ADVANCING METAPROTEOMICS RESEARCH VIA COMMUNITY-BASED INFORMATICS DEVELOPMENT & DISSEMINATION

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

Galaxy Community Conference (Montpellier, France) 30th June 2017
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**

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

**Metaproteomics**

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

**Search Database**

- **Size**
  - SINGLE
- **Complexity**
  - SINGLE + Contaminants

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)
Metaproteomic analysis using the Galaxy framework.
Metaproteomics in Galaxy

- **Database Generation**
  - Public database repositories

- **Database Search & Strategies**
  - ProteinPilot™

- **Functional Analysis**
  - Inputs for MEGAN

- **Taxonomy Analysis**
  - 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
- MetaProteome Analyzer

---

**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.

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
**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
  - Database Searching Strategies
  - SIPROS3

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

**Taxonomy Analysis**
- Inputs for MEGAN
  - UniPept
  - MetaProteome Analyzer

Extended repertoire of tools and used them in workflows.
SATELLITE Workshop @ ABRF Annual Meeting
March 25 2017
San Diego, CA

Workshop @ ASMS Annual Meeting
June 7 2017
Indianapolis, IN

Workshop @ GCC Annual Meeting
June 28 2017
Montpellier, France

• [https://galaxyproteomics.github.io/abrf2017/](https://galaxyproteomics.github.io/abrf2017/)
• [https://galaxyproteomics.github.io/asms2017/](https://galaxyproteomics.github.io/asms2017/)
• [https://galaxyproteomics.github.io/gcc2017/](https://galaxyproteomics.github.io/gcc2017/)
Metaproteomics Gateway

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. http://z.umn.edu/mphack2016.

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

z.umn.edu/metaproteomicsgateway
Embracing Complexity and Diversity:
Metaproteomics within the Galaxy Framework

GCC 2016 - Bloomington, IN
29th June 2016

Next steps...

- Seek inputs from leading metaproteomics researchers from around the world
- Identify tools and workflows
- Prioritize the tools and workflows
- Approach the Inter-Galactic Utilities Commission for a hackathon
- Test and make robust workflows available through Docker and other avenues.

Questions / Suggestions?
What Next? - Metaproteomics Projects

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

Project goals & scope
- Experiment Design

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

OCEAN METAPROTEOMICS Meeting in Woods Hole, MA
(organized by Mak Saito)
Metaproteomics Projects

International Metaproteomics Symposium 2017
Alghero, Italy
Conclusions

- Developed **metaproteomics workflows within Galaxy platform** and utilized workflows in research studies.
- Conducted **community-based development via Contribution-Fest** to extend the repertoire of tools and used them in workflows.
- Conducted **workshops** at national and international annual conferences.
- Set up a **test public instance metaproteomics gateway**. Plans to upgrade this resource.
- Seeking out clinical and environmental projects to develop and implement tools and workflows for advanced meta-genomics/transcriptomics/proteomics analysis within Galaxy.
ACKNOWLEDGEMENTS

The Galaxy-P Team at University of Minnesota

Funding

galaxyp.org

Follow us on: twitter.com/usegalaxyp
ACKNOWLEDGEMENTS

Biologists / collaborators

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

Minnesota Supercomputing Institute

James Johnson
Tom McGowan
Michael Milligan

University of Minnesota

TIM GRIFFIN
Praveen Kumar
Candace Guerrero
Subina Mehta
Shane Hubler

Dave Clements
Galaxy Project

Josh Elias
Stanford University

Thomas Doak
Indiana University

Matt Chambers
Vanderbilt University, Nashville, TN

Harald Barsnes &
Marc Vaudel
University of Bergen, Norway

Bart Mesuere
Felix Van der Jeugt
VIB, Ugent, Belgium

Judson Hervey
Naval Research Institute
Washington, D.C.

Alessandro Tanca
Porto Conte Ricerche, Italy

Bjoern Gruening
Clemens Blank
Bérénice Batut
University of Freiburg, Freiburg, Germany

Carolin Kolmeder
University of Helsinki, Finland

Thilo Muth
Bernhard Renard
Robert Koch Institut

University of Bergen,
Norway

University of Freiburg,
Freiburg, Germany

VIB, Ugent, Belgium

University of Helsinki,
Finland

Galaxy Project

GalaxyP