Metaproteomics Powered By Metatranscriptomics: Towards a Multi-omic Functional Microbiome Analysis

Pratik Jagtap[1]; Praveen Kumar[1]; Francesco Delogu[2]; Benoit J Kunath[2]; Sujun Li[3]; Marie Crane[4]; Subina Mehta[1]; Ray Sajulga[1]; Emma Leith[1]; James E. Johnson[5]; Yuzhen Ye[3]; Bérénice Batut[6]; Haixu Tang[3]; Phil B Pope[2]; Magnus Ø Arntzen[2]; Timothy Griffin[1]

1. Department of Biochemistry, Biophysics and Molecular Biology, University of Minnesota, Minneapolis, MN 55455 2. Faculty of Chemistry, Biotechnology and Food Science, NMBU, As, Norway; 3.School of Informatics, Computing, and Engineering, Indiana University,, Bloomington, IN; 4. Macalester College, Saint Paul, MN; 5. Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, MN; 6. Bioinformatics Group, University of Freiburg, Freiburg, Germany

INTRODUCTION

- Microbiomes play a critical role in health and disease in human hosts and in environmental ecosystems.
- Characterizing their microbial composition and functional role requires several complementary 'omics techniques. Metagenomics provides taxonomic information about the microbiome, while metatranscriptomics and metaproteomics provide functional information by quantifying microbial RNA and protein-expression levels, respectively.
- We have incorporated bioinformatics workflows within the Galaxy framework that use metatranscriptomics for improving metaproteomic results. We demonstrate the use of these workflows on a large multi-replicate, time course dataset.
- When combined, the outputs from these multi-omics methods can provide insights into the mechanistic nuances of complex microbial communities.

DATASET, TOOLS & WORKFLOWS.

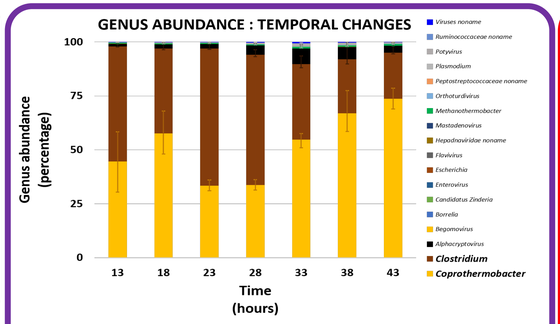
DATASET

- A 100 µl aliquot of an enriched community from a biogas reactor was transferred to 27 anaerobic bottles containing a rich medium and 10g/L of cellulose as sole carbon source 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

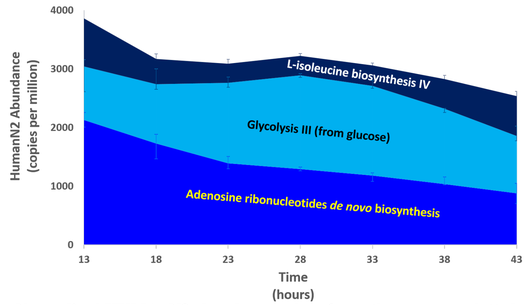
- The new and improved ASaiM workflow processes paired raw metatranscriptomics data and produces accurate and precise taxonomic and functional annotations and taxonomically related metabolism information.
- ASaiM workflow consists of:
  - FastQ Groomer and CutAdapt – for processing with quality control and trimming of adaptors.
  - SortmeRNA - Filtering out ribosomal RNA fragments from the metatranscriptomics data.
  - MetaPhlAn2 - For taxonomic assignment and categorization. For taxonomy visualization, KRONA and GraPhlAn were used.
  - HUMAnN2 - Functional analyses with metabolic assignment and pathway reconstruction.
  - Group abundance tool, for group abundances of gene families that were obtained from HUMAnN2
  - Functional and taxonomic interaction was determined by combining HUMAnN2 and MetaPhlAn2 outputs

GENUS ABUNDANCE : TEMPORAL CHANGES



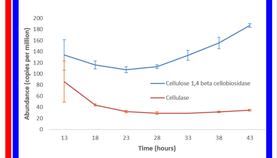
- Percentages of genus abundance from MetaPhlAn2 were merged for all time points. Error bars represent the standard error across the three replicates for each time point.
- *Coprothermobacter* and *Clostridium* were observed to be the most abundant.

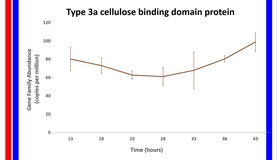
FUNCTIONAL PATHWAYS



- Normalized HUMAnN2 abundance values for pathways were plotted against time course data. Three pathways (out of many) have been represented here. Error bars represent the standard error across the three replicates for each time point.
- The stacked bar graph shows that the three pathways follow a distinct progression during the time course of cellulose degradation.

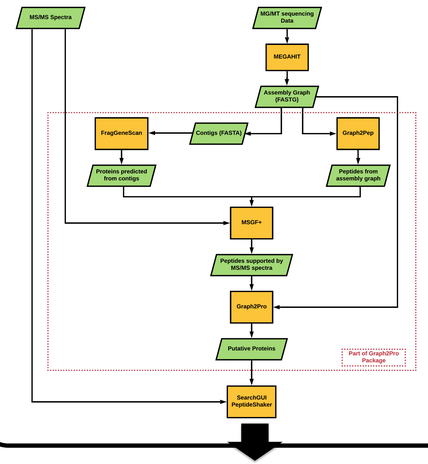
GENES INVOLVED IN CELLULOSE METABOLISM





- Quantitative analysis of gene family outputs from HUMAnN2 shows upregulation of cellulose.
- Gene encoding for the cellulose-binding domain protein shows an initial decrease and subsequent increase during cellulose degradation.

WORKFLOW TO GENERATE METAPROTEOMICS DATABASE



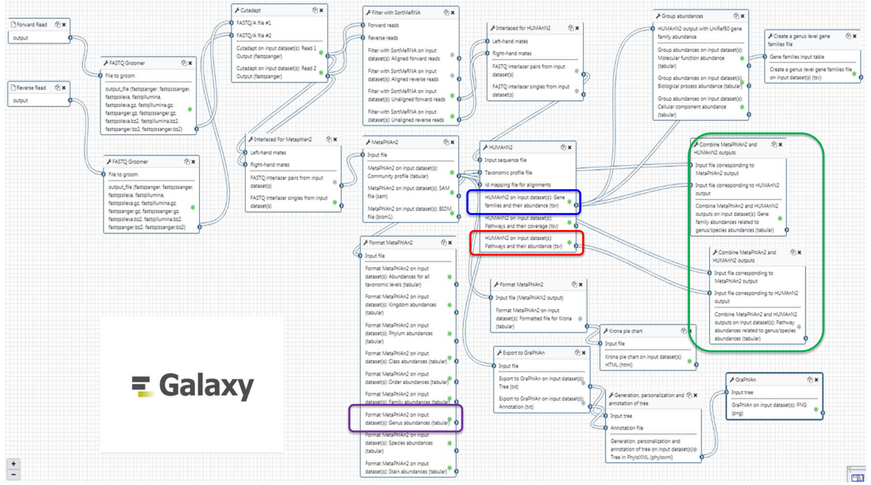
IDENTIFICATION STATISTICS FROM METATRANSCRIPTOMICS-DERIVED PROTEIN DATABASE SEARCHES.

	T1 (8hr)	T4 (23hr)	T6 (33hr)	T7 (38hr)
# Proteins in dB*	32,107	29,600	28,948	31,223
# Peptide-Spectral Matches	~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

** FASTQ File from Time point T2 (13 hours) was used to generate protein FASTA file since T1 FASTQ files were not available. The peak lists from the time points T1, T4, T6 and T7 were searched against Graph2Pro-generated non-redundant protein database with 109,538 protein entries.*

SEE POSTER # TP435 FOR METAPROTEOMICS RESULTS

WORKFLOW FOR METATRANSCRIPTOMICS ANALYSIS

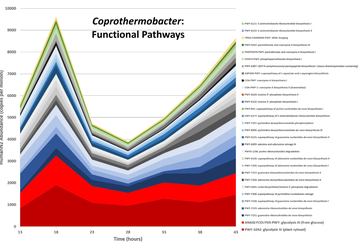


METATRANSCRIPTOMICS WORKFLOW OUTPUTS

	T2 (13hr)	T3 (18hr)	T4 (23hr)	T5 (28hr)	T6 (33hr)	T7 (38hr)	T8 (43hr)	Total
# Gene families	20,682	17,555	13,263	11,590	17,714	18,634	15,987	26,331
# Pathways	402	343	254	237	284	364	298	453
# Slim GO Terms (Molecular Function)	169	201	206	137	166	200	169	234
# Genera	12	13	12	12	12	9	12	18

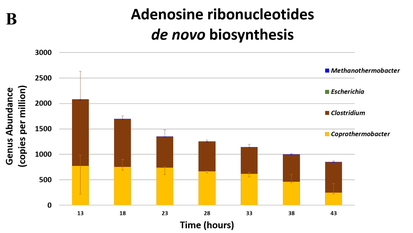
FUNCTIONS ASSOCIATED WITH A SELECTED TAXON

A



TAXA ASSOCIATED WITH A SELECTED FUNCTION

B



A. Glycolysis is observed to be the most abundant functional pathway across time points in *Coprothermobacter*.
B. Contribution of genera to adenosine ribonucleotides *de novo* biosynthesis across time points.

CONCLUSIONS

- We demonstrate the use of a quantitative metatranscriptomics Galaxy workflow for taxonomy and functional quantification across time points on a large dataset.
- The modified ASaiM workflow includes preprocessing of data before generation of quantitative taxonomic and functional outputs.
- Metatranscriptomics results on a cellulose degradation dataset offers deep insights into functional dynamics of the microbiome.
- We are currently in process of developing a tool – 'MUNDO' – which will perform statistical analysis and aids in visual interpretation of ASaiM outputs.

- ACKNOWLEDGEMENTS: *NCI-ITCR grant 1U24CA199347 ; NSF grant 1458524; 2018 Norwegian Centennial Chair Program Seed Grant; XSEDE research allocation BIO170096; European Galaxy Team.*