOR INTERACTIVE, STATISTICAL AND FUNCTIONAL MICROBIOME ANALYSIS

Subina Mehta<sup>1</sup>; Ray Sajulga<sup>1</sup>; Caleb W Easterly<sup>1</sup>; Francesco Delogu<sup>2</sup>; Benoit J Kunath<sup>2</sup>; Praveen Kumar<sup>1</sup>; Marie Crane<sup>3</sup>; Emma Leith<sup>1</sup>; James E. Johnson<sup>1</sup>; Thomas McGowan<sup>1</sup>; Joel Rudney<sup>1</sup>; Phil B Pope<sup>2</sup>; Magnus Ø Arntzen<sup>2</sup>; Timothy J. Griffin<sup>1</sup>; Pratik D Jagtap<sup>1</sup>

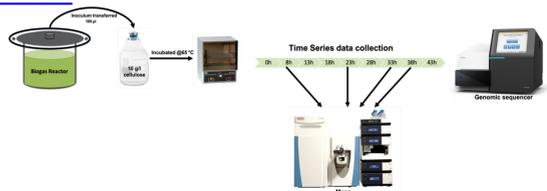
1. University of Minnesota, Minneapolis, MN; 2. NMBU - Norwegian University of Life Sciences, Ås, Norway; 3. Macalester College, Saint Paul, MN

## INTRODUCTION

- Metaproteomics is the study of proteins expressed by a community of organisms. Metaproteomics and metatranscriptomics, in particular, are used to determine the microbiome's functional state along with its taxonomy.
- Current metaproteomics workflows are focused on identification of taxonomy and microbiome functions. We believe that adding quantitative information along with taxonomy-function interaction will offer a deeper insight into microbiome dynamics.
- As a solution, the Galaxy-P team has developed a comprehensive workflow – seamlessly integrating protein identification tools with quantitation, taxonomy and functional annotation and interactive visualization.
- The imetaQuantome workflow analyzes the state of a microbiome by leveraging complex taxonomic and functional hierarchies to summarize peptide-level quantitative information.

## DATASET, TOOLS & WORKFLOWS.

### DATASET



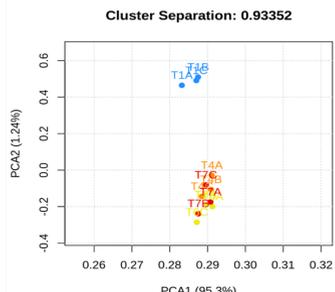
- 100 µl of inoculum of an enriched community from a biogas reactor was transferred to 27 anaerobic bottles containing the rich ATCC medium 1943 with cellobiose substituted for 10g/L of cellulose and incubated at 65 °C.
- Three bottles were collected at 9 different time points (0, 8, 13, 18, 23, 28, 33, 38 and 43 h) and processed in triplicates. Metatranscriptomic analysis was performed on all time points.

### METHODS

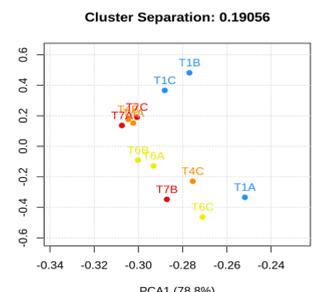
- Metatranscriptomics data was analyzed using ASaiM workflow and a metaproteomic database was created using Graph2Pro workflow (see poster # TP438)
- Peptide spectral matching (PSM) was performed by using SearchGUI (v3.3.10.1) and Peptideshaker (v1.16.36.2) software. The peptides obtained from the PSM report was quantified using FlashLFQ (v0.1.111) and taxonomy and functional annotation was performed using Unipept 4.0 software.
- The intensity output from the FlashLFQ was further normalized using limma tool.
- Tabular outputs: Intensity file (FlashLFQ), Taxonomy (Unipept) and Function (Unipept) were used as inputs for imetaQuantome.
- Most of the data processing and interpretation was done using the tools and workflows in the Galaxy Platform.
- NOTE:** The intensity of all proteins from all species associated with a specific function are summed to calculate the abundance of that function.

	T1 (8hr)	T4 (23hr)	T6(33hr)	T7(38hr)
# PSM	~330,000	~380,000	~350,000	~390,000
# Peptides	~66,000	~78,000	~81,000	~96,000
# Proteins	9,147	10,883	10,248	8,571
Genera Identified	540	590	606	694
GO terms	18,070	18,295	18,098	18,840

### FUNCTION: PCA PLOT

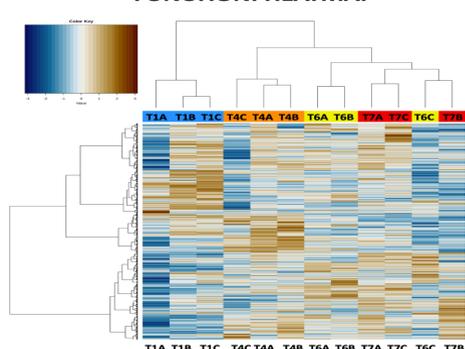


### TAXONOMY: PCA PLOT

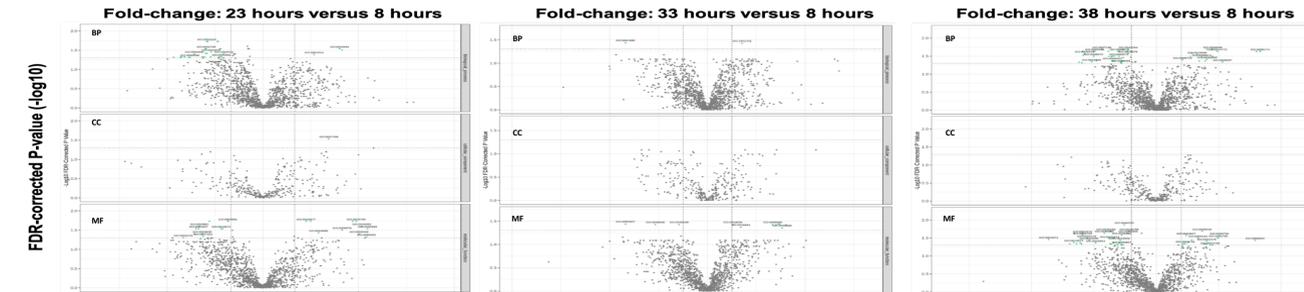


- metaQuantome visualization of the timepoints T1 (8hr), T4 (23hr), T6 (33hr) and T7 (38hr) (along with replicates A,B,C) in a PCA plot.
- Functional abundance values separate time point T1 (8 hr) from other time points (which are clustered together). Taxonomy abundance alone does not separate the time points, thus highlighting the importance of understanding functional state of the microbiome.

### FUNCTION: HEATMAP

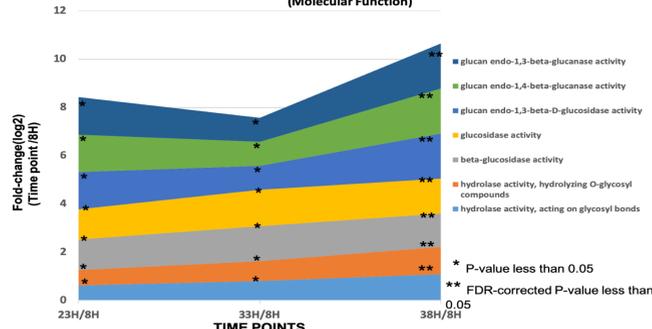


### FUNCTION: VOLCANO PLOTS



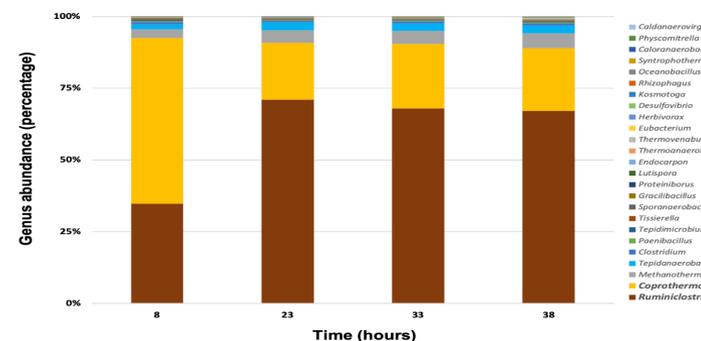
metaQuantome visualization of the timepoints T1 (8hr), T4 (23hr), T6 (33hr) and T7 (38hr) in a volcano plot (GO terms molecular function, cellular component and biological function). Plots are generated using fold-change values of each time with T1 to determine the differentially expressed GO term. Fold change values and the P-values associated helped in identification of differentially expressed functional groups.

### Glucosidases (Molecular Function)



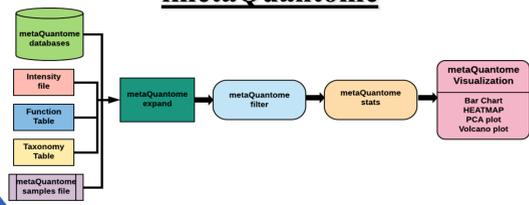
- Differentially expressed GO terms based on volcano plot output.
- The data shows that Glucosidases, - enzymes that break down complex carbohydrates – are upregulated from 8 hours to 38 hours.

### GENUS ABUNDANCE: TEMPORAL CHANGES

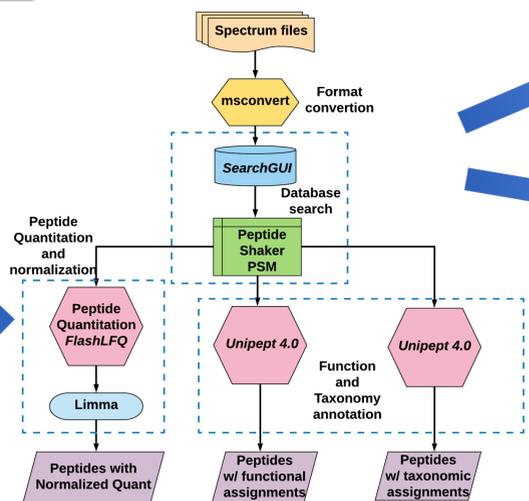


- MSI Quant outputs from FlashLFQ at peptide level were merged with Unipept taxonomy results.
- Coprothermobacter* and *Ruminiclostridium* were observed to be the most abundant genera.

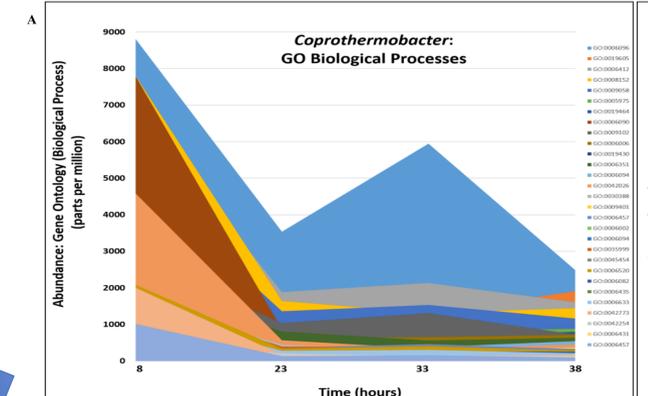
## imetaQuantome



## Metaproteomics Workflow

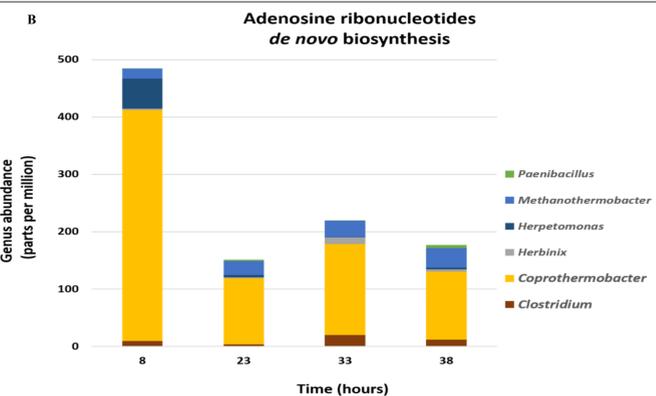


### FUNCTIONS ASSOCIATED WITH A SELECTED TAXON



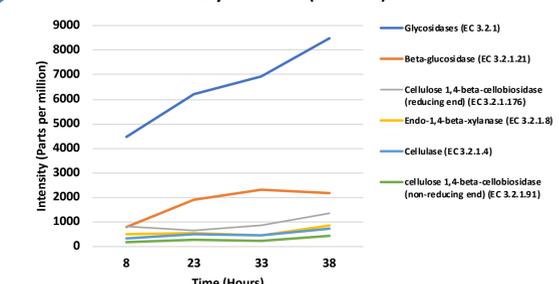
A. Glycolysis(GO:0006096) is observed to be the most abundant GO term across time points in *Coprothermobacter*.

### TAXA ASSOCIATED WITH A SELECTED FUNCTION



B. Contribution of genera to Adenosine ribonucleotides *de novo* biosynthesis across time points.

### Glycosidases (EC 3.2.1)



Quantitative analysis of proteins (EC numbers) shows upregulation of Glycosidases.

## CONCLUSIONS

- We demonstrate the use of a complete quantitative metaproteomics Galaxy workflow for taxonomy and functional quantification across time points on a large dataset.
- The imetaQuantome workflow includes peptide spectral matching and generation of quantitative taxonomic and functional outputs.
- Even a preliminary interpretation of results offers a deep insights into functional dynamics of the microbiome. Currently, we are in the process of adding interactive visualization capabilities to imetaQuantome workflow.

## ACKNOWLEDGEMENTS

- This work was supported by funding from grant NCI-ITCR grant 1U24CA199347, NSF grant 1458524 and 2018 Norwegian Centennial Chair Program Seed Grant.
- XSEDE research allocation BIO170096 and use of the Jetstream cloud-based computing resource for scientific computing (<https://jetstream-cloud.org/>) maintained at Indiana University. Support from the Minnesota Supercomputing Institute for maintenance and update of the Galaxy instances.
- Bjoern Gruening and the European Galaxy community for the help in the support during Galaxy implementation.
- Brook Nunn (University of Washington, Seattle, WA), Alessandro Tanca (Porto Conte Ricerche, Italy), Carolin Kolmeder and Nadia Szeinbaum (Georgia Tech, Atlanta, GA) for discussion during the development of imetaQuantome.

SEE POSTER TP438 FOR METATRANSCRIPTOMICS RESULTS