Adding Fun(ction) to Microbiome Analysis:
Metatranscriptomics & Metaproteomics Workflows in Galaxy

Saskia Hiltemann
Erasmus Medical Center & Erasmus University, Rotterdam
Subina Mehta
Pratik Jagtap
James Johnson
Timothy Griffin
University of Minnesota, Minneapolis
Table Mode
Presentation Mode

W-1-1-2-meta_workflows
Adding Fun(ct)ion to Microbiome Analysis: Metatranscriptomics and Metaproteomics Workflows in Galaxy

Chat  Participants  Q&A

Pratik Jagtap
I cannot hear or see anything.
8:24 AM | Today

Subina Mehta
probably log off and log in?
8:26 AM | Today

Pratik Jagtap
Got it!
8:28 AM | Today

Chat:

Saskia Hiltemann
View your results and discuss your table.
Ask your questions in chat or Google Doc.
[10 minutes]

Cam Off  Mic On  Share Screen  More  Quit Event
Microbiomes

**Microbiome**: Microbial genetic potential and response

Multiple studies have shown correlation of microbial composition with physiological conditions.
Potential to unravel the mechanistic details of microbial interactions with host / environment.
microbial taxa vary while metabolic pathways remain stable within a healthy population

Adding Fun(ction) to Microbiome Analysis: Metatranscriptomics & Metaproteomics Workflows in Galaxy

Functional microbiome analysis which estimates the functional groups expressed by microbial community, enables researchers to look beyond taxonomic composition and correlation with the condition under study. Using microbial community RNASeq data and subsequent metatranscriptomics workflows to elucidate the functional complement of the microbiome is gaining interest in the field. The hands-on training workshop will introduce researchers to the basic concepts and tools from the published ASaiM workflow (Batut et al. Gigascience (2018) doi: 10.1093/gigascience/giy057.). Attendees will also be able to run workflows on small test datasets. Lastly, workshop trainers will update attendees on the latest developments in Galaxy tools and workflows for functional microbiome and multi-omics analysis.

We will be following a tutorial available on the Galaxy Training Network (GTN), a unique, community-based training resource for Galaxy tools and workflows across many applications.

Get set up
Step 1: Log into Galaxy

1. Go to Galaxy: https://proteomics.usegalaxy.eu
2. Click on “Login or Register” in the top menu bar and log in to your account
Step 2: Join the workshop Queue

1. Visit this url to join the dedicated queue for the workshop: https://proteomics.usegalaxy.eu/join-training/gcc-meta/

1. You should see the following confirmation:
Introduction to Metatranscriptomics
These slides follow the tutorial on the GTN training website

https://training.galaxyproject.org

**Topic:** Metagenomics

**Tutorial:** Metatranscriptomics analysis using microbiome RNA-seq data (short)
Requirements
Before diving into this slide deck, we recommend you have a look at:

Introduction to Galaxy Analyses
Questions

• How to analyze metatranscriptomics data?

• What information can be extracted from metatranscriptomics data?

• How to assign taxa and function to the identified sequences?
Why study the microbiome?

- Health care research
  - Humans are full of microorganisms
  - Skin, gut, oral cavity, nasal cavity, eyes, ..
  - Affects health, drug efficacy, etc.
- Sometimes referred to as your **second genome**
  - ~10 times more cells than you
  - ~100 times more genes than you
  - ~1000s different species
Why study the microbiome?

- Environmental studies
  - Microbes in the soil affect plants and animals
  - Improve agriculture
This Tutorial: ASaiM pipeline

• Quality Control
  • Assess Quality
  • Trim and Filter raw reads
  • Filter ribosomal RNA (rRNA)

• Community profiling (Who?)
  • Determine composition of sample
  • Visualisation

• Functional Analysis (What?)

doi: 10.1093/gigascience/giy057
CELLULOSE DEGRADATION IN A BIOGAS REACTOR

Biogas-plant (60°C)  
Fredrikstad, Norway

Lab-scale reactor (55°C)

Anaerobic bottles (65°C)

Food waste  
Manure

Food waste  
Manure

Cellulose

Genomic sequencer

Mass Spectrometer

0h
8h
13h
18h
23h
28h
33h
38h
43h

T2
T4
T6
T7

Magnus Arntzen  
NMBU, Norway

DOI: 10.1038/s41396-018-0290-y
Let’s get started!
Step 1: Import Shared History

1. Go to Shared Data -> Histories at the top menu bar
Step 3: Import the History

1. Search for “metatranscriptomics”
2. Click on the history named “Input_Metatranscriptomics_GTN”
Step 3: Import the History

1. Search for “metatranscriptomics”
2. Click on the history named “Input_Metatranscriptomics_GTN”
Step 4: Import History

1. Click on the **plus** button at the top-right of the screen
Step 5: Name your History

1. Enter a title for your history
   a. Choose a name that makes sense to you, no wrong answers here
2. Click on the “Import” button
Step 6: Check the History

1. Check that your active history has 5 datasets in it, like below:
Workflow 1: Preprocessing
Data Preprocessing

1. Quality Control
   a. Assess Quality of reads
   b. Trim reads to improve quality
   c. Filter out ribosomal RNA

2. Format data for downstream analysis

We will now start the workflow, then explain each of these steps in more detail
Import Workflow?
Step 1: Import Shared Workflow

1. Go to “Shared Data -> Workflows” on top menu bar
Step 2: Import Workflow

1. Search for “metatranscriptomics”
2. Find the workflow named “Workflow 1: Preprocessing [Metatranscriptomics]”
3. Click on dropdown menu (triangle icon), and select “Import”
Step 3: Success!

1. You should see a green box like below
2. Click on “start using this workflow” to go to your workflow list

3. Alternative: Click on “Workflows” in top menu bar to view your workflow list
Step 4: Run Workflow

1. Find the workflow in your list
2. Select “Run” from the dropdown list (triangle icon next to name)
Step 4: Set Parameters

1. Make sure to select the correct FASTQ files for **forward** and **reverse**
2. Click on “**Run Workflow**”
The workflow steps
Input Format: FastQ Files

- Four lines per read

Example:

```
@FORJUSP02AJWD1
CCGTCATTCTAAGTTTAACCTTGCGGCCGTACTCCCCAGGCCGT+
AAAAAAAAAAAA:99:??@:FFAAAAACCAA:BB@@?A?
```
FastQ: Quality score

- Each character denotes a different Phred score

<table>
<thead>
<tr>
<th>ASCII code</th>
<th>33</th>
<th>59</th>
<th>64</th>
<th>73</th>
<th>104</th>
<th>126</th>
</tr>
</thead>
<tbody>
<tr>
<td>!&quot;#$%&amp;'()*+.-/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sanger</td>
<td>0</td>
<td>26</td>
<td>31</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solexa</td>
<td>-5</td>
<td>0</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illumina 1.3+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illumina 1.5+</td>
<td>0</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illumina 1.8+</td>
<td>26</td>
<td>31</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Phred scores are logarithmic

<table>
<thead>
<tr>
<th>Phred Quality Score</th>
<th>Probability of incorrect base call</th>
<th>Base call accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1 in 10</td>
<td>90%</td>
</tr>
<tr>
<td>20</td>
<td>1 in 100</td>
<td>99%</td>
</tr>
<tr>
<td>30</td>
<td>1 in 1000</td>
<td>99.9%</td>
</tr>
<tr>
<td>40</td>
<td>1 in 10,000</td>
<td>99.99%</td>
</tr>
</tbody>
</table>
Preprocessing
In this tutorial we start with some preprocessing steps
Preprocessing Tools
In this tutorial we start with some preprocessing steps
Quality Reports: FastQC
Generate a web report with quality metrics of your FastQ file
Quality Reports:
FastQC
Many different QC plots
Example: Per-base sequence Quality plot
Quality Reports:
FastQC
Many more plots
See QC tutorial for more information
Quality Reports: MultiQC

- Combine multiple FastQC reports into one report
- Also for outputs of other tools
- Great when sequencing large numbers of samples
Read Trimming and Filtering: Cutadapt

- Trim low-quality bases from reads
- Filter reads based on length, mean quality score,..
- Remove adapters/primers

Many tools:
- CutAdapt
- TrimGalore
- Trimmomatic
SortMeRNA

- Most RNA sequences will be ribosomal RNA (rRNA)
- Great for taxonomical assignment (who is there?)
- Not informative for functional analysis (What are they doing?)
- Filter out rRNA before doing functional analysis
FastQ interlacer

- Paired-end data often comes in two separate FastQ files
- One file with *forward* reads, one with *reverse* reads
FastQ interlacer

- Some tools require a single interlaced FastQ file
- Galaxy has tools for interlacing and deinterlacing FastQ files
View your results and discuss with your table

Ask your questions in chat or Google Doc

[10 minutes]
Workflow 2: Community Profiling
Community Profile

• We want to identify which organisms are present in our sample, and their relative abundances

• MetaPhlan2 tool for identification
• Krona tool and Graphlan tool for visualisation
Import Shared Workflow

1. Go to Shared Data -> Workflows
2. Search for “metatranscriptomics”
3. Import workflow “Workflow2: Community Profile [Metatranscriptomics]”
Run Workflow

1. Go to Workflows
2. Run “Workflow2: Community Profile [Metatranscriptomics]”
3. Check input files “Interlaced non rRNA reads” is selected
4. Click “Run Workflow”
The workflow steps
MetaPhlAn2

- Estimates the presence and relative abundance of microbial cells
- Maps reads against a set of marker sequences
- Caveat: this tool is designed for DNA-seq
  - Be careful interpreting abundances when using this tool with transcriptomics data
Krona

Visualization of community composition, interactive plot
GraPhlAn

Cladogram visualization
Genus Abundance

Tutorial: one timepoint

Figure Over multiple timepoints
View your results and discuss with your table

Ask your questions in chat or Google Doc

[5 minutes]
Workflow 3: Functional Annotation
Function

- PATHWAYS
- GENE ONTOLOGY
  - BIOLOGICAL PROCESS
  - MOLECULAR FUNCTION
  - CELLULAR COMPONENT
- GENE FAMILY
Galaxy Workflow

1) Go to ‘Shared Data’ and click on ‘Histories’

2) Search for ‘galaxyp’ tagged histories and click on ‘WF4:Functional information’

3) Import the history using the ‘+’ sign
Import Shared Workflow

1. Go to Shared Data -> Workflows
2. Search for “metatranscriptomics”
Run Workflow

1. Go to **Workflows**
2. Run **“Workflow3: Functional Information (quick)[Metatranscriptomics]”**
3. Check input files **“T1A_humann2_gene_family_abundance.tsv and T1A_humann2_pathway_abundance.tsv”** is selected
4. Click **“Run Workflow”**
Workflow

Input from SortmeRNA
- Interlaced non-rRNA reads

Input from MetaPhlAn2
- Community Profile

HUMAnN2

Renormalize
- Gene families
- Pathways

Group Abundances

Unpack pathway abundance to show genes included
HUMAnN2

- Profiles presence/absence and abundance of microbial community.
- Efficiently characterizes microbial metabolic pathways.

**Input:**
- Interlaced non-rRNA reads
- Taxonomic profile (MetaPhlAn2)

**Output:**
- Gene families and their abundance
- Pathways and their coverage
- Pathways and their abundance
HUMAnN2 Tiered Search

1.) Meta’omic sequences (DNA/RNA reads)

2.) Initial screen through MetaPhlAn2 → known microbial species
   - Database: merging pangenomes of identified species

3.) Nucleotide-level mapping against database

4.) Unaligned reads from #2 searched against protein DB (UniRef90 or UniRef50) through accelerated translated search

Result: Gene family and pathway abundances

Figure 1 - Nat Methods; Franzosa et al., 2018
Gene Families Abundances

<table>
<thead>
<tr>
<th>Gene Family</th>
<th>Abundance-RPKs</th>
<th>Abundance-RPKs</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNMAPPED</td>
<td>931987000000000</td>
<td></td>
</tr>
<tr>
<td>UniRef50_962593: Beta-lactamase TEM</td>
<td>487671563258074</td>
<td>487671563258074</td>
</tr>
<tr>
<td>UniRef50_962593: Beta-lactamase TEMlg_Clostridium_s Clostridium_thermocellum</td>
<td>42205747425928</td>
<td>42205747425928</td>
</tr>
<tr>
<td>UniRef50_RSFV01</td>
<td></td>
<td>392732031560915</td>
</tr>
<tr>
<td>UniRef50_RSFV01: Clostridium_s Clostridium_thermocellum</td>
<td>392732031560915</td>
<td>392732031560915</td>
</tr>
<tr>
<td>UniRef50_A3OC67</td>
<td>331734939517261</td>
<td>331734939517261</td>
</tr>
</tbody>
</table>

RPK (reads per kilobase) = sum of alignment scores
Gene Families to Functional Annotation

- MetaCyc Reactions
- KEGG Orthogroups
- Pfam domains
- Enzyme commission (EC) categories
- Gene Ontology (GO)
- Informative GO
- Slim GO
Group Abundances

humann2_regroup_table

Group HUMAnN2 to GO slim terms
Gene Families to Functional Annotation

Group HUMAnN2 to GO slim terms

<table>
<thead>
<tr>
<th>Gene Family</th>
<th>GO slim terms</th>
<th>Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>UniRef50</td>
<td>recr. activity</td>
<td>113.197</td>
</tr>
<tr>
<td></td>
<td>recr. activity</td>
<td>113.197</td>
</tr>
<tr>
<td></td>
<td>nucleotide binding</td>
<td>45885.889</td>
</tr>
<tr>
<td></td>
<td>nucleotide binding</td>
<td>39725.409</td>
</tr>
<tr>
<td></td>
<td>nucleotide binding</td>
<td>7019.934</td>
</tr>
<tr>
<td></td>
<td>nucleotide binding</td>
<td>140.545</td>
</tr>
<tr>
<td></td>
<td>molecular function</td>
<td>3382.394</td>
</tr>
</tbody>
</table>
### Molecular Function

<table>
<thead>
<tr>
<th>GO id</th>
<th>GO name</th>
<th>Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0009690</td>
<td>phosphorylation</td>
<td>5338.993</td>
</tr>
<tr>
<td>GO:0009690</td>
<td>phosphorylation signal transduction system</td>
<td>6315.734</td>
</tr>
<tr>
<td>GO:0009705</td>
<td>carbohydrate metabolic process</td>
<td>7601.510</td>
</tr>
<tr>
<td>GO:0009705</td>
<td>carbohydrate metabolic process signal transduction</td>
<td>6349.320</td>
</tr>
<tr>
<td>GO:0009705</td>
<td>generation of precursor metabolites and energy</td>
<td>1442.602</td>
</tr>
<tr>
<td>GO:0009705</td>
<td>generation of precursor metabolites and energy</td>
<td>5977.512</td>
</tr>
</tbody>
</table>

### Biological Process

<table>
<thead>
<tr>
<th>GO id</th>
<th>GO name</th>
<th>Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0009757</td>
<td>cellular component</td>
<td>13272.618</td>
</tr>
<tr>
<td>GO:0009757</td>
<td>cellular component</td>
<td>9409.291</td>
</tr>
<tr>
<td>GO:0009757</td>
<td>cellular component</td>
<td>168.874</td>
</tr>
<tr>
<td>GO:0009757</td>
<td>extracellular region</td>
<td>867.015</td>
</tr>
<tr>
<td>GO:0009757</td>
<td>extracellular region</td>
<td>627.760</td>
</tr>
<tr>
<td>GO:0009757</td>
<td>extracellular region</td>
<td>20.262</td>
</tr>
<tr>
<td>GO:0009757</td>
<td>intracellular</td>
<td>35104.490</td>
</tr>
<tr>
<td>GO:0009757</td>
<td>intracellular</td>
<td>17564.816</td>
</tr>
<tr>
<td>GO:0009757</td>
<td>intracellular</td>
<td>23.022</td>
</tr>
<tr>
<td>GO:0009757</td>
<td>intracellular</td>
<td>13773.386</td>
</tr>
</tbody>
</table>

### Cellular Component

<table>
<thead>
<tr>
<th>GO id</th>
<th>GO name</th>
<th>Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0009757</td>
<td>cellular component</td>
<td>13272.618</td>
</tr>
<tr>
<td>GO:0009757</td>
<td>cellular component</td>
<td>9409.291</td>
</tr>
<tr>
<td>GO:0009757</td>
<td>cellular component</td>
<td>168.874</td>
</tr>
<tr>
<td>GO:0009757</td>
<td>extracellular region</td>
<td>867.015</td>
</tr>
<tr>
<td>GO:0009757</td>
<td>extracellular region</td>
<td>627.760</td>
</tr>
<tr>
<td>GO:0009757</td>
<td>extracellular region</td>
<td>20.262</td>
</tr>
<tr>
<td>GO:0009757</td>
<td>intracellular</td>
<td>35104.490</td>
</tr>
<tr>
<td>GO:0009757</td>
<td>intracellular</td>
<td>17564.816</td>
</tr>
<tr>
<td>GO:0009757</td>
<td>intracellular</td>
<td>23.022</td>
</tr>
<tr>
<td>GO:0009757</td>
<td>intracellular</td>
<td>13773.386</td>
</tr>
</tbody>
</table>

The output shows the distribution of GO terms across different categories: Molecular Function, Biological Process, and Cellular Component, with their respective counts of abundance.
Unpack pathway abundances to show genes included

- Renormalize the gene and pathway abundance in copies per million or relative abundance.
- This tool unpacks the pathways abundance by including gene families.
FUNCTION: CELLULOSE DEGRADATION

- Quantitative analysis of gene family outputs from HumanN2 shows upregulation of cellulase.
FUNCTIONS ASSOCIATED WITH A SELECTED TAXON

Coprothermobacter: Functional Pathways

HumanN2 Abundance (copies per million)

Time (hours)

- PWY-6215: 5-aminoimidazole-4-carboxamide ribonucleotide biosynthesis
- PWY-6213: 5-aminoimidazole-4-carboxamide ribonucleotide biosynthesis
- TRNA CHARGING: PWY-6065 charging
- PWY-6245: pantotetone and coenzyme A biosynthesis
- PWY-6305: pantotetone and coenzyme A biosynthesis
- PWY-6695: 3-hydroxy-3-methylglutaric acid degradation
- ASLAM: PWY-6541: superpathway of l-aspartate and l-glutamate biosynthesis
- CDA-PHY-6542: coenzyme A biosynthesis
- CDA-PHY-6542: coenzyme A biosynthesis
- PWY-7145: Glutamate dehydrogenase
- PWY-6508: superpathway of guanine nucleotides de novo biosynthesis
- PWY-6105: adenine and adenosine salvage
- PWY-6105: adenine and adenosine salvage
- PWY-6105: adenine and adenosine salvage
- PWY-6105: adenine and adenosine salvage
- PWY-7150: guanine nucleotides de novo biosynthesis
- PWY-7150: guanine nucleotides de novo biosynthesis
- PWY-7150: guanine nucleotides de novo biosynthesis
- PWY-7150: guanine nucleotides de novo biosynthesis
- ARAGLYCOLYSIS-PHY: glycogenolysis II (from glucose)
- PWY-1042: glycogenolysis IV (plant cytosol)
TAXA ASSOCIATED WITH A SELECTED FUNCTION

Adenosine ribonucleotides

de novo biosynthesis

Genus Abundance (copies per million)

Time (hours)

- Methanothermobacter
- Escherichia
- Clostridium
- Coprothermobacter
TABULAR OUTPUTS FROM ASaiM WORKFLOW

TAXONOMY (WHO?)

• KINGDOM
• PHYLUM
• CLASS
• ORDER
• FAMILY
• GENUS
• SPECIES
• STRAIN

FUNCTION (WHAT?)

• PATHWAYS
• GENE ONTOLOGY
  • BIOLOGICAL PROCESS
  • MOLECULAR FUNCTION
  • CELLULAR COMPONENT
• GENE FAMILY

ABUNDANCE
View your results and discuss with your table

Ask your questions in chat or Google Doc

[10 minutes]
Metaproteomics
**Microbiome: Microbial genetic potential and response**

Multiple studies have shown **correlation of microbial composition with physiological conditions**. Also used to study **interaction with environment**.

**Metagenomics:** Identifies species present within complex community (16S rRNA and Whole Genome Sequencing).

DNA from samples. **16S rRNA** (economical) or **Shotgun sequencing** (expensive).

Multiple studies that **correlate taxonomy with observed phenotype**.

**Metaproteomics:**

The large-scale characterization of the entire protein complement of environmental microbiota at a given point in time.

Proteins from samples.

Potential to unravel the **mechanistic details of microbial interactions with host / environment** by analyzing the **functional dynamics of the microbiome**.
Metaproteomics: A primer

Peptide fractionation coupled to tandem mass spectrometry (MS/MS)
Matching amino acid sequences to MS/MS data

Raw MS/MS spectrum

Protein sequence and/or DNA sequence database search

Peptide sequence match

Protein identification

Direct identification of 1000s proteins from complex mixtures
**Metaproteomics**

**Bond and Wilmes 2004**

“The large-scale characterization of the entire protein complement of environmental microbiota at a given point in time”


**Bond and Wilmes 2015**

“Through the application of metaproteomics to different microbial consortia over the past decade, we have learnt much about key functional traits in the various environmental settings where they occur.”

Metaproteomics Workflow

**DATABASE GENERATION**
- FASTQ
- Protein / Peptide FASTA
- Spectra

**DATABASE SEARCH & STRATEGIES**
- Search Algorithm
- Peptides

**FUNCTIONAL ANALYSIS**
- Known Function
- Hypothetical Function
- Unknown Function
- Shared Taxonomy
- Unassigned Taxonomy
- Unique Peptides

**TAXONOMY ANALYSIS**
Metaproteomics Workflow

https://galaxyproject.github.io/training-material/topics/proteomics/tutorials/metaproteomics/tutorial.html

Disseminating Metaproteomic Informatics Capabilities and Knowledge Using the Galaxy-P Framework
Proteomes 2018, 6(1), 7; https://doi.org/10.3390/proteomes6010007
Metaproteomics Workflow

https://galaxyproject.github.io/training-material/topics/proteomics/tutorials/metaproteomics/tutorial.html

Disseminating Metaproteomic Informatics Capabilities and Knowledge Using the Galaxy-P Framework
Proteomes 2018, 6(1), 7; https://doi.org/10.3390/proteomes6010007
Metaproteomics Workflow

DATABASE GENERATION

FASTQ

Protein / Peptide FASTA

Search Algorithm

Spectra

DATABASE SEARCH & STRATEGIES

Peptides

Spectral counts OR Intensity data

QUANTITATIVE ANALYSIS

FUNCTIONAL ANALYSIS

Known Function

Proteins

Hypothetical Function

Unknown Function

Shared Taxonomy

Unassigned Taxonomy

Unique Peptides

TAXONOMY ANALYSIS
metaQuantome Workflow

PEPTIDE QUANTIFICATION

- Peptide Quantitation
  - FlashLFQ
  - Limma

- Peptides with Normalized Quant

PEPTIDE IDENTIFICATION

- msconvert
  - Format conversion

- SearchGUI
  - Database search

- Peptide Shaker PSM

FUNCTION TAXONOMY

- Unipept 4.0
  - Function and Taxonomy annotation

- Peptides w/ taxonomic assignments
metaQuantome OUTPUTS

metaproteomic data input

peptides

MS1 intensity

function annotation

taxonomic annotation

metaQuantome

data exploration

differential abundance

cluster analysis

FUNCTION: VOLCANO PLOTS

Fold-change: 33 hours versus 8 hours

FUNCTION: HEATMAP

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. Metaproteomics analysis on 4 data points.
metaQuantome (METAPROTEOMICS) WORKFLOW OUTPUTS

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

FUNCTION: PCA PLOT
Cluster Separation: 0.93352

TAXONOMY: PCA PLOT
Cluster Separation: 0.19056

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.
HEATMAP ANALYSIS

METATRANSCRIPTOMICs

Functional and taxonomic abundance values separate time point T4 (23 hr) from other time points.

METAPROTEOMICS

Functional abundance values cluster time points T4 (23 hr), T6 (28 hr) and T7 (38 hr), while taxonomy values do not.
**TAXONOMY: GENUS ABUNDANCE**

**metaQuantome approach**

- **Genus**: Other, Methanothermobacter, Thermoclostridium, Hungateiclostridium, Coprothermobacter

**Metagenome binning approach**

- **Hungateiclostridium**
- **Coprothermobacter strains**

*Hungateiclostridium and Coprothermobacter do not seem to change significantly during the time points under study.*
Seventy-nine GO terms were found to be differentially expressed in both timepoints T6 and T7 as compared to T4.
Functions Associated With Cellulose Degradation in Hungateiclostridium

**METATRANSCRIPTOMICS**
Cellulose degradation (*Hungateiclostridium*)

**METAPROTEOMICS**
Cellulose degradation (*Hungateiclostridium*)

![Graphs illustrating abundance over time](image-url)
Accessing tools and Workflows

METAGENOMICS:
Toolshed: z.umn.edu/metagenomics_toolshed
Galaxy Training Network: https://training.galaxyproject.org/training-material/topics/metagenomics/

METATRANSCRIPTOMICS:
Workflow: http://z.umn.edu/MTWF2020

METAPROTEOMICS:
Workflow: z.umn.edu/MPWF2020
Galaxy Training Network: http://z.umn.edu/gtn-metaproteomics

Also available on: https://proteomics.usegalaxy.eu/ and Metaproteomics Gateway: z.umn.edu/metaproteomicsgateway

galaxyp.org/contact
UNIVERSITY OF MINNESOTA
Timothy Griffin
Praveen Kumar
Subina Mehta
Ray Sajulga
Caleb Easterly
Marie Crane
Emma Leith
Adrian Hegeman (Co-I)
Art Eschenlauer
Mark Esler
Andrew Rajczewski
Candace Guerrero

MINNESOTA SUPERCOMPUTING INSTITUTE
James Johnson
Thomas McGowan
Michael Milligan

University of Minnesota
Timothy Griffin
Praveen Kumar
Subina Mehta
Ray Sajulga
Caleb Easterly
Marie Crane
Emma Leith
Adrian Hegeman (Co-I)
Art Eschenlauer
Mark Esler
Andrew Rajczewski
Candace Guerrero

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

ACKNOWLEDGMENTS

GalaxyP

Josh Elias
Stanford University

Harald Barsnes
Marc Vaudel
University of Bergen, Norway

Magnus Øverlie Arntzen
Francesco Delogu
NMBU, Oslo, Norway

Carolin Kolmeder
University of Helsinki, Finland

Thilo Muth
Bernhard Renard
Robert Koch Institut

Anamika Krishanpal
Priyabrata Panigrahi
Persistent Systems Limited

Bjoern Gruening
Bérénice Batut
University of Freiburg, Freiburg, Germany

Ira Cooke and Maria Doyle
Melbourne, Australia

Twitter.com/usegalaxy

galaxyp.org