FUNCTIONAL MICROBIOME ANALYSIS USING METAPROTEOMICS WITHIN GALAXY FRAMEWORK

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

20th October 2017
FUNCTIONAL MICROBIOME ANALYSIS USING METAPROTEOMICS WITHIN GALAXY FRAMEWORK

• *Functional microbiome analyses*

• *Metaproteomics, Galaxy and Projects*

• *Extending tools for metaproteomics*

• *Future Plans*
The Importance of the MICROBIOME by the Numbers

- **90%**: Up to 90% of all disease can be traced back to the gut and health of the microbiome.
- **10–100 trillion**: Number of symbiotic microbial cells harbored by each person, primarily bacteria in the gut, that make up the human microbiome.
- **>10,000**: Number of different microbial species researchers have identified living in the human body.
- **100 to 1**: The microbes in our microbiome outnumber the cells in our genome by about 100 to 1.
- **3.3 million**: Number of non-redundant genes in the human gut microbiome.
- **22,000**: Additional number of genes in the human gene catalog.
- **99.9%**: Percentage individual humans are identical to one another in terms of host genome.
- **80%-90%**: Percentage individual humans are different from one another in terms of the microbiome.

MICROBIOME RESEARCH

**GUT MICROBIOME**

- **Chronic Disease, Autoimmune, Obesity, Diabetes**
- **Mood Disorders, Depression, Anxiety, Autism**
- **Age**
- **Mode of birth**
- **Maternal Gut Flora**
- **Maternal and Paternal Nutrient sufficiency**
- **Breastfed?**
- **Colicky?**
- **Environmental Exposures**

---

http://draxe.com/microbiome/
**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.**
FUNCTIONAL MICROBIOME ANALYSIS USING METAPROTEOMICS WITHIN GALAXY FRAMEWORK

- Functional microbiome analyses
- Metaproteomics, Galaxy and Projects
- Extending tools for metaproteomics
- Future Plans
“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.”

Bond and Wilmes 2015

Ploeg et al. 2018

“Metaproteomics can provide functional insights into microbial consortia composition and activity.”

**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 (10 K to 100K sequences)
  - Complexity: SINGLE + Contaminants

**Metaproteomics**

- **Search Database**
  - Size: Large (1 million and above)
  - Complexity: 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.
• A web-based bioinformatics data analysis platform.
• Software accessibility and usability.
• Share-ability of tools, workflows and histories.
• Reproducibility and ability to test and compare results after using multiple parameters.
• Ability to assimilate disparate software into integrated workflows.
Software tools can be used in a sequential manner to generate analytical workflows that can be reused, shared and creatively modified.

For example, Protein Database Downloader downloads UniProt protein FASTA databases of various organisms.
Metaproteomic analysis using the Galaxy framework.
PEPTIDES TO TAXONOMY AND FUNCTION

- **Protein inference conundrum in metaproteomics**
- **TAXONOMY:** Lowest common ancestor used for taxonomic classification
- **FUNCTION:** Peptides $\rightarrow$ Proteins $\rightarrow$ Function
Plaque microcosms were grown from carious lesions in 12 children.

Extensive inter-individual variation in dental plaque composition.

Paired design – microcosms grown in presence/absence of sucrose.

Sucrose induces acid production in dental caries (dysbiosis).

16S rRNA sucrose microcosms retained inter-individual taxonomic variation.

---

**NS:** Absence of sucrose  **WS:** Presence of sucrose  
**P:** Plaque samples

**Rudney et al., BMC Microbiome** DOI: 10.1186/s40168-015-0136-z
**METAPROTEOMICS PROJECTS**

**SUCROSE-INDUCED DYSBIOSIS IN DENTAL CARIES.**

Rudney et al., *BMC Microbiome* DOI: 10.1186/s40168-015-0136-z

1. **RAW data conversion**
2. **Database generation**
3. **Database Search against Human Oral Microbiome Database (HOMD)**
4. **Convert microbial peptide lists to FASTA format**
5. **BLAST-P Analysis (30K to 60K peptides)**
6. **MEGAN Analysis**
METAPROTEOMICS PROJECTS
SUCROSE-INDUCED DYSBIOSIS IN DENTAL CARIES.

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
Metaproteomics Projects

Sucrose-induced dysbiosis in dental caries.

- Statistical analysis identified proteins that were expressed at a higher abundance in presence of sugar.

- Selected reaction monitoring (SRM) was used to quantify conserved peptides from three enzymes whose abundance increased in sucrose-induced dysbiosis: NADP-specific glutamate dehydrogenase (EC 1.4.1.4); Ketol-acid reductoisomerase (EC 1.1.1.86); L-lactate dehydrogenase (EC 1.1.1.27).
**METAPROTEOMICS PROJECTS**

**ONGOING PROJECTS**

**Bhargava Lab**
- **MICROBIOME OF HUMAN CERVICAL-VAGINAL FLUID** was analyzed after collecting samples using residual Pap test fixatives from a pool of 40 normal samples and 5 individual women with normal cytology.
- Understanding the nature of the microorganisms and the functional role of the expressed proteins is an important step to develop diagnostic tools for gynecological conditions and malignancies.

**MEGAN Analysis**
- **776 Microbial Peptides**
  - 53 genera. 71 species
  - Function: 232 peptides assigned to 107 InterPro2GO proteins

**Skubitz Lab**
- **MICROBIOME OF HUMAN CERVICAL-VAGINAL FLUID** was analyzed after collecting samples using residual Pap test fixatives from a pool of 40 normal samples and 5 individual women with normal cytology.
- Understanding the nature of the microorganisms and the functional role of the expressed proteins is an important step to develop diagnostic tools for gynecological conditions and malignancies.

**MEGAN Analysis**
- **1190 Microbial Peptides**
  - 30 genera. 42 species
  - Function: 939 peptides assigned to 137 InterPro2GO proteins

**BRONCHOALVEOLAR LAVAGE FLUID (BALF) MICROBIOME**: Microbiome of fluid proximate to the site of lung injury – acquired by using a bronchoscopy procedure – was analyzed.

- BALF from subjects with Acute Respiratory Disease (ARDS) patients and lung injury after Hematopoietic Stem Cell Transplant (HSCT) were analyzed.
FUNCTIONAL MICROBIOME ANALYSIS USING METAPROTEOMICS WITHIN GALAXY FRAMEWORK

• *Functional microbiome analyses*

• *Metaproteomics, Galaxy and Projects*

• *Extending tools for metaproteomics*

• *Future Plans*
Metaproteomics in Galaxy

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

Extend repertoire of tools and use them in workflows.

- **Public database repositories**
- **Inputs for MEGAN**
- **ProteinPilot™**
- **UniPept**
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.
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.
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 Core http://z.umn.edu/mphack2016.

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

z.umn.edu/metaproteomicsgateway
FUNCTIONAL MICROBIOME ANALYSIS USING METAPROTEOMICS WITHIN GALAXY FRAMEWORK

• *Functional microbiome analyses*
• *Metaproteomics, Galaxy and Projects*
• *Extending tools for metaproteomics*
• *Future Plans*
What Next?
Challenges & Opportunities

Navigating through metaproteomics data: A log of plant data

The impact of sequence database choice on metaproteomic results in gut microbiota

Critical decisions in metaproteomics: achieving high confidence protein annotations in a sea of unknowns

Enhancing metaproteomics – The value of models and defined environmental microbial systems

A decade of metaproteomics: Where we stand and what the future holds
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
Statistical tools and visualization

**Metaproteomics Projects**

**Database Generation**
- FASTQ

**Database Search & Strategies**
- Peptides
- FASTA
- Search Algorithm
- Spectra

**Functional Analysis**
- Peptides
- Known Function
- Proteins

**Taxonomy Analysis**
- Unassigned
- Shared
- Unique

**Projects**

- Metaproteomics
  - Databases
  - Algorithms
Metaproteomics Projects

**DATABASE GENERATION**
- FASTQ

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

**FUNCTIONAL ANALYSIS**
- Peptides
- Known Function
- Proteins
- Unassigned

**QUANTITATIVE ANALYSIS**
- Hypothetical Protein
- Unassigned

**TAXONOMY ANALYSIS**
- Unique
- Shared
- Unassigned
Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes.

Heintz-Buschart et al NATURE MICROBIOLOGY 2, 16180 (2016) | DOI: 10.1038/nmicrobiol.2016.180
Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes.
Heintz-Buschart et al NATURE MICROBIOLOGY 2, 16180 (2016) | DOI: 10.1038/nmicrobiol.2016.180
What Next? - Metaproteomics Projects

**Project goals & scope**

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

**Experiment Design**

- Methods optimized in the laboratory.

**Sample preparation**

**Data acquisition**

- Methods optimized in the laboratory.

**Tools and workflows**

- Metaproteomics contribution-fest

**Optimize and run Workflows Data Analysis**

- Run analysis Using Galaxy.

**Manuscript writing and submission**

- Community effort

---

*Note: The image includes a world map background with associated text boxes and icons.*
‘The ultimate goal is to develop workflows that enable the functional characterization of diverse microbiomes...
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

Funding

galaxyp.org