

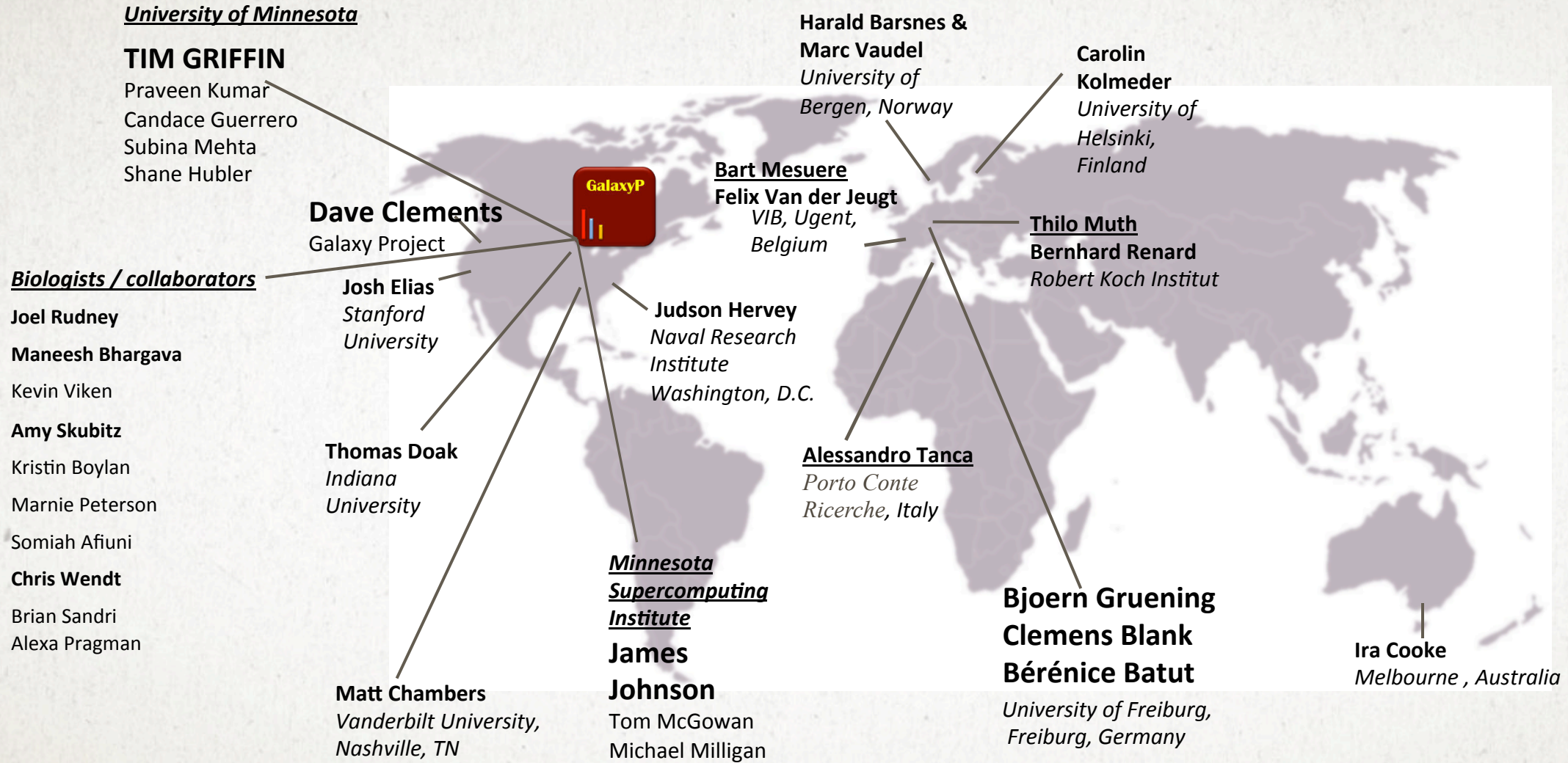
ADVANCING METAPROTEOMICS RESEARCH VIA COMMUNITY-BASED INFORMATICS DEVELOPMENT & DISSEMINATION



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***

COLLABORATIVE EFFORT



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**.

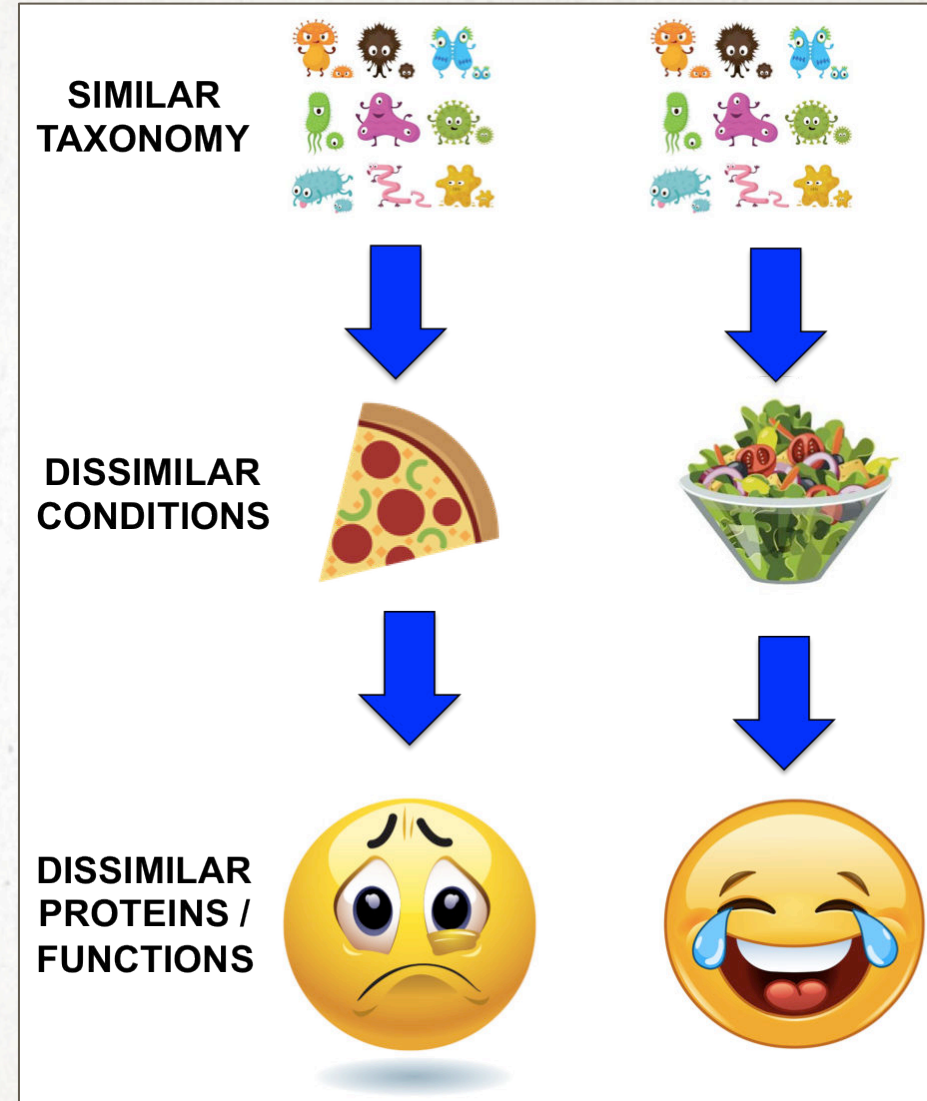
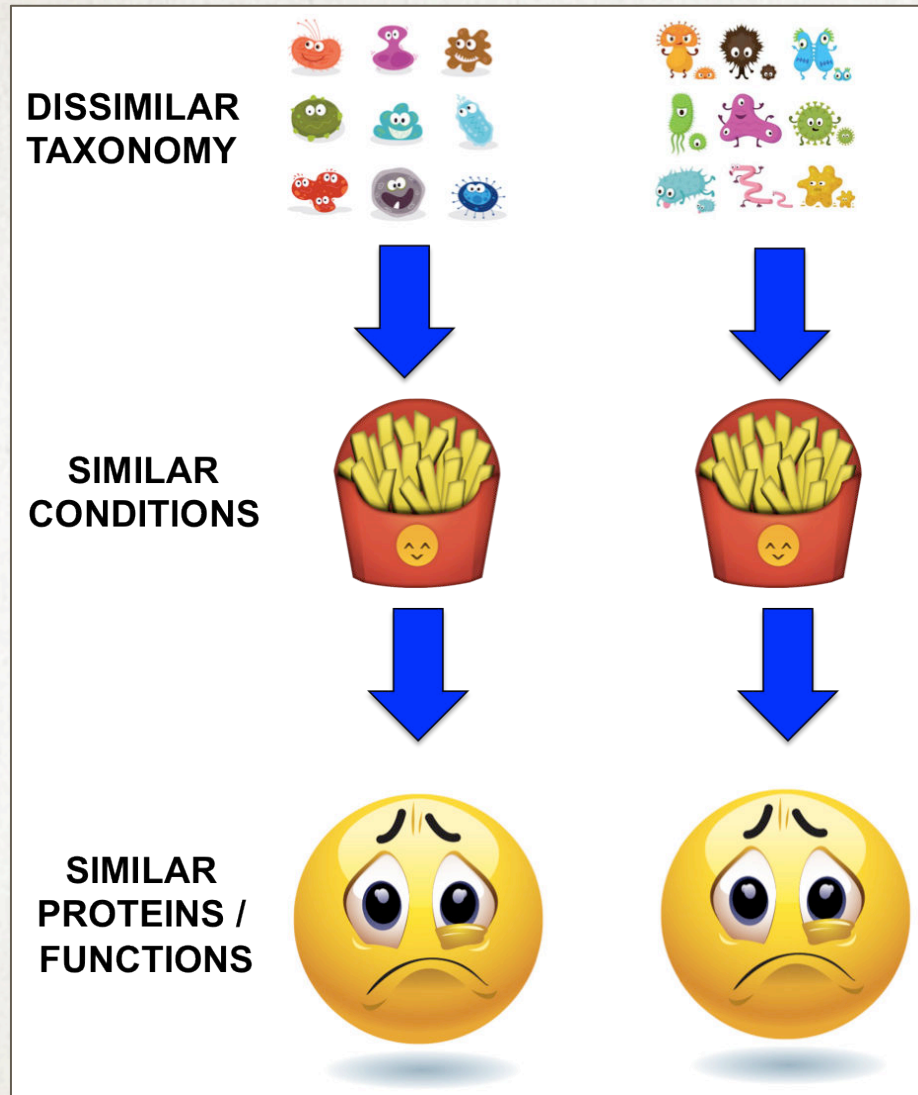
Metaproteomics:

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**.

TAXONOMY AND FUNCTIONAL EXPRESSION



METAPROTEOMIC CHALLENGES

SINGLE-ORGANISM PROTEOMICS



METAPROTEOMICS



SEARCH DATABASE

- | | |
|--------------|---|
| - SIZE | SMALL TO MEDIUM SIZE (10 K TO 100K SEQUENCES) |
| - COMPLEXITY | SINGLE + CONTAMINANTS |

LARGE (1 MILLION AND ABOVE)

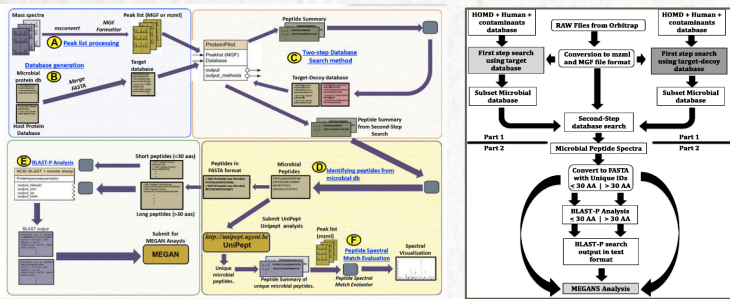
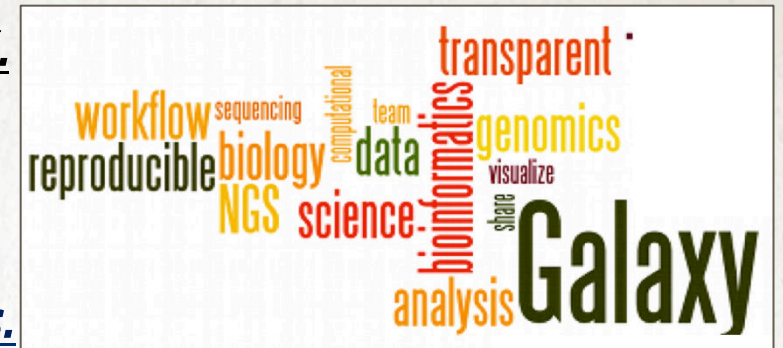
MULTI-ORGANISM DATABASE WITH HOMOLOGOUS PROTEINS

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

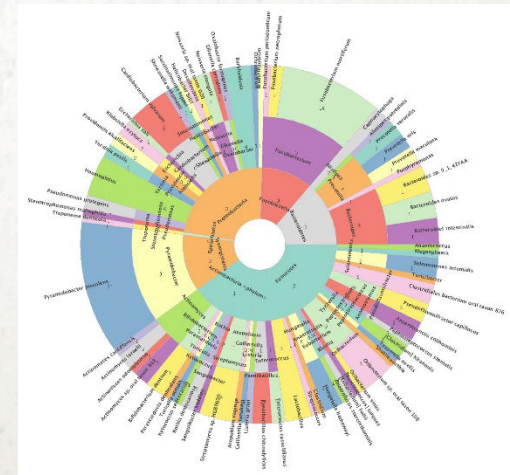
METAPROTEOMICS IN GALAXY

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

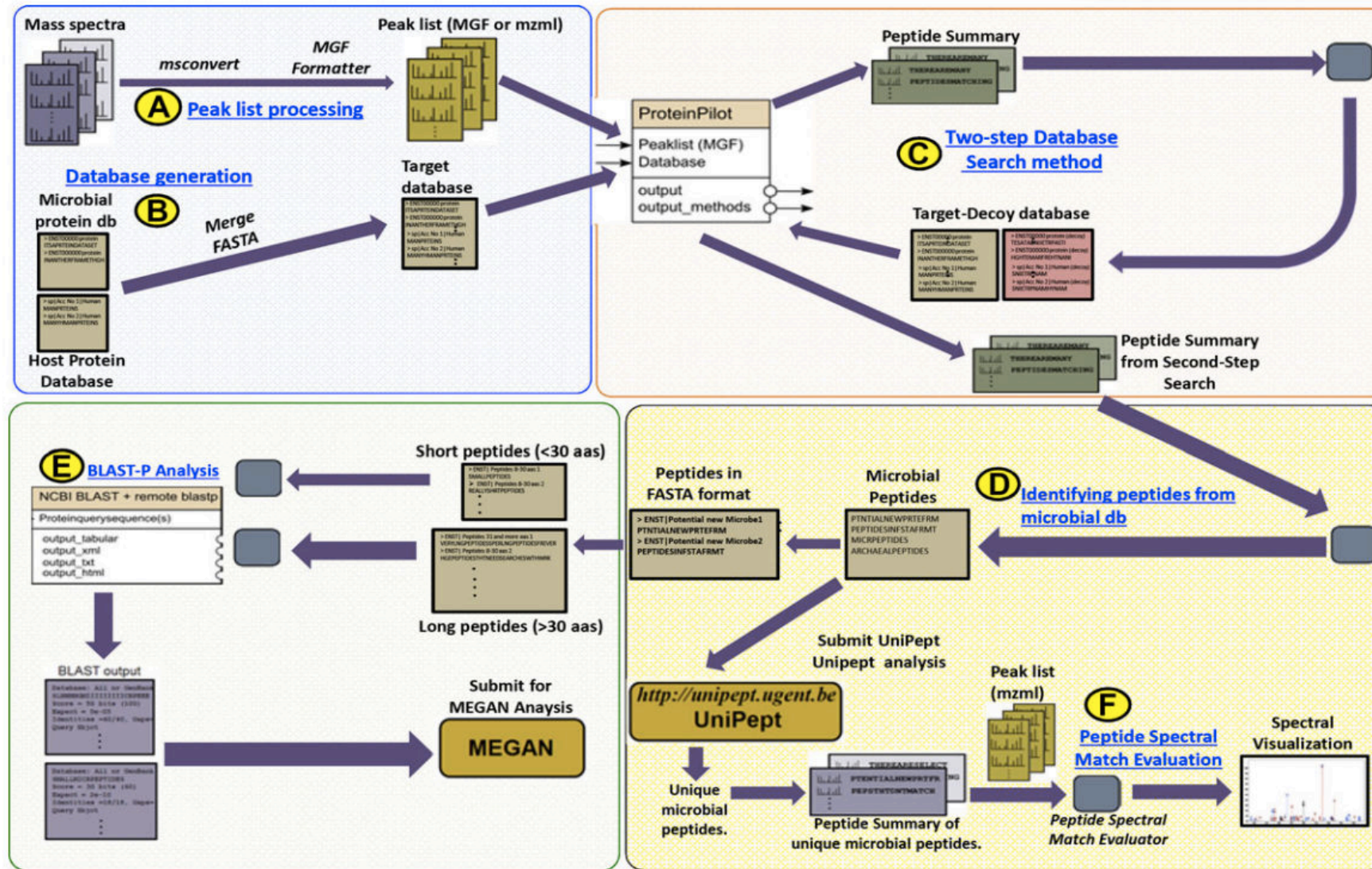


- **DEVELOPED METAPROTEOMICS WORKFLOWS WITHIN GALAXY**
Metaproteomic analysis using the Galaxy framework. Proteomics. (2015) doi:10.1002/pmic.201500074.

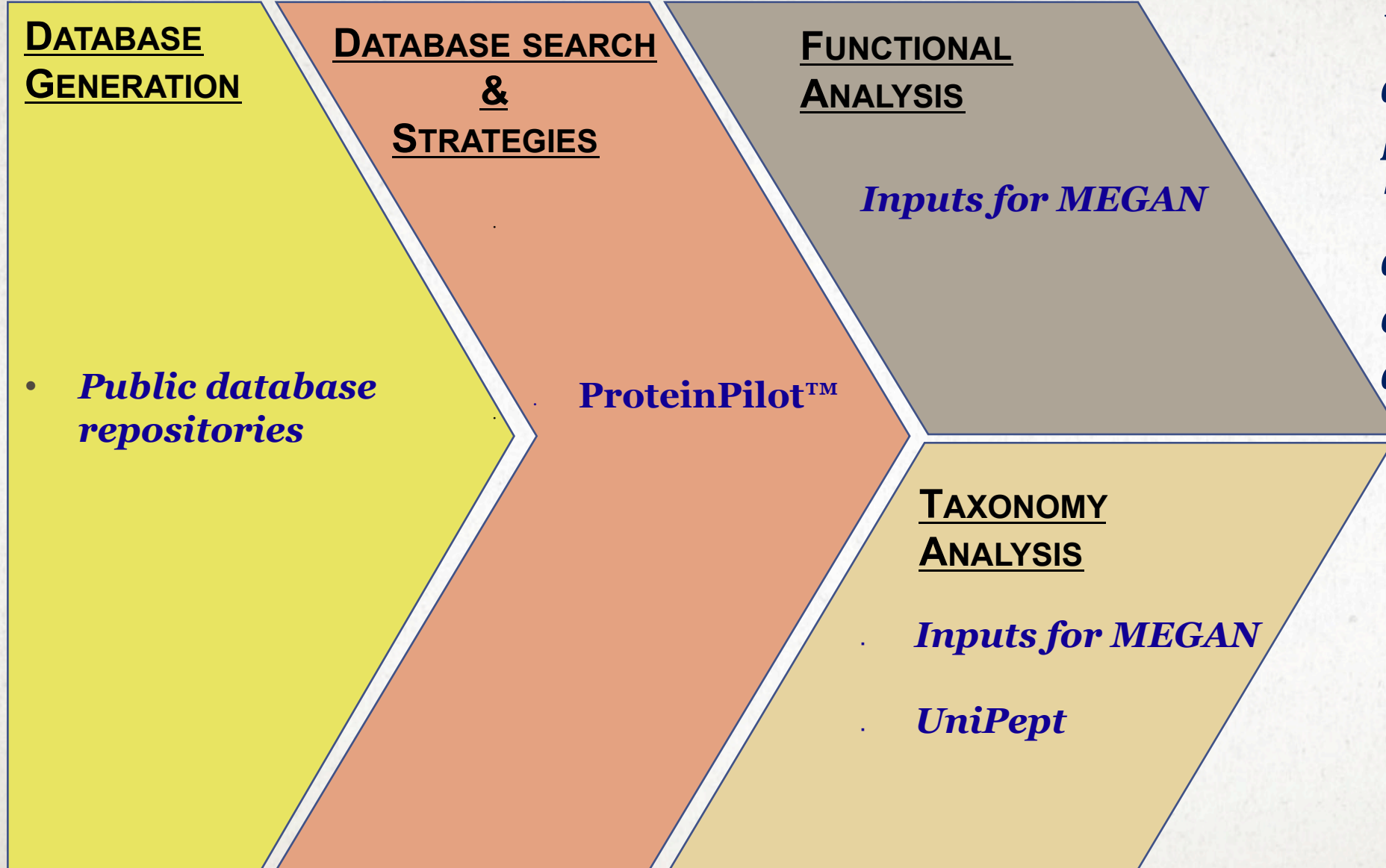
- **UTILIZED WORKFLOWS IN RESEARCH STUDIES**
- **Rudney et al., BMC Microbiome DOI: 10.1186/s40168-015-0136-z**
- **Bhargava et al., Bronchoalveolar Lavage Fluid Metaproteome in Acute Respiratory Failure (In Preparation)**
- **Skubitz et al., Metaproteomic analysis of human cervical-vaginal fluid in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



METAPROTEOMICS WORKFLOWS IN GALAXY



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.

COMMUNITY-BASED DEVELOPMENT

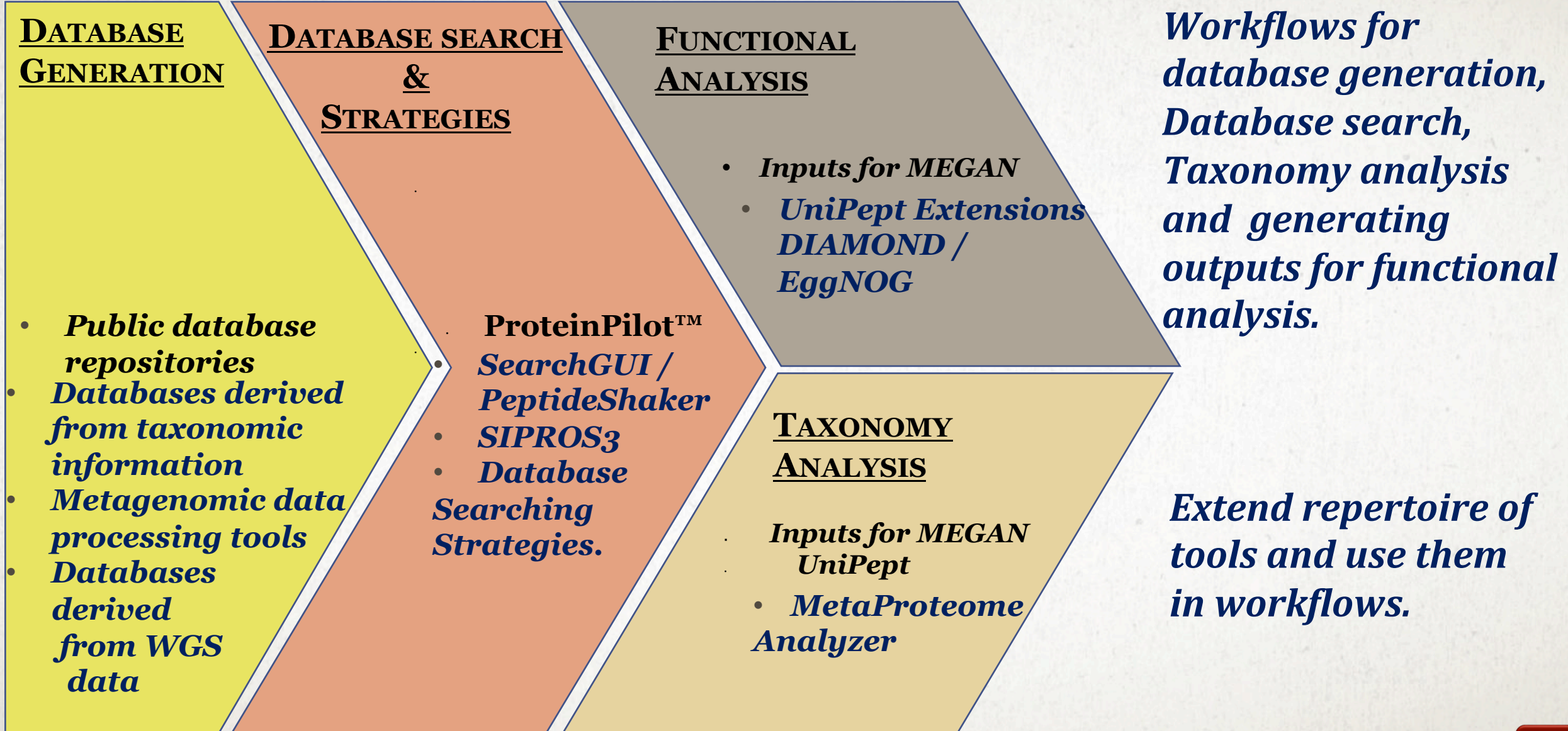
- 1 Email communication with metaproteomics researchers and developers around the world



- 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



CONTRIBUTION-FEST

1 Email communication with metaproteomics researchers and developers around the world



A world map with blue dots indicating the locations of contributors. The dots are labeled with names: Joel Rudney, James Johnson, Dave Clements, Praveen Kumar, Josh Elias, Thomas Doak, Judson Hervey, Bart Mesuere, Felix Van der Jeugt, Alessandro Tanca, Carolin Kolmeder, Thilo Muth, Bjoern Gruening, Clemens Blank, Bérénice Batut, and Ira Cooke.

Joel Rudney
James Johnson
Dave Clements
Praveen Kumar
Josh Elias
Thomas Doak
Judson Hervey
Bart Mesuere
Felix Van der Jeugt
Alessandro Tanca
Carolin Kolmeder
Thilo Muth
Bjoern Gruening
Clemens Blank
Bérénice Batut
Ira Cooke

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”.

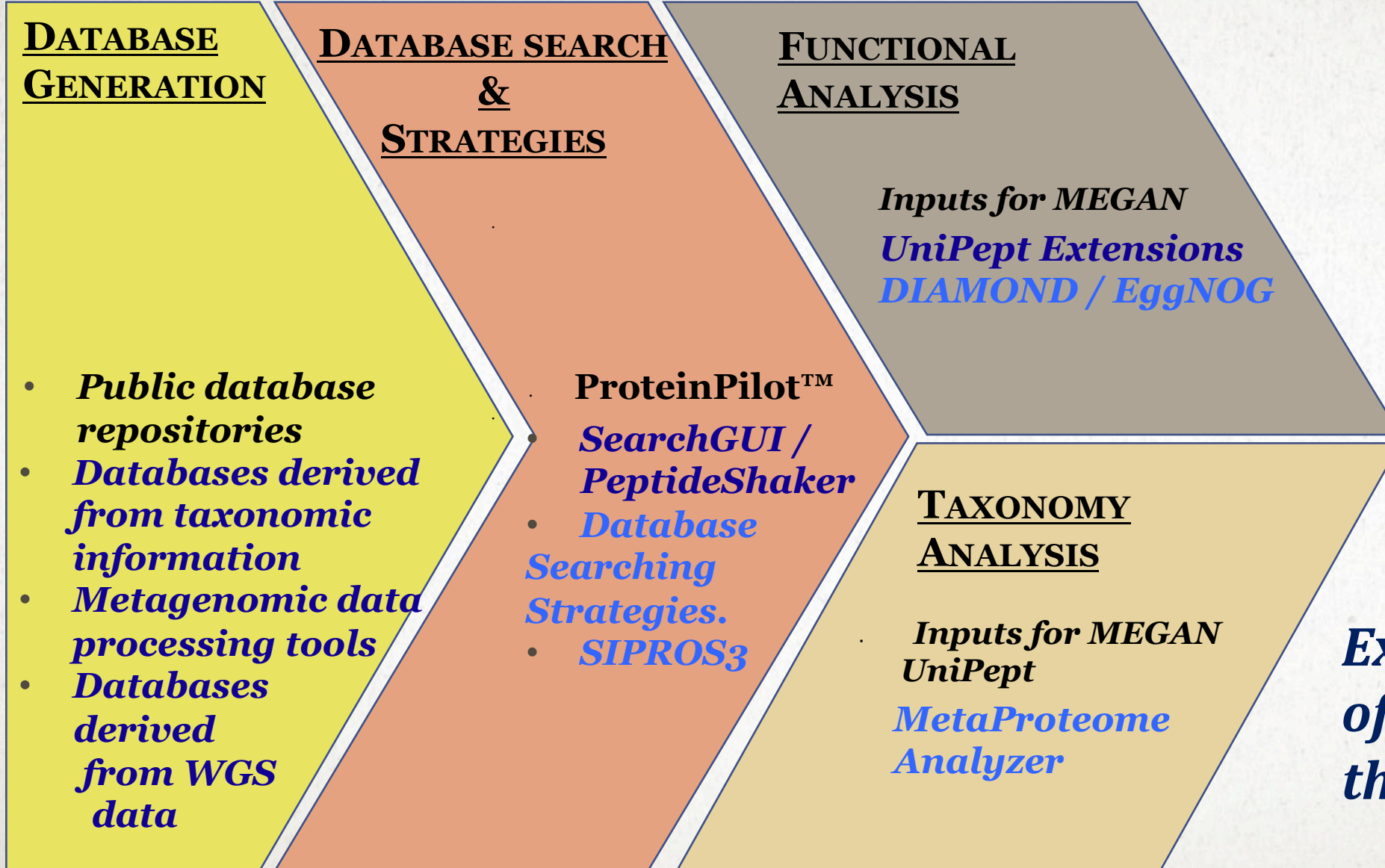
4 “Contribution Fest” (led by Bjoern Gruening) discussed about packaging and testing of tools for Galaxy.

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

6 Tools available for worldwide Galaxy users via Galaxy Toolshed

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

METAPROTEOMICS IN GALAXY

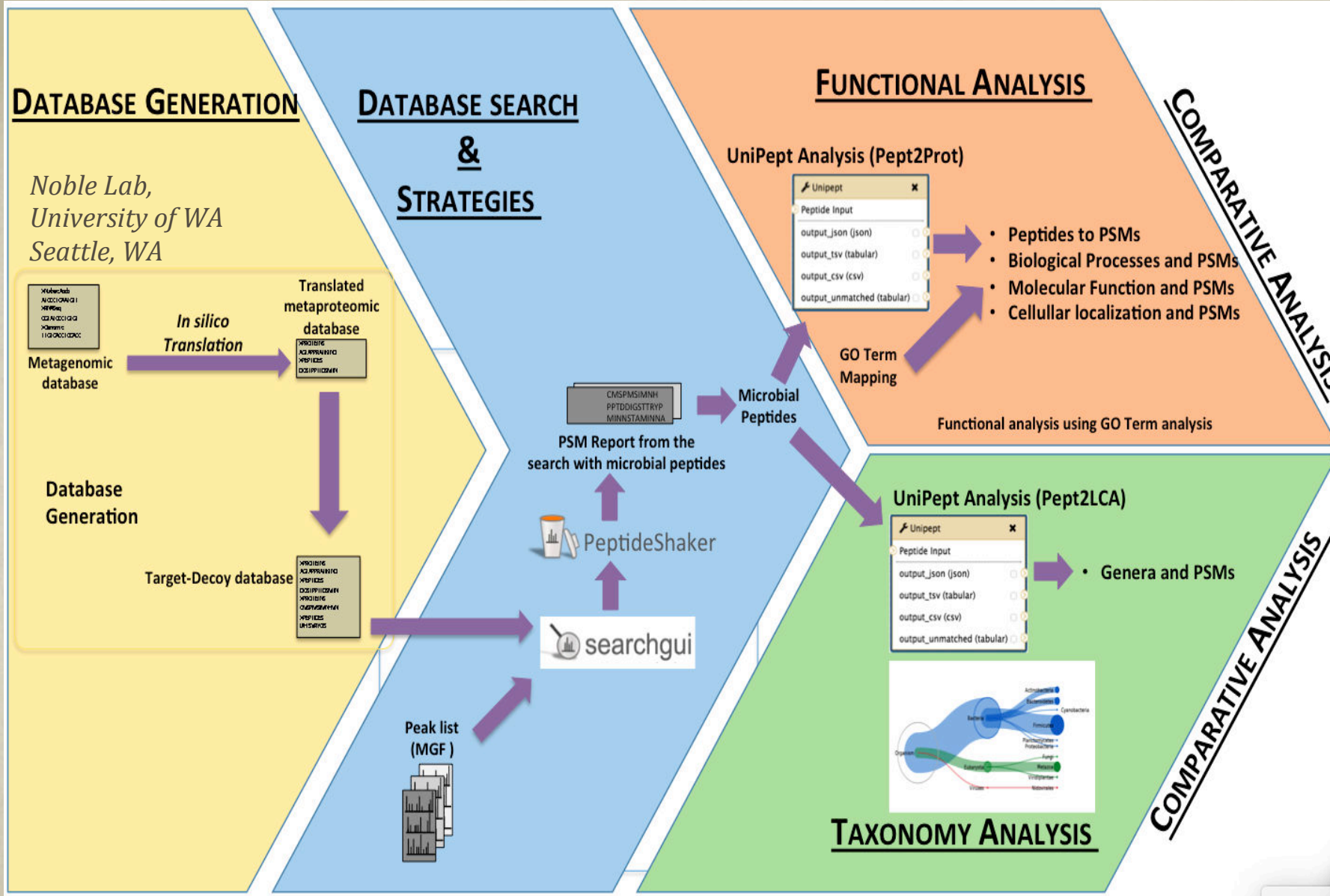


Extended repertoire of tools and used them in workflows.

**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/asms2017/>
- <https://galaxyproteomics.github.io/gcc2017/>



METAPROTEOMICS GATEWAY

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy' logo, 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', 'User', and a grid icon. The right side of the bar shows 'Using 0 bytes'. The left sidebar contains a 'Tools' section with a search bar and a list of tool categories: Get Data, Collection Operations, Text Manipulation, Filter and Sort, Join, Subtract and Group, Convert Formats, Extract Features, Fetch Sequences, Fetch Alignments, Statistics, Graph/Display Data, FASTA manipulation, MetaProteomics, MS Data Conversion, Sequence Database Tools, Protein/Peptide Search, Algorithms, and Uniprot retrieve taxonomy for peptides. Below these are 'Workflows' and a link to 'All workflows'. The main content area has a large blue and green header with the text 'An Accessible Metaproteomics Resource for Researchers' and the URL 'http://z.umn.edu/metaproteomicsgateway'. Below this, there are sections for 'Instructions for accessing Jetstream metaproteomics gateway' (with URL 'http://z.umn.edu/accessmpgwjetstream') and 'Instructions for accessing workflows and training material'. The latter section includes links to instructional materials from the ABRF 2017 satellite workshop and a list of slides and documentation from the ABRF 2017 Workshop. A footer note mentions collaboration between Galaxy-P team, Jetstream instance, and members of the Metaproteomics Consortium. The right sidebar shows a 'History' section with a search bar and a message that the history is empty, with links to 'load your own data' or 'get data from an external source'.

← → ↻ ⓘ 129.114.16.192

Galaxy Analyze Data Workflow Shared Data Visualization Help User ⌵

Tools

search tools

Get Data
[Collection Operations](#)
[Text Manipulation](#)
[Filter and Sort](#)
[Join, Subtract and Group](#)
[Convert Formats](#)
[Extract Features](#)
[Fetch Sequences](#)
[Fetch Alignments](#)
[Statistics](#)
[Graph/Display Data](#)
[FASTA manipulation](#)
[MetaProteomics](#)
[MS Data Conversion](#)
[Sequence Database Tools](#)
[Protein/Peptide Search](#)
[Algorithms](#)
[Uniprot](#) retrieve taxonomy for peptides

Workflows
▪ [All 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/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, Jetstream instance jetstream-cloud.org and the members of the Metaproteomics Consortium. <http://z.umn.edu/mphack2016>.

History

search datasets

Unnamed history
(empty)

This history is empty. You can [load your own data](#) or [get data from an external source](#)

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

z.umn.edu/metaproteomicsgateway



BACK TO THE FUTURE: LIGHTNING TALK AT GCC 2016

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

Project goals & scope Experiment Design

- Clinical study?
Environmental study?
- 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

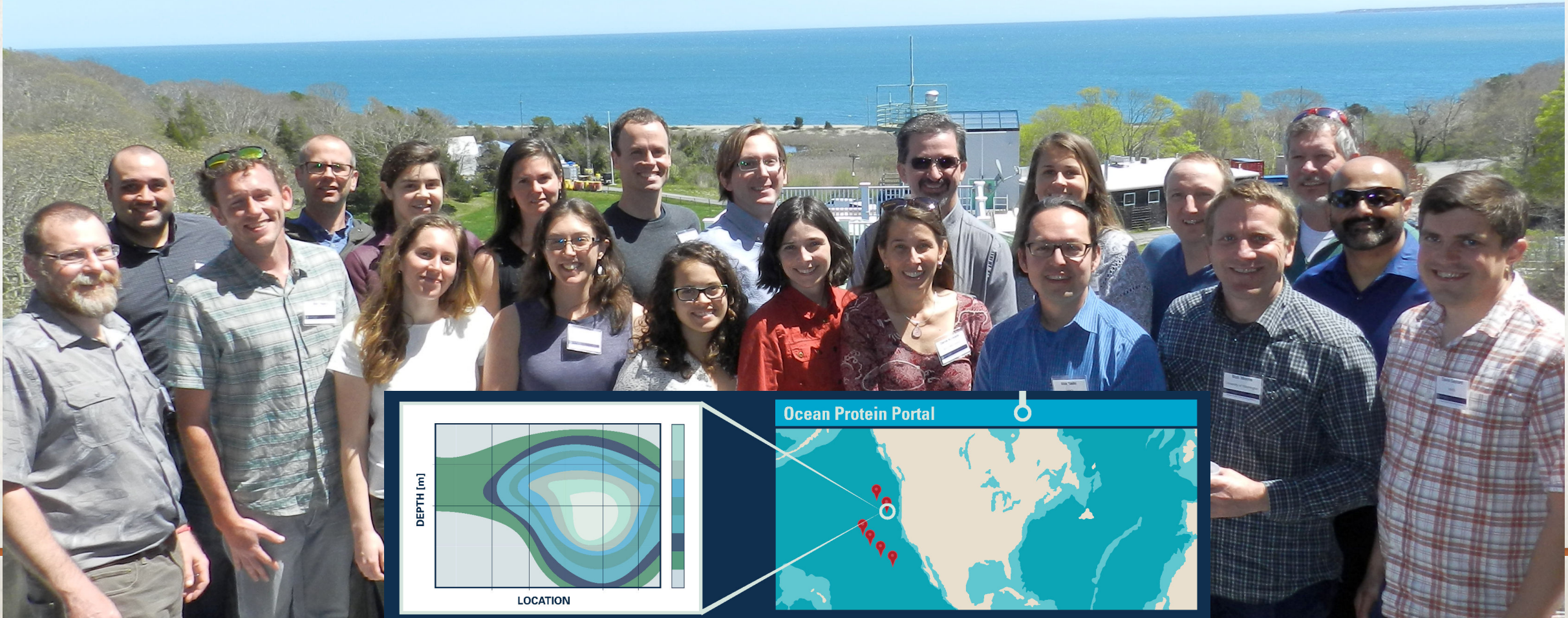
- Metaproteomics contribution-fest

Manuscript writing and submission

- Community effort

METAPROTEOMICS PROJECTS

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

University of Minnesota

TIM GRIFFIN

Praveen Kumar
Candace Guerrero
Subina Mehta
Shane Hubler

Dave Clements

Galaxy Project



Biologists / collaborators

Joel Rudney

Maneesh Bhargava

Kevin Viken

Amy Skubitz

Kristin Boylan

Marnie Peterson

Somiah Afiuni

Chris Wendt

Brian Sandri

Alexa Pragman

Josh Elias
Stanford
University

Thomas Doak
Indiana
University

Matt Chambers
Vanderbilt University,
Nashville, TN

Judson Hervey
Naval Research
Institute
Washington, D.C.

Minnesota
Supercomputing
Institute

**James
Johnson**
Tom McGowan
Michael Milligan

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

Bart Mesuere
Felix Van der Jeugt
VIB, Ugent,
Belgium

**Carolin
Kolmeder**
University of
Helsinki,
Finland

Thilo Muth
Bernhard Renard
Robert Koch Institut

Alessandro Tanca
Porto Conte
Ricerche, Italy

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

Ira Cooke
Melbourne , Australia