METAPROTEOMICS FOR THE FUTURE: DEMOCRATIZING FUNCTIONAL ANALYSIS OF MICROBIOMES VIA COMMUNITY-BASED INFORMATICS DEVELOPMENT AND DISSEMINATION

PRATIK JAGTAP, BJOERN GRUENING, JIM JOHNSON, ALESSANDRO TANCA, THOMAS DOAK, W. JUDSON HERVEY, THILO MUTH, BERNHARD Y. RENARD, CAROLIN KOLMEDER, JOSH ELIAS, BART MESUERE, DAVE CLEMENTS, THOMAS MCGOWAN, CLEMENS BLANK, PRAVEEN KUMAR, JOEL RUDNEY AND TIM GRIFFIN
**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**.
**Metaproteomic Challenges**

**Single-Organism Proteomics**

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

**Metaproteomics**

- **Search Database**
  - Size: Large (1 million and above)
  - Complexity: Multi-organism database with homologous proteins

- Disparate tools and multiple processing steps.

- **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
**Metaproteomics in Galaxy**

- **GALAXY: WEB-BASED BIOINFORMATICS DATA ANALYSIS PLATFORM.**
  - Facilitates software usability for bench scientists
  - Ensures Reproducibility
  - Ability to assimilate disparate software into integrated workflows.

- **DEVELOPED METAPROTEOMICS WORKFLOWS WITHIN GALAXY**

- **UTILIZED WORKFLOWS IN RESEARCH STUDIES**
  - Rudney et al., BMC Microbiome DOI: 10.1186/s40168-015-0136-z
  - Bhargava et al., Bronchoalveolar Lavage Fluid Metaproteome in Acute Respiratory Failure (In Preparation)
  - Skubitz et al., Metaproteomic analysis of human cervical-vaginal fluid in residual Pap tests – Insights into the cervical microbiome (In Preparation)
Inputs for MEGAN

DATABASE
GENERATION

DATABASE SEARCH
& STRATEGIES

FUNCTIONAL
ANALYSIS

TAXONOMY
ANALYSIS

• Public database repositories

Inputs for MEGAN

ProteinPilot™

Inputs for MEGAN

UniPept

Workflows for database generation, Database search, Taxonomy analysis and generating outputs for functional analysis.

Extend repertoire of tools and use them in workflows.
1 Email communication with metaproteomics researchers and developers around the world

2 Set up a userlist and discussed needs for the community.

3 Prioritized tools and workflows and decided on a date for the “Contribution Fest”.
**Metaproteomics in Galaxy**

- **Database Generation**
  - Public database repositories
  - Databases derived from taxonomic information
  - Metagenomic data processing tools
  - Databases derived from WGS data

- **Database Search & Strategies**
  - ProteinPilot™
  - SearchGUI / PeptideShaker
  - SIPROS3
  - Database Searching Strategies.

- **Functional Analysis**
  - Inputs for MEGAN
  - UniPept Extensions
  - DIAMOND / EggNOG

- **Taxonomy Analysis**
  - Inputs for MEGAN
  - UniPept
  - Prophane

Workflows for database generation, Database search, Taxonomy analysis and generating outputs for functional analysis.

Extend repertoire of tools and use them in workflows.
“Contribution Fest” (led by Bjoern Gruening) discussed about packaging and testing of tools for Galaxy.

1. Email communication with metaproteomics researchers and developers around the world.

2. Set up a user mail and discussed needs for the community.

3. Prioritized tools and workflows and decided on a date for “Contribution Fest”.

4. Github issues generated; tools were developed & packaged, CONDA tested and added to Galaxy Toolshed.

5. Tools available for worldwide Galaxy users via Galaxy Toolshed.

http://z.umn.edu/mphack2016
Extended repertoire of tools and used them in workflows.
The Galaxy platform for metaproteomic characterization of microbiomes

SATELLITE Workshop @ ABRF Annual Meeting
March 25, 2017

https://galaxyproteomics.github.io/abrf2017/
An Accessible Metaproteomics Resource for Researchers

http://z.umn.edu/metaproteomicsgateway

Gateway on JetStream with documentation, tools & workflows used at the ABRF workshop
Challenges & Opportunities

Enhancing metaproteomics—The value of models and defined environmental microbial systems
Florian-Alexander Herbst, Vanessa Lohmann, Andreas Thiele, Martin von Bergen, and Par Halkjaer Nielsen

Next-generation proteomics faces new challenges in environmental biotechnology
Jean Argnaraud

Navigating through metaproteomics data: A logbook of database searching
Thilo Mehn, Carsten A. Gehmmer, Jörg Retzlaff, Sven J. Mehl, Marcus Venzdo, René W. Joffre, Volker S. Forschner, Uli Hetzel, Christian Steglich, and Gerhard Markert

A decade of metaproteomics: Where we stand and what the future holds
Paul Wirmser, Ameena Hamez-Buchheit, and Philip L. Borda
Challenges & Opportunities

- Learnings from metagenomic approaches: Statistical tools and visualization

- Quantification: DIA and emerging methods

- Multi-omics Meta-transcriptomics and Metabolomics

- Quality Control Instrumentation and Spectral Quality Estimation

- Targeted metaproteomics
‘The ultimate goal is to develop workflows that enable the functional characterization of not only the microbiome of the human surfer, but also the rich microbiome of the underlying ocean…while sailing through the ocean of data’
## Galaxy-P @ ASMS 2017

### https://galaxyproject.org/events/2017-asms/ Presentations by the **Galaxy-P team:**

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Topic</th>
<th>Presenter</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tue Jun 6</td>
<td>10:30am-2:30pm</td>
<td><strong>TP 362</strong>: Multi-Omic Informatics in the Cloud: Galaxy-P takes a ride on JetStream</td>
<td>Tim Griffin</td>
<td>Exhibit Hall</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>TP 364</strong>: A Galaxy-based, Multi-Staged, two-step searching Pipeline for improved peptide identification in Next-generation Proteomic Studies</td>
<td>Praveen Kumar</td>
<td>Exhibit Hall</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>TP 365</strong>: Enhancing the Multi-Omics Visualization Platform (MVP) plug-in for Galaxy-based Applications</td>
<td>Thomas McGowan</td>
<td>Exhibit Hall</td>
</tr>
<tr>
<td>Wed Jun 7</td>
<td>10:30am-2:30pm</td>
<td><strong>WP 333</strong>: Galaxy-centered lab workflows to organize metabolomics analyses</td>
<td>Arthur C Eschenlauer</td>
<td>Exhibit Hall</td>
</tr>
</tbody>
</table>

**WORKSHOP 4**

**GALAXY-P IN THE CLOUD:**

**PROTEOMIC INFORMATICS ON JETSTREAM**

**WEDNESDAY, June 7**

**5:45-7:00 PM, Room 135-136**

Follow us on: [twitter.com/usegalaxyp](https://twitter.com/usegalaxyp)
ACKNOWLEDGEMENTS

galaxyp.org

**Biologists / collaborators**
- Joel Rudney
- Maneesh Bhargava
- Kevin Viken
- Amy Skubitz
- Kristin Boylan
- Marnie Peterson
- Somiah Afiuni
- Chris Wendt
- Brian Sandri
- Alexa Pragman

**UNIVERSITY OF MINNESOTA**
- TIM GRIFFIN
  - Praveen Kumar
  - Candace Guerrero
  - Subina Mehta
  - Shane Hubler

**University of Bergen, Norway**
- Harald Barsnes
- Marc Vaudel

**University of Freiburg, Germany**
- Judson Hervey
- Naval Research Institute
- Washington, D.C.

**VIB, Ugent, Belgium**
- Alessandro Tanca
- Porto Conte Ricerche, Italy

**University of Helsinki, Finland**
- Carolin Kolmeder

**University of Berg, Australia**
- Ira Cooke

**GalaxyP**
- Dave Clements
  - Galaxy Project
- Josh Elias
  - Stanford University
- Thomas Doak
  - Indiana University
- Matt Chambers
  - Vanderbilt University, Nashville, TN
- Minnesota Supercomputing Institute
  - James Johnson
  - Tom McGowan
  - Michael Milligan

**Funding**

**National Institutes of Health**

**GalaxyP**

**University of Bergen, Norway**
- Harald Barsnes
- Marc Vaudel

**University of Freiburg, Germany**
- Judson Hervey
- Naval Research Institute
- Washington, D.C.

**VIB, Ugent, Belgium**
- Alessandro Tanca
- Porto Conte Ricerche, Italy

**University of Helsinki, Finland**
- Carolin Kolmeder

**University of Berg, Australia**
- Ira Cooke