METAPROTEOMICS: Promoting Functional Analysis of Microbiome

Pratik Jagtap
University of Minnesota
METAPROTEOMICS:
Promoting Functional Analysis of Microbiome

• METAPROTEOMICS AND FUNCTIONAL MICROBIOME ANALYSIS
• METAPROTEOMICS WORKFLOWS WITHIN GALAXY PLATFORM
• QUANTITATIVE METAPROTEOMICS
• TARGETED HOST-MICROBIOME ANALYSIS
• EDUCATION AND PROMOTION
Microbiome Research

In Numbers

100 Trillion
Symbolic microbes live in and on every person and make up the human microbiota.

95%
of our microbiota is located in the GI tract.

150:1
The human body has more microbes than there are stars in the milky way.

The genes in your microbiome outnumber the genes in our genome by about 150 to one.

The surface area of the GI tract is the same size as 1.3 more microbes than human cells.

The gut microbiota can weigh up to 2kg.

Each individual has a unique gut microbiota, as personal as a fingerprint.

Viruses: Bacteria
5:1
in the gut microbiota.

2.5
The number of times your body’s microbes would circle the earth if positioned end to end.

>10,000
Number of different microbial species that researchers have identified living in and on the human body.

You have
The gut microbiota can weigh up to 2kg.

No. 1310
You have 2.5 times the number of bacterial cells in your gut.

https://worldmicrobiomeday.com/resources/

https://www.nature.com/articles/d41586-020-00193-3
Potential to unravel the mechanistic details of microbial interactions with host / environment by analyzing the functional dynamics of the microbiome.
“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 Analytical Challenges

**Single-Organism Proteomics**

- **Search Database Size**: Small to medium size (10K to 100K sequences)
- **Single + Contaminants**

**Metaproteomics**

- **LARGE (1 million and above)**
- **Multi-Organism Database with Homologous Proteins**

## Search Algorithms Being Developed to Address Large and Complex Database Searches
- Disparate tools and multiple processing steps.
- **Protein Grouping at Multi-Organism Level**
- **Identification Statistics Affected by Large Databases**
- **Taxonomy Based on Unique Peptide Identifications**
- **Functional Analysis Based on Proteins Identified**
Metaproteomics Workflow

**DATABASE GENERATION**
- FASTQ
- FASTA
- Mass Spectrometry Data

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

**FUNCTIONAL ANALYSIS**
- Known Function
- Proteins

**TAXONOMY ANALYSIS**
- Unassigned
- Shared
- Unique

**Peptides**
- Known Function
- Proteins
- Unassigned
- Shared
- Unique

**Search Algorithm**
Software tools can be used in a sequential manner to generate analytical workflows that can be reused, shared and creatively modified.

Blank et al/ 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

**Quantitative Analysis**
- Spectral counts OR Intensity data

**Functional Analysis**
- Known Function
- Proteins
- Hypothetical Function
- Unknown Function
- Shared Taxonomy
- Unassigned Taxonomy
- Unique Peptides

Taxonomy Analysis
**metaQuantome**

metaQuantome enables quantitative analysis of the taxonomic and functional state of a microbiome.

metaQuantome on Galaxy Training Network

- https://training.galaxyproject.org/training-material/topics/proteomics/tutorials/metaquantome-data-creation/tutorial.html
- https://training.galaxyproject.org/training-material/topics/proteomics/tutorials/metaquantome-function/tutorial.html
- https://training.galaxyproject.org/training-material/topics/proteomics/tutorials/metaquantome-taxonomy/tutorial.html
Case Study: Sucrose-induced oral dysbiosis

Mass spectral data was acquired from plaque samples from twelve subjects at high risk for dental caries grown in biofilm reactor in the presence (With Sucrose, or WS) and absence of sucrose (No Sucrose, or NS) (12 in each group, 24 total samples)

Mass spectra were searched against the Human Oral Microbiome database (HOMD) to identify microbial peptides.

Quantitation, functional annotation, and taxonomic assignment was performed in Galaxy; metaQuantome was used to analyze the results.
Principal Component Analysis

**FUNCTION: PCA PLOT**

**TAXONOMY: PCA PLOT**
metaQuantome deciphers the contribution of taxa to a functional process.

**CARBOHYDRATE METABOLISM**

**Control**

**In presence of sugar**

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Proportion of peptide intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>In presence of</td>
<td></td>
</tr>
</tbody>
</table>

**SUGAR**

Sugar $\rightarrow$ Lactate $\rightarrow$ Acidity increases $\rightarrow$ *Fusobacterium* does not thrive in high acidic conditions

**CONTROL**

Carbohydrates $\rightarrow$ Weak acids $\rightarrow$ Low acidity $\rightarrow$ *Fusobacterium* thrives in mild acidic conditions
QUANTITATION METHODS

DIA-MS offers benefits for quantitative metaproteomics

DIA-MS has better proteome coverage

DIA-MS offers higher accuracy

DIA-MS offers quantitatative reproducibility

DIA-MS has lesser missing values

Sources:
Brook L. Nunn
Susan T. Weintraub
Workflow Overview for Peptides and Proteins Detection and Quantification.

Pooled samples (hard-spun pellets, n = 16/group)

DCL  DCH  CFL  CFH

Matching of MS/MS data to Large 16S rRNA guided sequence databases (MetaNovo)

Peptides and Proteins Detection & Quantitation

<table>
<thead>
<tr>
<th></th>
<th>PROTEINS</th>
<th>PEPTIDES</th>
</tr>
</thead>
<tbody>
<tr>
<td>FragPipe</td>
<td>5873</td>
<td>69,906</td>
</tr>
<tr>
<td></td>
<td>(129)</td>
<td>(335)</td>
</tr>
<tr>
<td>MQ</td>
<td>3803</td>
<td>36,409</td>
</tr>
<tr>
<td></td>
<td>(161)</td>
<td>(300)</td>
</tr>
<tr>
<td>16S rRNA</td>
<td>54,287</td>
<td>2292</td>
</tr>
<tr>
<td></td>
<td>(1879)</td>
<td>microbial peptides</td>
</tr>
</tbody>
</table>
Verified Microbial Peptides

- Peptides and Proteins Detection & Quantitation
- Microbial Peptide Verification
- Taxonomy & Functional Analysis
# Verified Microbial Peptides

## 87 Verified Microbial Peptides

### 24 Taxonomy-Specific Candidate Peptides
(Known taxonomy)

<table>
<thead>
<tr>
<th>Organism</th>
<th>Peptide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pseudomonas</em></td>
<td>DWLDSLQR NLQQAALAYPPP MSVVIAGER WMYSADOTPLAGER</td>
</tr>
<tr>
<td><em>Escherichia</em></td>
<td>TAAVLLDTKGEIR NERMTMSQLER IMEAEKLSGADAIK</td>
</tr>
<tr>
<td><em>Streptococcus</em></td>
<td>MKIGNLLGGAVR QDPPSYSYISSR</td>
</tr>
<tr>
<td><em>Paracoccus</em></td>
<td>TAEFADEYPGLADR</td>
</tr>
<tr>
<td><em>Mycobacterium</em></td>
<td>ERWADTPIDNNSPLA沃尔</td>
</tr>
<tr>
<td><em>Sphingomonas</em></td>
<td>NANIIVEGR</td>
</tr>
</tbody>
</table>

### 20 Peptides Of Ambiguous Taxonomy
(3 Proteins)

<table>
<thead>
<tr>
<th>Peptide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTLVTVSSASTK SASTKGPSVFPLAPSSK SSASTKGPSVFPLAPSSK TVSSASTKGPSVFPLAPSSK</td>
</tr>
<tr>
<td>LSEGGLVQPGGSLR LSEGGLVQPGGSLR EVQLESEGGLVQPGGSLR</td>
</tr>
<tr>
<td>STFQQMWISKEEYDEGSIVHRK ASLSTFQQMWISKEEYDEGSIVHRK SLSTFQQMWISKEEYDEGSIVHRK</td>
</tr>
<tr>
<td>STFQQMWISKEEYDEGSIVHRK STFQQMWISKEEYDEGSIVHRK</td>
</tr>
<tr>
<td>TFQMQWISKEEYDEGSIVHRK LSTFQQMWISKEEYDEGSIVHRK</td>
</tr>
<tr>
<td>QMWISKEEYDEGSIVHR FQQMWISKEEYDEGSIVHR</td>
</tr>
</tbody>
</table>

### Additional Peptides for Targeted Analysis

<table>
<thead>
<tr>
<th>Peptide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKADSVPYTI ILDPHYWNSTLR TGGDFSINR VMDSGDGVSTVPIEGYALPHAIRL WGSDFDYKDLSR ATVEELHLEG TTGVMDSDGVSTVPIEGYALPHAIRL LGQDVFPGVTR LVADSITSQLER GITNTSHYEDTPTR QAQAVWGADDK LAEDEIERR AYLTIAK LIPNNOQGK IPAPSGHEEGR RAETELOFCQR ALGMQSGAEAIHRR YLVEPNVKOGK FIVPTDPAK MAGDGQAFALQPSTQQGQK VIPEDLGKLT AALGAYDLR</td>
</tr>
</tbody>
</table>

Katherine Do
**The vision: quantitative assay for host-microbial protein dynamics in clinical CF samples**

<table>
<thead>
<tr>
<th>Source</th>
<th>Peptide sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>KLEDGYFFGAR</td>
</tr>
<tr>
<td></td>
<td>SNSGVRLGYAR</td>
</tr>
<tr>
<td></td>
<td>FAVHVESVLNRPQPEYR</td>
</tr>
<tr>
<td></td>
<td>RFSADEQFFSVGQAASSAHSSK</td>
</tr>
<tr>
<td>Microbial</td>
<td>SSTPSSPTGSSDSGHHHIGWER</td>
</tr>
<tr>
<td></td>
<td>VYRVPDFGWER</td>
</tr>
<tr>
<td></td>
<td>VVDSLAPSITNVQVGK</td>
</tr>
<tr>
<td></td>
<td>ATGHEFAVVK</td>
</tr>
<tr>
<td></td>
<td>LLQVDPEAR</td>
</tr>
</tbody>
</table>

Targeted mass spectrometry-based assays for quantifying levels of host-microbe proteins of interest
METAPROTEOMICS: STRENGTHS & CHALLENGES

FUNCTIONAL COMPOSITION
HOST-MICROBIOME INTERACTION
QUANTIFYING BIOMASS CONTRIBUTIONS

PROTEIN EXTRACTION
VARIABILITY & ABUNDANCE
LARGE DATABASES FALSE POSITIVES
ANALYSIS OF FUNCTIONS EXPRESSED BY MICROBIOMES

Online Workshop
15-24 November 2021

Funded by American Society of Microbiology (ASM) and Indo-United States Science and Technology Forum (IUSSTF).

Hosted by CSIR-IMTech, Chandigarh and Galaxy-P Team, Minneapolis.

https://galaxyproject.org/events/2021-11-microbiomes/home/
ANALYSIS OF FUNCTIONS EXPRESSED BY MICROBIOMES

Online Workshop (15-24 November 2021) Hosted by CSIR-IMTech, Chandigarh and Galaxy-P Team, Minneapolis

WEEK-LONG ONLINE WORKSHOP

- Talks By Eminent Researchers
- Introduction to Training Material
- Participants run workflows based on instructions
- Q & A Session with Instructors
- Quiz Sessions to test prior sessions

DESCRIBE GALAXY IN ONE WORD...
- Reproducible
- accessibility
- helpful!
- Wet and dry research
- Easy to use
- starting-point
- User-friendly
- Awesome
- Marvellous
- Reproducibility
- Saviour
- metagenomics
- easy data analyzer
- Smooth

DESCRIBE METAGENOMICS IN ONE WORD...
- unknown world
- Communities
- Interesting
- comprehensive
- Microbiome
- Delineate
- SISPA and NGS Techni
- diversity
- Virome
- Unravel
- genes collection
- eDNA
- all-in-one
- Pool of information
- Abundance

DESCRIBE METATRANSCRIPTOMICS IN ONE WORD...
- Functional expression
- Master of Dry Lab
- Global RNA
- RNA pool
- RNAseq
- Half way
- mRNA
- Collective analysis
- expression profiling
- Many researcher...
- expression profiling
- Microbes action plan
- Expression

DESCRIBE METAPROTEOMICS IN ONE WORD...
- Function
- Microbes at work
- The Future
- Proteomes
- Many
- Functions
- Expression
- complex data
- Application of metagenom
ANALYSIS OF FUNCTIONS EXPRESSED BY MICROBIOMES

Online Workshop (15-24 November 2021) Hosted by CSIR-IMTech, Chandigarh and Galaxy-P Team, Minneapolis

<table>
<thead>
<tr>
<th>Survey Category</th>
<th>Performance Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of Talks</td>
<td>93.75</td>
</tr>
<tr>
<td>Satisfied with the workshop</td>
<td>93.75</td>
</tr>
<tr>
<td>Will attend future workshops</td>
<td>93.75</td>
</tr>
<tr>
<td>Quality of the Quizzes</td>
<td>98.75</td>
</tr>
<tr>
<td>Quality of Hands-on Workshop</td>
<td>86.25</td>
</tr>
<tr>
<td>Will use Galaxy for Research</td>
<td>83.75</td>
</tr>
</tbody>
</table>
The Galaxy-P team seeks to continue integrating promising new software tools and workflows from leading laboratories into the Galaxy platform and encourages researchers to share their interests so that newer bioinformatics workflows can be made accessible via this resource.
EDUCATION PORTAL @ METAPROTEOMICS INITIATIVE

https://metaproteomics.org/education/

Want to stay up to date about the metaproteomics field and the Initiative? Become a member via www.metaproteomics.org!

@MetaP_Init
info@metaproteomics.org
Metatranscriptomics on Galaxy Training Network (ASaiM-MT)

https://training.galaxyproject.org/training-material//topics/metagenomics/tutorials/metatranscriptomics/tutorial.html

ASaiM

Batut et al Gigascience. 2018 7(6)
doi: 10.1093/gigascience/giy057
Accessing tools and Workflows

METAGENOMICs:
Toolshed: [z.umn.edu/metagenomics_toolshed](z.umn.edu/metagenomics_toolshed)
Galaxy Training Network: [https://training.galaxyproject.org/training-material/topics/metagenomics/](https://training.galaxyproject.org/training-material/topics/metagenomics/)

METATRANSCRIPTOMICs:
Workflow: [http://z.umn.edu/MTWF2020](http://z.umn.edu/MTWF2020)

METAPROTEOMICs:
Workflow: [z.umn.edu/MPWF2020](z.umn.edu/MPWF2020)
Galaxy Training Network: [http://z.umn.edu/gtn-metaproteomics](http://z.umn.edu/gtn-metaproteomics)

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

galaxyp.org/contact