METAPROTEOMICS: Promoting Functional Analysis of Microbiome through ONLINE EDUCATIONAL RESOURCES.

Pratik Jagtap

University of Minnesota

Google Scholar: z.umn.edu/pjgs
PubMed: z.umn.edu/pjagtapreferences
Email: pjagtap@umn.edu
Twitter: pratikomics usegalaxyp
METAPROTEOMICS: Promoting Functional Analysis of Microbiome through online educational resources.

- METAPROTEOMICS AND FUNCTIONAL MICROBIOME ANALYSIS
- METAPROTEOMICS WORKFLOWS WITHIN GALAXY PLATFORM
- QUANTITATIVE METAPROTEOMICS
- TARGETED HOST-MICROBIOME ANALYSIS
- EDUCATION AND PROMOTION
METAPROTEOMICS:
Promoting Functional Analysis of Microbiome through online educational resources.

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

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

90%
of disease can be linked in some way back to the gut and health of the microbiome

5:1
Viruses:Bacteria in the gut microbiota

You have 1.3X more microbes than human cells

The surface area of the GI tract is the same size as 2 tennis courts

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

2kg
The gut microbiota can weigh up to 2kg

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

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.

MICROBIAL TAXA VARY WHILE METABOLIC PATHWAYS REMAIN STABLE WITHIN A HEALTHY POPULATION

METAPROTEOMICS:
Promoting Functional Analysis of Microbiome through online educational resources.

• METAPROTEOMICS AND FUNCTIONAL MICROBIOME ANALYSIS
• METAPROTEOMICS WORKFLOWS WITHIN GALAXY PLATFORM
• QUANTITATIVE METAPROTEOMICS
• TARGETED HOST-MICROBIOME ANALYSIS
• EDUCATION AND PROMOTION
Metaproteomics Analytical Challenges

**Search Database**

<table>
<thead>
<tr>
<th>Size</th>
<th>Complexity</th>
<th>Search Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small to medium size</td>
<td>Single + Contaminants</td>
<td>SINGLE-ORGANISM</td>
</tr>
<tr>
<td>(10 K to 100K sequences)</td>
<td></td>
<td>PROTEOMICS</td>
</tr>
<tr>
<td>LARGE (1 million and above)</td>
<td></td>
<td>METAPROTEOMICS</td>
</tr>
<tr>
<td>Multi-organism database with homologous proteins</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Search algorithms being developed to address large and complex database searches
- Protein grouping at multi-organism level
- Identification statistics affected by large databases
- Taxonomy based on unique peptide identifications
- Functional analysis based on proteins identified

Disparate tools and multiple processing steps.
Metaproteomics Workflow

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

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

**FUNCTIONAL ANALYSIS**
- Known Function
- Proteins

**TAXONOMY ANALYSIS**
- Unique
- Shared
- Unassigned
Software tools can be used in a sequential manner to generate analytical workflows that can be reused, shared and creatively modified.
Metaproteomics Publications

**PROTEOMICS**


**SCIENTIFIC REPORTS**


**Journals**


METAPROTEOMICS: Promoting Functional Analysis of Microbiome through online educational resources.

- METAPROTEOMICS AND FUNCTIONAL MICROBIOME ANALYSIS
- METAPROTEOMICS WORKFLOWS WITHIN GALAXY PLATFORM
- QUANTITATIVE METAPROTEOMICS
- TARGETED HOST-MICROBIOME ANALYSIS
- EDUCATION AND PROMOTION
### 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
- Hypothetical Function
- Unknown Function
- Shared Taxonomy
- Unassigned Taxonomy
- Unique Peptides
- Taxonomy Analysis
While many taxonomic members of microbiomes are known, there are still some undetected taxa.

Functional unknowns: Almost 40% of genes are without a match in functional databases.

Need for quantitation of taxonomy and functions expressed by microbiomes.

Image: Talk by Dr. Robert Hettich (ASMS 2019 Workshop)

https://www.pnnl.gov/projects/soil-microbiome/research
metaQuantome enables quantitative analysis of the taxonomic and functional state of a microbiome.

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.
metaQuantome ANALYSIS: PRINCIPAL COMPONENT ANALYSIS

FUNCTION: PCA PLOT

TAXONOMY: PCA PLOT
**Differentially Abundant Taxa**

<table>
<thead>
<tr>
<th>Taxon Name</th>
<th>Rank</th>
<th>Log2 FC (WS/NS)</th>
<th>FDR (q-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Streptococcus</td>
<td>Genus</td>
<td>2.0</td>
<td>0.0053</td>
</tr>
<tr>
<td>Lactobacillales</td>
<td>Order</td>
<td>1.4</td>
<td>0.0087</td>
</tr>
<tr>
<td>Bacilli</td>
<td>Class</td>
<td>1.4</td>
<td>0.0018</td>
</tr>
<tr>
<td>Gammaproteobacteria</td>
<td>Class</td>
<td>-3.5</td>
<td>0.0048</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>Phylum</td>
<td>-4.0</td>
<td>0.0007</td>
</tr>
<tr>
<td>Fusobacterium</td>
<td>Genus</td>
<td>-5.0</td>
<td>0.0026</td>
</tr>
<tr>
<td>Fusobacteriales</td>
<td>Order</td>
<td>-5.1</td>
<td>0.0031</td>
</tr>
</tbody>
</table>

**Heatmap Cluster Analysis**
metaQuantome deciphers the contribution of taxa to a functional process.

**CARBOHYDRATE METABOLISM**

**Control**
- **SUGAR**
  - Sugar → Lactate
  - Acidity increases
  - *Fusobacterium* does not thrive in high acidic conditions

**In presence of sugar**
- **CONTROL**
  - Carbohydrates → Weak acids
  - Low acidity
  - *Fusobacterium* thrives in mild acidic conditions
QUANTITATION METHODS

DIA-MS OFFERS BENEFITS FOR QUANTITATIVE METAPROTEOMICS

DIA-MS HAS BETTER PROTEOME COVERAGE

- DDA-MS: 57.7%
- DIA-MS: 96.7%

DIA-MS OFFERS QUANTITATIVE REPRODUCIBILITY

- Bacteriophage Mix1
  - Log2 peptide intensity (DDA)
- Bacteriophage Mix2
  - Log2 peptide intensity (DDA)
- Bacteriophage Mix3
  - Log2 peptide intensity (DDA)
- Bacteriophage Mix4
  - Log2 peptide intensity (DDA)

DIA-MS OFFERS HIGHER ACCURACY

- T4 Bacteriophage: 8.2%
- Bacillus subtilis: 9.4%
- Escherichia coli: 8.6%
- Salmonella typhimurium: 5%

DIA-MS HAS LESHER MISSING VALUES

- DDA-Spc: 56.3%
- DDA-MS1: 54.5%
- DIA: 58.8%

Brook L. Nunn
Susan T. Weintraub
METAPROTEOMICS: Promoting Functional Analysis of Microbiome through online educational resources.

- METAPROTEOMICS AND FUNCTIONAL MICROBIOME ANALYSIS
- METAPROTEOMICS WORKFLOWS WITHIN GALAXY PLATFORM
- QUANTITATIVE METAPROTEOMICS
- TARGETED HOST-MICROBIOME ANALYSIS
- EDUCATION AND PROMOTION
Workflow Overview for Peptides and Proteins Detection and Quantification.

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

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

Peptides and Proteins Detection & Quantitation

<table>
<thead>
<tr>
<th>PROTEINS</th>
<th>PEPTIDES</th>
</tr>
</thead>
<tbody>
<tr>
<td>5873</td>
<td>69,906</td>
</tr>
<tr>
<td>(129)</td>
<td>(335)</td>
</tr>
<tr>
<td>3803</td>
<td>36,409</td>
</tr>
<tr>
<td>(161)</td>
<td>(300)</td>
</tr>
<tr>
<td>54,287</td>
<td></td>
</tr>
<tr>
<td>(1879)</td>
<td></td>
</tr>
</tbody>
</table>

2292 microbial peptides
Overview of Microbial Peptide Verification

Peptides and Proteins Detection & Quantitation

Microbial Peptides
2292 microbial peptides

Microbial Peptide Verification

PepQuery

PepQuery Verified Peptides

VERIFIED

680

NON-VERIFIED

1612

searchgui

peptide-shaker
Overview of Microbial Peptide Verification

Verified Microbial Peptides

680 microbial peptides

FragPipe

MaxQuant

SGPS

searchgui

peptide-shaker

33

5

17

2
Verified Microbial Peptides

Peptides and Proteins Detection & Quantitation

Microbial Peptide Verification

Taxonomy & Functional Analysis
Verified Microbial Peptides
Verified Microbial Peptides

87 Verified Microbial Peptides

24 Taxonomy-Specific Candidate Peptides
(Known taxonomy)

- Pseudomonas
- Moraxella
- Escherichia
- Staphylococcus
- Streptococcus
- Prevotella
- Paracoccus
- Stenotrophomonas
- Mycobacterium
- Fusobacterium
- Sphingomonas
- Propionibacterium

20 Peptides Of Ambiguous Taxonomy
(3 Proteins)

- GTLVTVSSASTK
- SASTKGPSVQFPLAPPSSK
- SSASTKGSVPSFPLAPPSSK
- TVSSASTKGSVPSFPLAPPSSK
- LESGGGLVQPQGSRL
- LLESGGGLVQPQGSRL
- EVQLLESGGGLVQPQGSRL
- FGGGTKTLVQLQPK
- LTVLQQPK
- VFGGGTKTLVQLQPK
- STFQMQMWSIKEEYESGPSIVHRK
- ASLSTFQMQMWSIKEEYESGPSIVHRK
- SLSTFQMQMWSIKEEYESGPSIVHRK
- STFQMQMWSIKEEYESGPSIVHRK
- TFQMQMWSIKEEYESGPSIVHRK
- LSTFQMQMWSIKEEYESGPSIVHRK
- QQMWSIKEEYESGPSIVHRK
- FQMQMWSIKEEYESGPSIVHRK

Additional Peptides for Targeted Analysis

- AKADSYPYTI
- ILDPNGYWNSTLR
- TGGGDFSNR
- VMDSGDGVSHTVPVEYGAFLAPHL
- WGSFDYKGDLRS
- ATVEELHLGI
- TTVGvMDSDGVSHTVPEYGAFLAPHL
- LGQDVVFPGVTR
- LVADSITQLER
- GITINTSHYEDTPTR
- QAVAGWGGADK
- LAEDELERR
- AYLLTIAK
- LIPNQGQK
- IPAPSGHEEGR
- RAETELOFCQR
- ALGMOGSAEAIHR
- YLVEPNVKQDGK
- FVNTDPAK
- MAGDGQFAFLQPOSTQQGQQK
- VIPDELGKLT
- AALGAYDLR

Katherine Do
Quantitation of Microbial Peptides

**Pseudomonas**
GO:0016021 integral component of membrane
IPR:IPR042094 Type II secretion system GspF domain superfamily

**Moraxella catarrhalis**
GO:0009279 cell outer membrane
GO:0005509 calcium ion binding
GO:0006811 ion transport
IPR006664 Outer membrane protein

**Moraxella catarrhalis**
IPR011250 Outer membrane protein / outer membrane enzyme PagP beta-barrel

**Moraxella**
GO:0009279 cell outer membrane
GO:0005509 calcium ion binding
GO:0006811 ion transport
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><strong>Human</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KLEDGYFGGAR</td>
</tr>
<tr>
<td></td>
<td>SNSGVRLDGYAR</td>
</tr>
<tr>
<td></td>
<td>FAVHVESVLNRVPQPEYR</td>
</tr>
<tr>
<td></td>
<td>RFSADEQFFSVGQAASSSAHSSK</td>
</tr>
<tr>
<td><strong>Microbial</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SSTPSSPTGTSDDSGGHHIGWER</td>
</tr>
<tr>
<td></td>
<td>VYRVPDFGVWER</td>
</tr>
<tr>
<td></td>
<td>VVDSLAPSITNVLVQGK</td>
</tr>
<tr>
<td></td>
<td>ATGHEFAVK</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: Promoting Functional Analysis of Microbiome through online educational resources.

• METAPROTEOMICS AND FUNCTIONAL MICROBIOME ANALYSIS
• METAPROTEOMICS WORKFLOWS WITHIN GALAXY PLATFORM
• QUANTITATIVE METAPROTEOMICS
• TARGETED HOST-MICROBIOME ANALYSIS
• EDUCATION AND PROMOTION
METAPROTEOMICS: STRENGTHS & CHALLENGES

FUNCTIONAL COMPOSITION

HOST-MICROBIOME INTERACTION

QUANTIFYING BIOMASS CONTRIBUTIONS

PROTEIN EXTRACTION

VARIABILITY & ABUNDANCE

LARGE DATABASES FALSE POSITIVES
TO BOLDLY GO WHERE METAPROTEOMICS HAS NOT REACHED BEFORE...

• Introduce Metaproteomics to New Researchers
• Make Tools and Workflows Accessible to Researchers
• Attract new perspectives and talent
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
- collective analysis
- RNAseq
- half way
- mRNA
- community expression
- expression profiling
- activity
- yet-to-learn
- RNA
- Moist researcher...
- Microbes action plan
- Expression

**DESCRIBE METAPROTEOMICS IN ONE WORD...**

- Functions
- Functions
- Function

- Microbes at work
- The Future
- Proteomes
- mass-spec
- 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

- Quality of Talks: 93.75%
- Satisfied with the workshop: 93.75%
- Will attend future workshops: 93.75%
- Quality of the Quizzes: 98.75%
- Quality of Hands-on Workshop: 86.25%
- Will use Galaxy for Research: 83.75%
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
Long term Goal

Disseminate Validated Tools and workflows

- Galaxy Project
- Galaxy Europe
- Galaxy Australia
- GitHub

Presentations and Publications

Training

Adoption

- Workflows accessible via usegalaxy*

- Galaxy Project
- Galaxy Europe
- Galaxy Australia

Hands-on Workshops @ Annual Conferences

On-Demand, Online Training

Biological Discovery
Tools and Workflows also available on:
https://proteomics.usegalaxy.eu/

Galaxy Training Network:
https://training.galaxyproject.org/training-material/topics/proteomics

Galaxy Europe: https://proteomics.usegalaxy.eu/

Contact: http://galaxyp.org/contact/
What’s On The Horizon?

**Quantitation**

- Predicted Peptide Library
  - Predicted fragmentation
  - Predicted retention time
- Empirically corrected Chromatogram Library
- Experimental wide-window DIA data

**Targeted Metaproteomics**

- MS DATA
  - DDA
  - DIA
  - Gas phase fractionation
    - Empirical narrow-window DIA data
- Peptide Quantitation
  - EncyclopeDIA
  - metaQuantome

**Characterization of Proteins with Unknown Function**

- metaQuantome
  - SIGNIFICANT TAXA, PROTEINS, FUNCTIONS
  - TABULAR OUTPUT WITH PEPTIDE m/z, RT INFORMATION
  - DATA EXPLORATION
    - Differential Abundance
    - Cluster Analysis
  - + multiple conditions + interactive visualization + targeted analysis integration

- HOMOLOGY WITH PROTEINS OF KNOWN STRUCTURE / FUNCTION
  - CONSERVED DOMAINS

- METAGENOME METADATA
  - METAGENOME ANNOTATION TABULAR OUTPUTS