GALAXY-P FOR METAPROTEOMICS:
FUNCTIONAL MICROBIOME ANALYSIS
VIA COMMUNITY-BASED INFORMATICS DEVELOPMENT &
DISSEMINATION

2nd International Metaproteomics Symposium (Alghero, Italy)
COLLABORATIVE EFFORT

galaxyp.org
# Metaproteomic Challenges

**Single-Organism Proteomics**

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

An Accessible Metaproteomics Resource for Researchers

http://z.umn.edu/metaproteomicsgateway

Instructions for accessing Jetstream metaproteomics gateway

http://z.umn.edu/accessmpgwjetstream

Instructions for accessing workflows and training material

Links to instructional materials from the ABRF 2017 satellite workshop describing how to access and use the metaproteomic workflows.

- Slides from the ABRF 2017 Workshop: http://z.umn.edu/mpgwsldes
- Documentation from the ABRF 2017 Workshop: http://z.umn.edu/mpgwdocs

Tools and workflows on this site are a result of collaboration between Galaxy-P team usegalaxyp.org, Jetstream instance jetstream-cloud.org and the members of the Metaproteomics Contribution Fest http://z.umn.edu/mphack2016.
**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)
Rudney et al., BMC Microbiome DOI: 10.1186/s40168-015-0136-z

- 2-D LC-MS/MS Metaproteome was analyzed for taxonomy and function.
- Taxonomically-diverse sucrose-pulsed microcosms clustered more closely by function. Function-based changes may be better indicators of dysbiosis.
Inputs for MEGAN

DATABASE

DATABASE SEARCH & STRATEGIES

FUNCTIONAL ANALYSIS

Inputs for MEGAN

TAXONOMY ANALYSIS

Inputs for MEGAN

UniPept

ProteinPilot™

• Public database repositories

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

Set up a userlist and discussed needs for the community.

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

Inputs for MEGAN
Database generation, database search, taxonomy analysis and generating outputs for functional analysis.

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

Extend repertoire of tools and use them in workflows.

---

**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
  - MetaProteome Analyzer
Email communication with metaproteomics researchers and developers around the world.

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

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

“Contribution Fest” (led by Bjoern Gruening) discussed about packaging and testing of tools for Galaxy.

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

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

Nunn Lab, University of WA
Seattle, WA

https://galaxyproteomics.github.io/abrf2017/
An Accessible Metaproteomics Resource for Researchers
http://z.umn.edu/metaproteomicsgateway

Instructions for accessing Jetstream metaproteomics gateway
http://z.umn.edu/accessmpgwjetstream

Instructions for accessing workflows and training material
Links to instructional materials from the ABRF 2017 satellite workshop describing how to access and use the metaproteomic workflows.
- Slides from the ABRF 2017 Workshop: http://z.umn.edu/mpgwslides
- Documentation from the ABRF 2017 Workshop: http://z.umn.edu/mpgwdocs

Tools and workflows on this site are a result of collaboration between Galaxy-P team Jetstream instance jetstream-cloud.org and the members of the Metaproteomics Co.

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

z.umn.edu/metaproteomicsgateway
**Project goals & scope**

**Experiment Design**

- Clinical study?
- Cancer-related?
- Metaproteomics
- Shotgun
- Metagenomics
- Metatranscriptomics?
- Metabolomics?
- Quantitative (label-free; DIA; labeled?)

---

**Sample preparation**

- Methods optimized in the laboratory.

---

**Data acquisition**

- Methods optimized in the laboratory.

---

**Optimize and run Workflows**

**Data Analysis**

- Run analysis Using Galaxy.

---

**Tools and workflows**

- Metaproteomics contribution-fest

---

**Manuscript writing and submission**

- Community effort

---

**Metaproteomics Projects**
Metaproteomics Projects

Ocean Metaproteomics
Woods Hole, MA
(Mak Saito)
ACKNOWLEDGEMENTS

galaxyp.org