

# Advanced MS Analysis: Metaproteomics

**Pratik Jagtap**  
**University of Minnesota**



*Learn more at [galaxyp.org](http://galaxyp.org)*  
[z.umn.edu/itcrgalaxyvideo](http://z.umn.edu/itcrgalaxyvideo)

***International Mass Spectrometry Conference***  
***28th August, 2022***

# Workshop acknowledgements

- **Instructors**

- Pratik Jagtap



- **Other contributors**

- Tim Griffin
- Subina Mehta
- James Johnson
- Andrew Rajczewski
- Reid Wagner
- Katherine Do
- usegalaxy.eu team
- Galaxy community



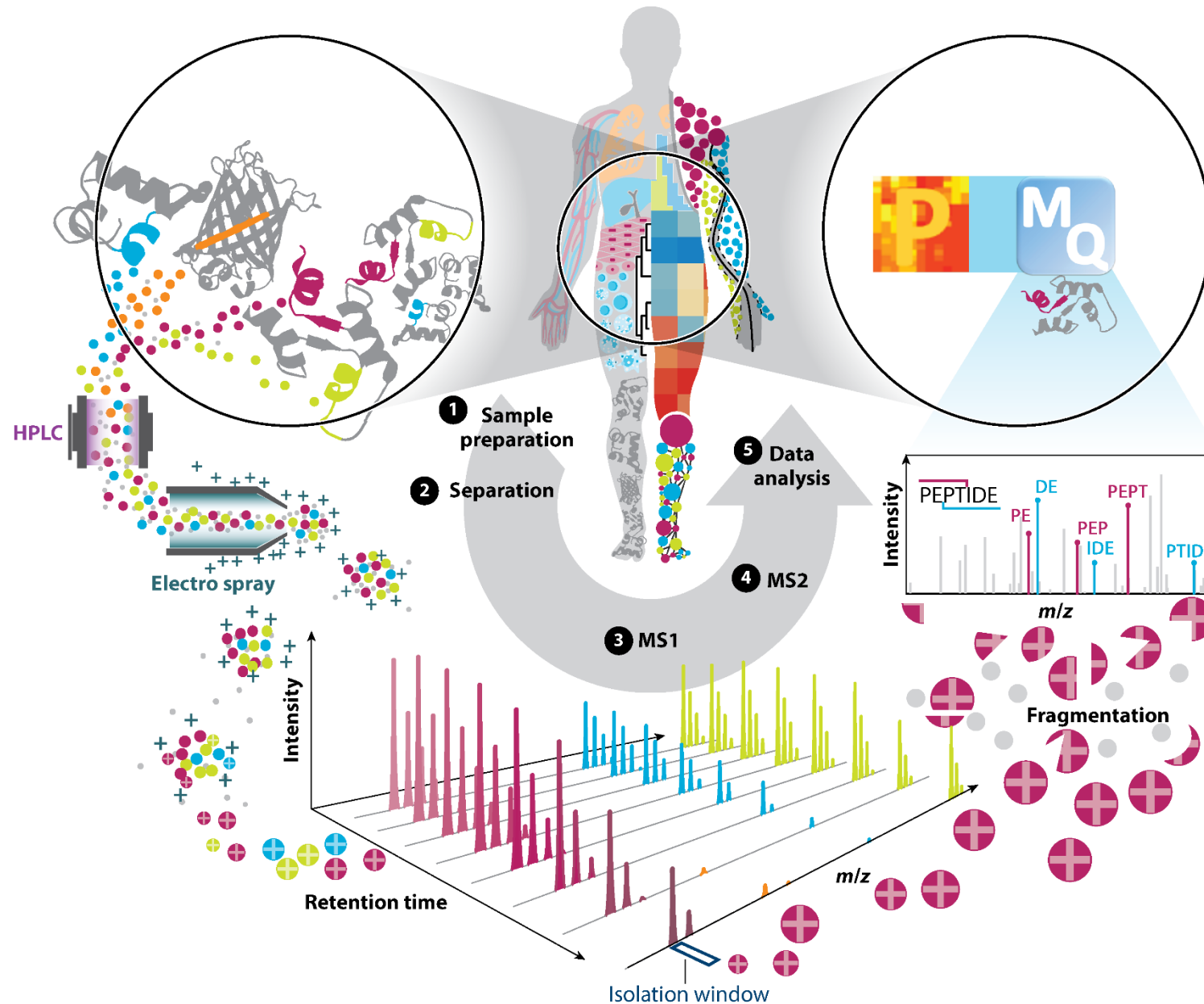
**NATIONAL CANCER INSTITUTE**  
Informatics Technology for  
Cancer Research



<https://training.galaxyproject.org/training-material/topics/proteomics/tutorials/metaproteomics/tutorial.html>

<https://gallantries.github.io/video-library/videos/proteomics/metaproteomics/tutorial/>

# MASS SPECTROMETRY AND PROTEOMICS



# MICROBIOME RESEARCH

## Microbiome

IN NUMBERS

**100 Trillion**

symbiotic microbes live in and on every person and make up the human microbiota

The human body has more microbes than there are stars in the milky way

**95%**

of our microbiota is located in the GI tract

**150:1**

The genes in your microbiome outnumber the genes in our genome by about 150 to one

The surface area of the **GI tract** is the same size as 2 tennis courts

**>10,000**

Number of different microbial species that researchers have identified living in and on the human body

You have

**1.3X**

more microbes than human cells

**2kg**

The gut microbiota can weigh up to 2Kg



Interfacing Food & Medicine

The microbiome is more medically accessible and manipulable than the human genome

**90%**

It is thought that of disease can be linked in some way back to the gut and health of the microbiome

**5:1**

**Viruses:Bacteria** in the gut microbiota

**2.5**

The number of times your body's microbes would circle the earth if positioned end to end



Each individual has a unique gut **microbiota**, as personal as a fingerprint



<https://worldmicrobiomeday.com/resources/>

<https://www.nature.com/articles/d41586-020-00193-3>

# MICROBIOME

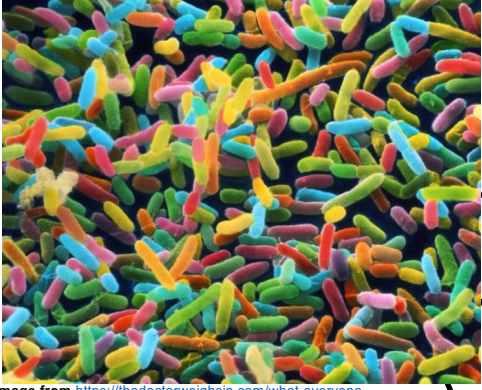
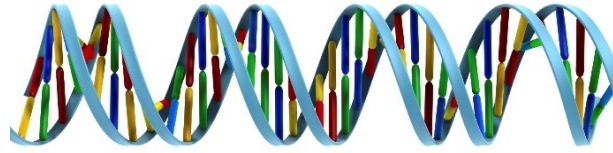


Image from <https://thedoctorweighsin.com/what-everyone-should-know-about-the-infant-microbiome/>

Potential to unravel the mechanistic details of microbial interactions with host / environment by analyzing the functional dynamics of the microbiome.



**METAGENOMICS**

**TAXONOMY**

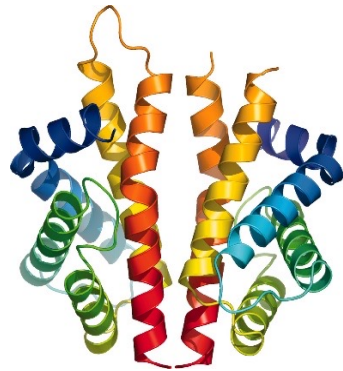
function



**METATRANSCRIPTOMICS**

**TAXONOMY**

**function**



**METAPROTEOMICS**

**TAXONOMY**

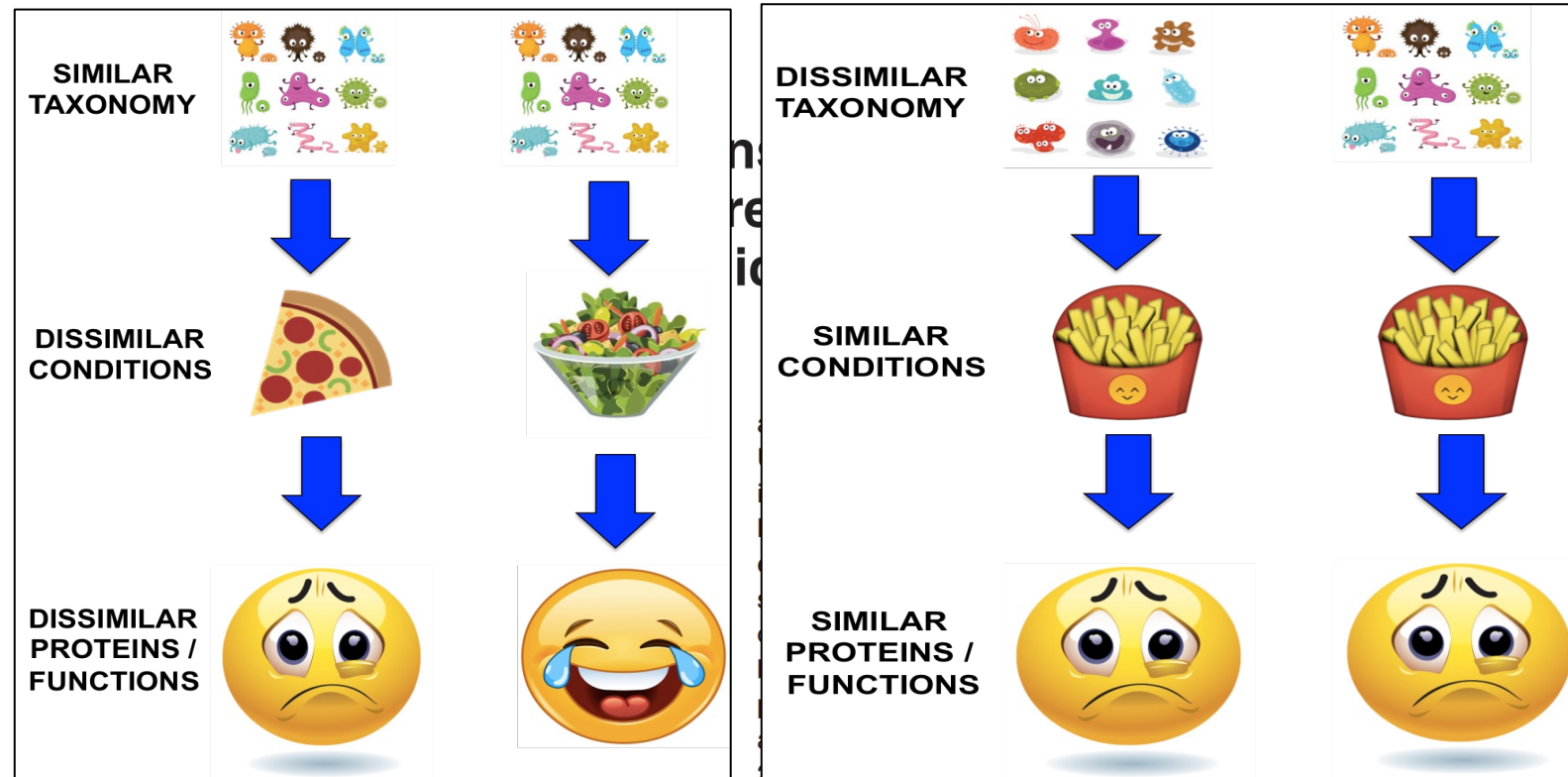
**FUNCTION**

# METAPROTEOMICS

Bond and Wilmes 2004

“The large-scale characterization of the entire protein complement of environmental microbiota at a given point in time”

*Environ. Microbiol.* 6, 911–920.

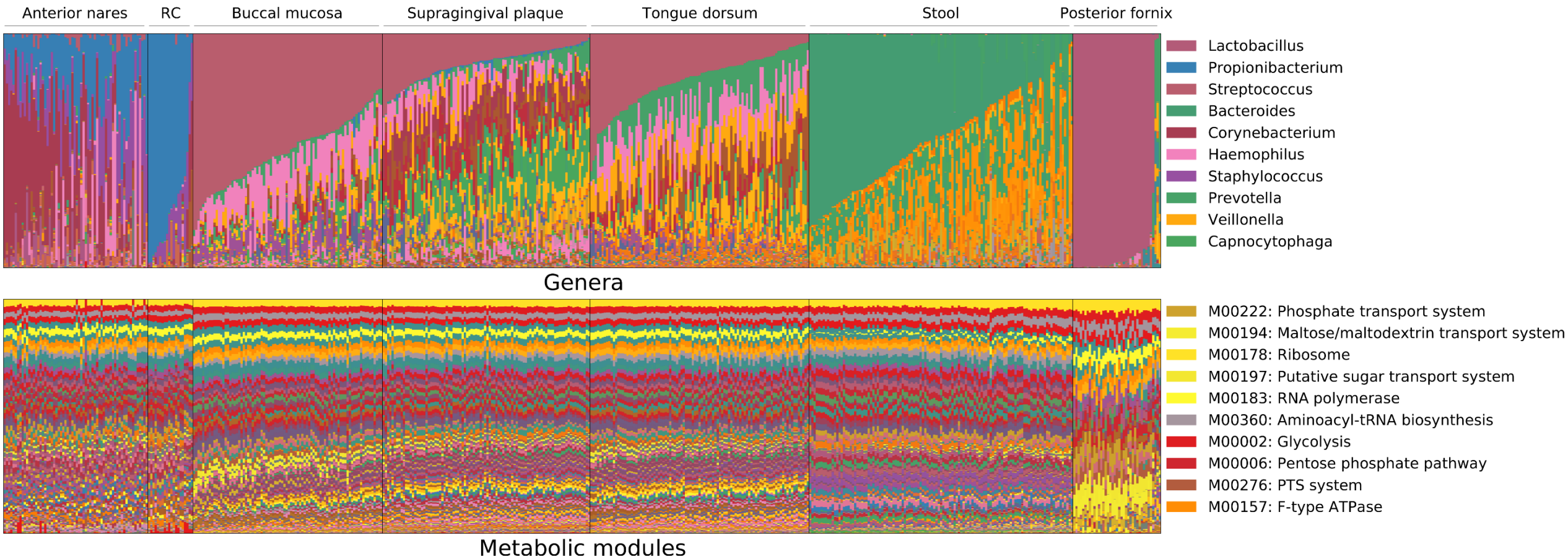


Bond and Wilmes 2015

“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.”

*Proteomics.* doi:10.1002/pmic.201500183.

# MICROBIAL TAXA VARY WHILE METABOLIC PATHWAYS REMAIN STABLE WITHIN A HEALTHY POPULATION



Human Microbiome Project Consortium (2012). *Nature*. 486(7402):207-214. doi: 10.1038/nature11234.

# METAPROTEOMICS ANALYTICAL CHALLENGES

## SINGLE-ORGANISM PROTEOMICS



## METAPROTEOMICS



### SEARCH

#### DATABASE SIZE

SMALL TO MEDIUM SIZE (10 K  
TO 100K SEQUENCES)

SINGLE + CONTAMINANTS

LARGE (1 MILLION AND ABOVE)

MULTI-ORGANISM DATABASE WITH HOMOLOGOUS  
PROTEINS

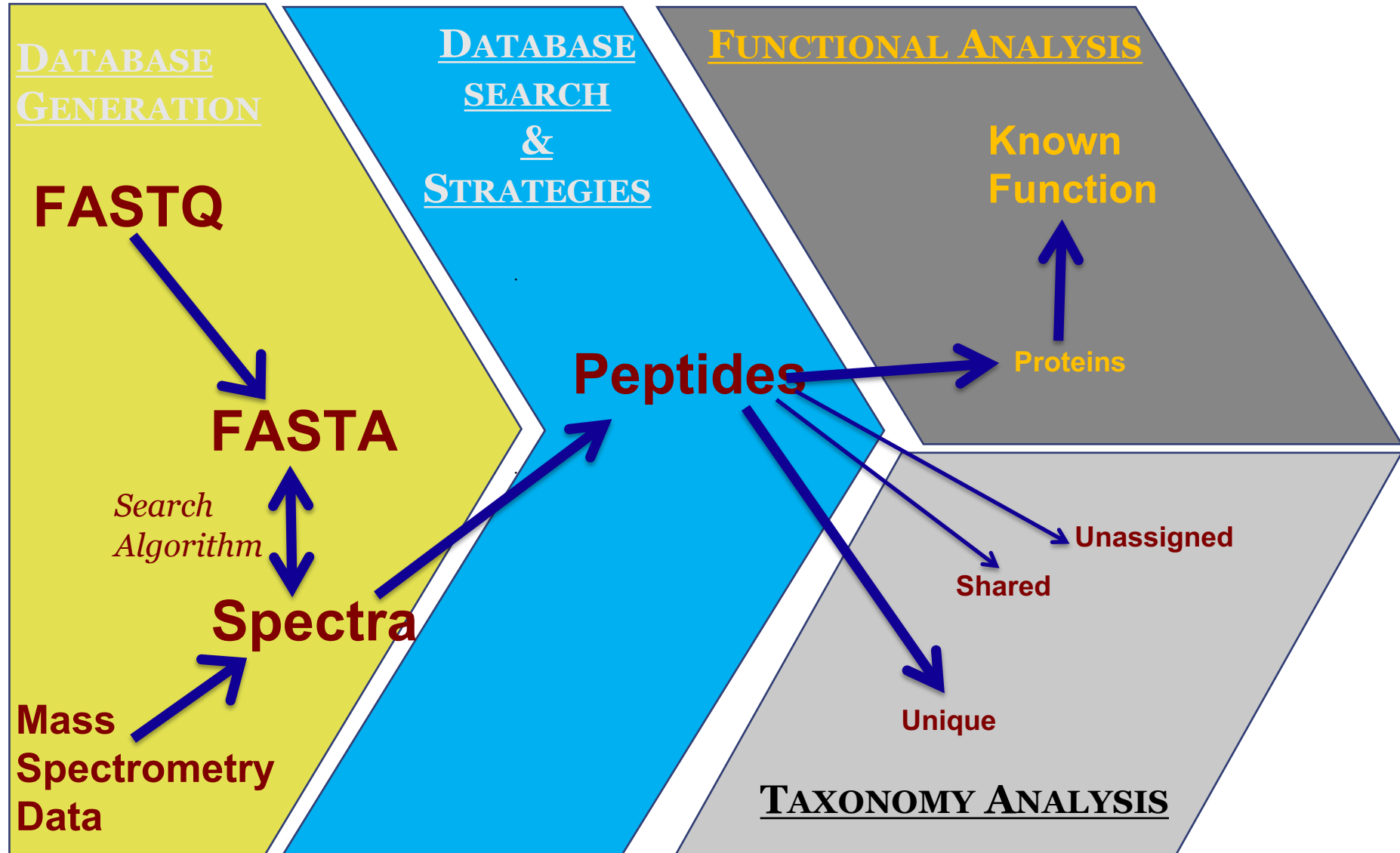
#### COMPLEXITY

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 WORKFLOW



# HANDS-ON SESSION

## Instructions



Please **Register** for creating an account with a valid email ID and Password at [usegalaxy.eu](https://usegalaxy.eu).

Once Registered, click on TIAAS to join the IMSC 2022 Galaxy session.  
[https://usegalaxy.eu/join-training/imsc\\_galaxy\\_training](https://usegalaxy.eu/join-training/imsc_galaxy_training)

Go to Shared Data Published Histories

### Published Histories

Advanced Search

Name	Annotation	Owner
Inputs for Metaproteomics-GTN		subina

Go to Shared Data Published Workflows

### Published Workflows

Advanced Search

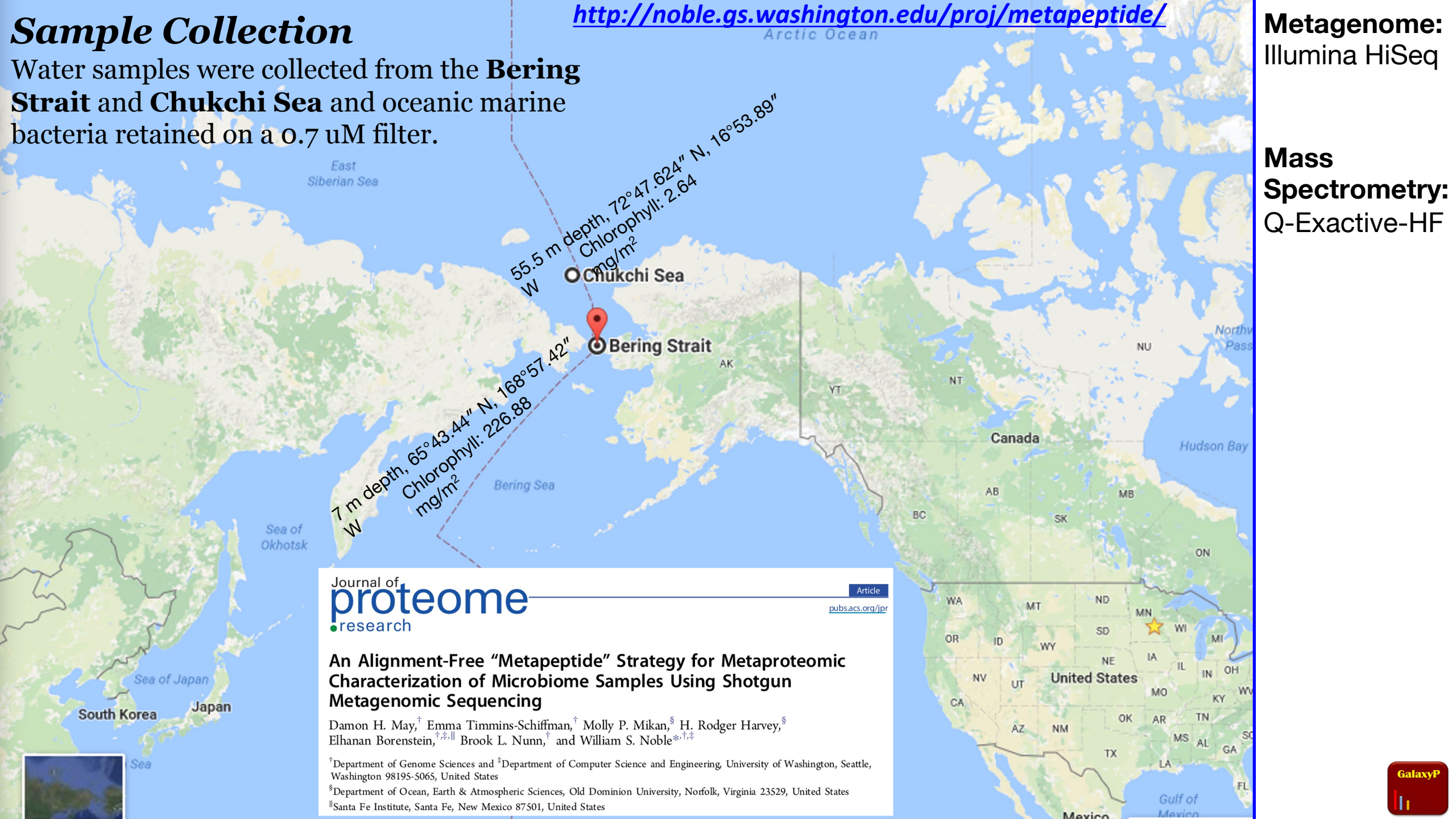
Name	Annotation	Owner
Metaproteomics_GTN_IMSC2022_Aug28		pratikjagtap

Run the workflow on active history

# Sample Collection

Water samples were collected from the **Bering Strait** and **Chukchi Sea** and oceanic marine bacteria retained on a 0.7  $\mu\text{M}$  filter.

<http://noble.gs.washington.edu/proj/metapeptide/>  
Arctic Ocean



**Metagenome:**  
Illumina HiSeq

**Mass Spectrometry:**  
Q-Exactive-HF

Journal of **proteome** research

Article  
[pubs.acs.org/jpr](https://pubs.acs.org/jpr)

### An Alignment-Free "Metapeptide" Strategy for Metaproteomic Characterization of Microbiome Samples Using Shotgun Metagenomic Sequencing

Damon H. May,<sup>†</sup> Emma Timmins-Schiffman,<sup>†</sup> Molly P. Mikan,<sup>§</sup> H. Rodger Harvey,<sup>§</sup> Elhanan Borenstein,<sup>†,‡,||</sup> Brook L. Nunn,<sup>†</sup> and William S. Noble<sup>\*,†,‡</sup>

<sup>†</sup>Department of Genome Sciences and <sup>‡</sup>Department of Computer Science and Engineering, University of Washington, Seattle, Washington 98195-5065, United States  
<sup>§</sup>Department of Ocean, Earth & Atmospheric Sciences, Old Dominion University, Norfolk, Virginia 23529, United States  
<sup>||</sup>Santa Fe Institute, Santa Fe, New Mexico 87501, United States



# SearchGUI

## SearchGUI matches MS/MS spectra to peptide sequences

- SearchGUI allows for multiple search engines to run simultaneously
- Specific digestion conditions can be selected
- Mass spectrometer parameters can be selected to maximize the efficacy of spectral matches
- Post-Translational Modifications (PTMs) can be added to the search parameters

### DATABASE GENERATION

**FASTQ**

**FASTA**

*Search Algorithm*

**Spectra**

**Mass Spectrometry Data**

The screenshot displays three panels of search parameters in SearchGUI:

- Protein Digestion Options:** Includes a 'Digestion' section with a 'Select Enzymes' dropdown, an 'Enzymes' section with a '1: Enzymes' dropdown (set to 'Trypsin'), and a 'Maximum Missed Cleavages' input field (set to '2').
- Precursor Options:** Includes a 'Precursor Ion Tolerance Units' dropdown (set to 'Parts per million (ppm)'), a 'Precursor Ion Tolerance' input field (set to '100'), and a 'Fragment Tolerance Units' dropdown (set to 'Daltons').
- Protein Modification Options:** Includes a 'Fixed Modifications' section with a 'Cysteineamidomethylation' button, and a 'Variable Modifications' section with an 'Oxidation of M' button.

12: Peptide Shaker on data 10: PSM Report



11: Peptide Shaker on data 10: mzidentML file



10: Search GUI on data 7, data 6, and others



# PeptideShaker

## PeptideShaker filters SearchGUI results.

- Search GUI results are filtered by FDR to yield most confident peptide spectral matches (PSMs)
- Peptide Shaker generates outputs such as Protein Report, Peptide Report and mzIdentML files for subsequent analysis.

### DATABASE GENERATION

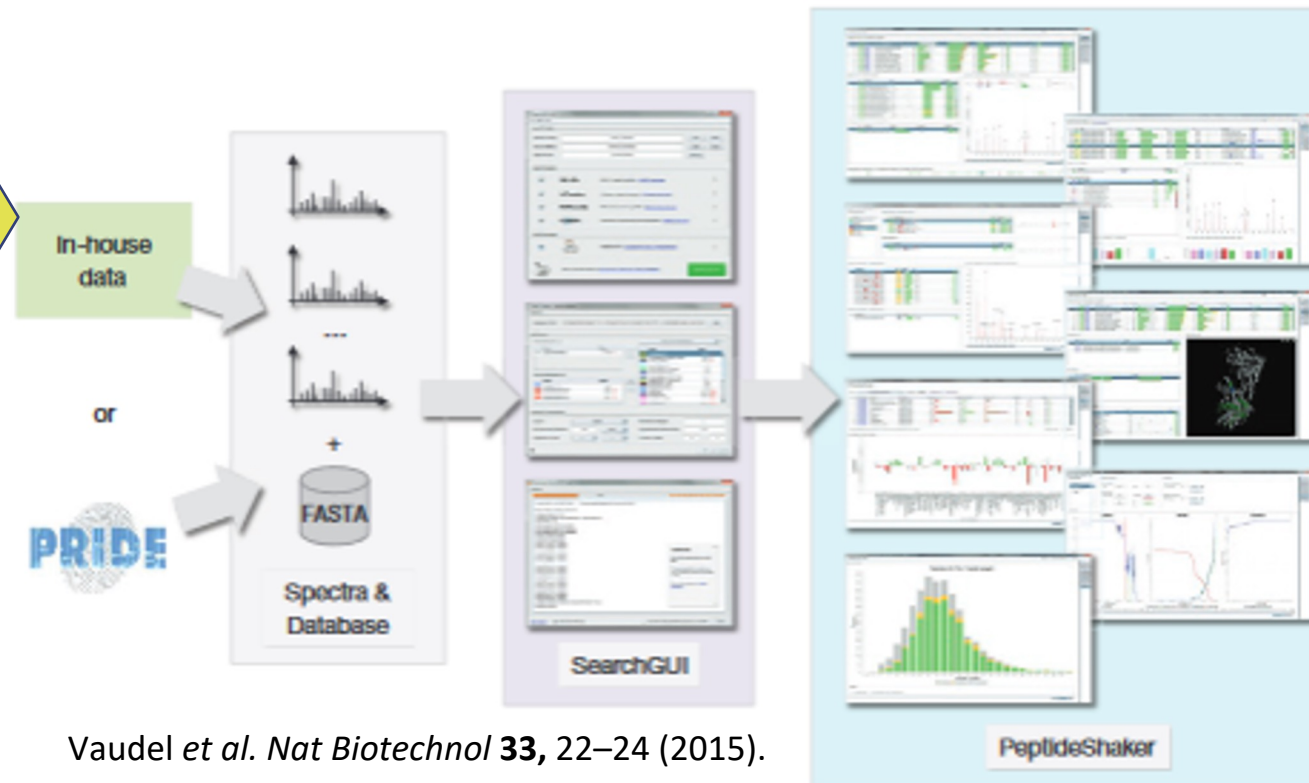
FASTQ

FASTA

Search Algorithm

Spectra

Mass Spectrometry Data



Vaudel et al. *Nat Biotechnol* **33**, 22–24 (2015).

12: Peptide Shaker on data 10: PSM Report



11: Peptide Shaker on data 10: mzidentML file

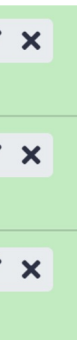
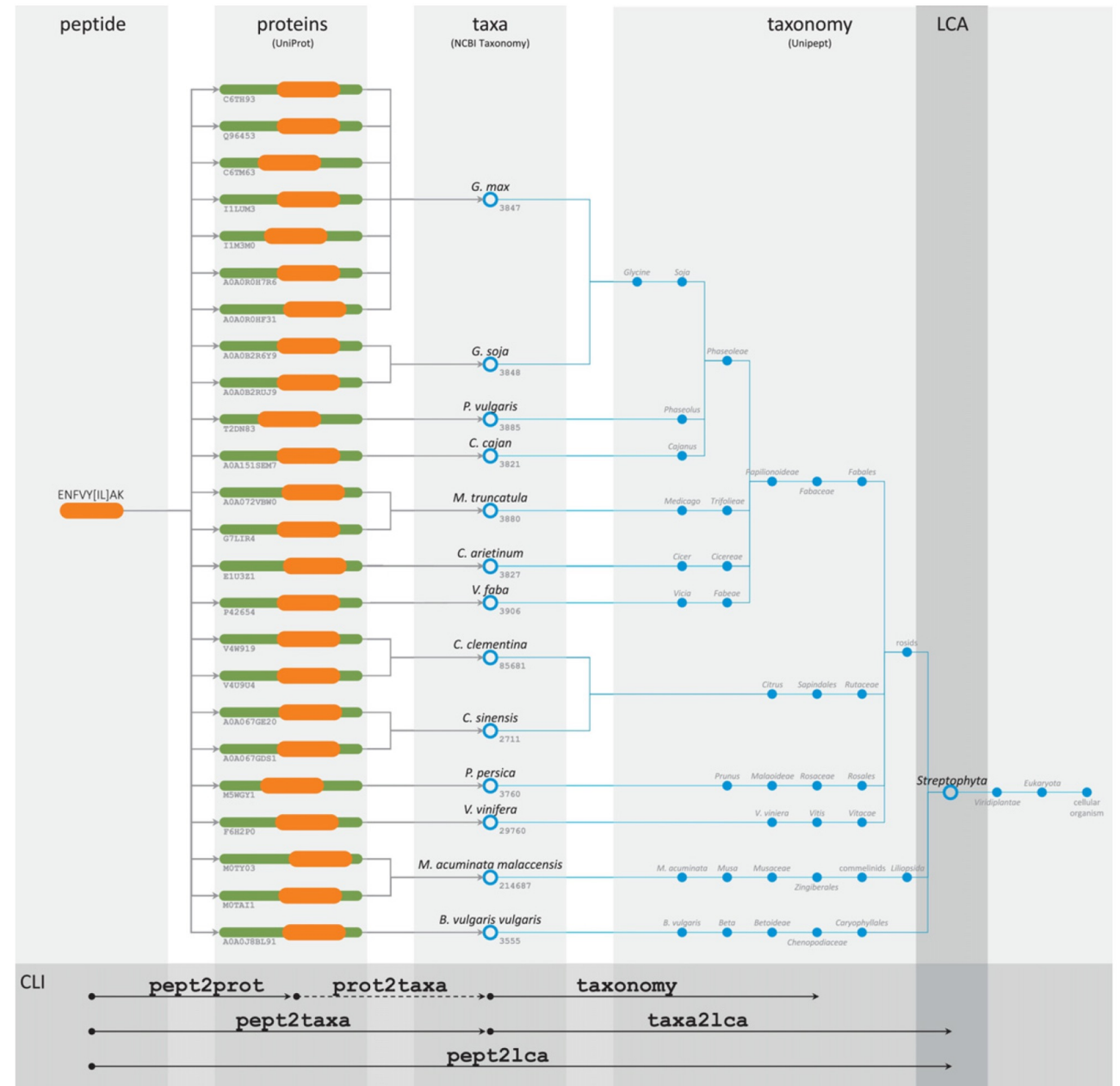
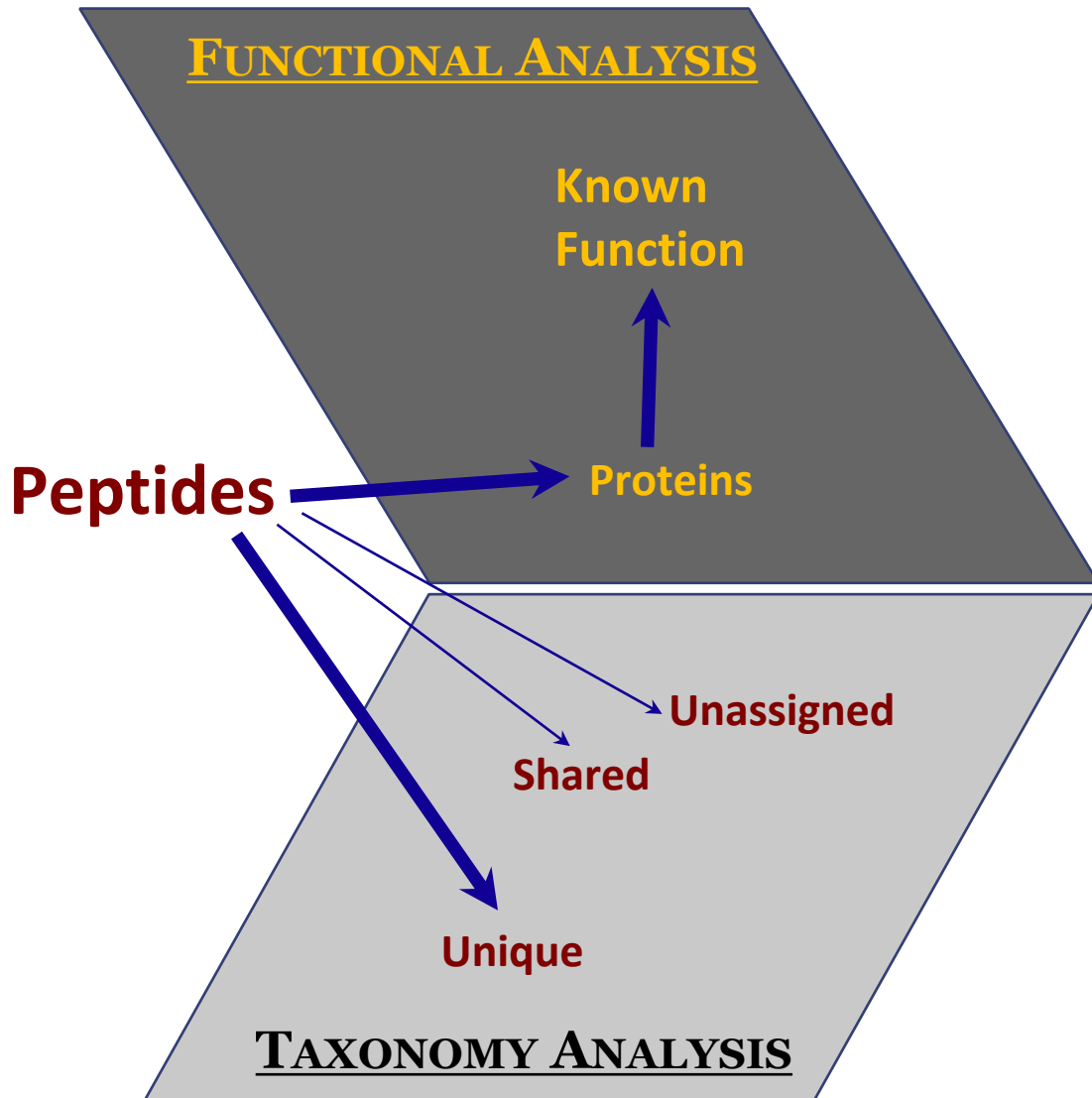


10: Search GUI on data 7, data 6, and others

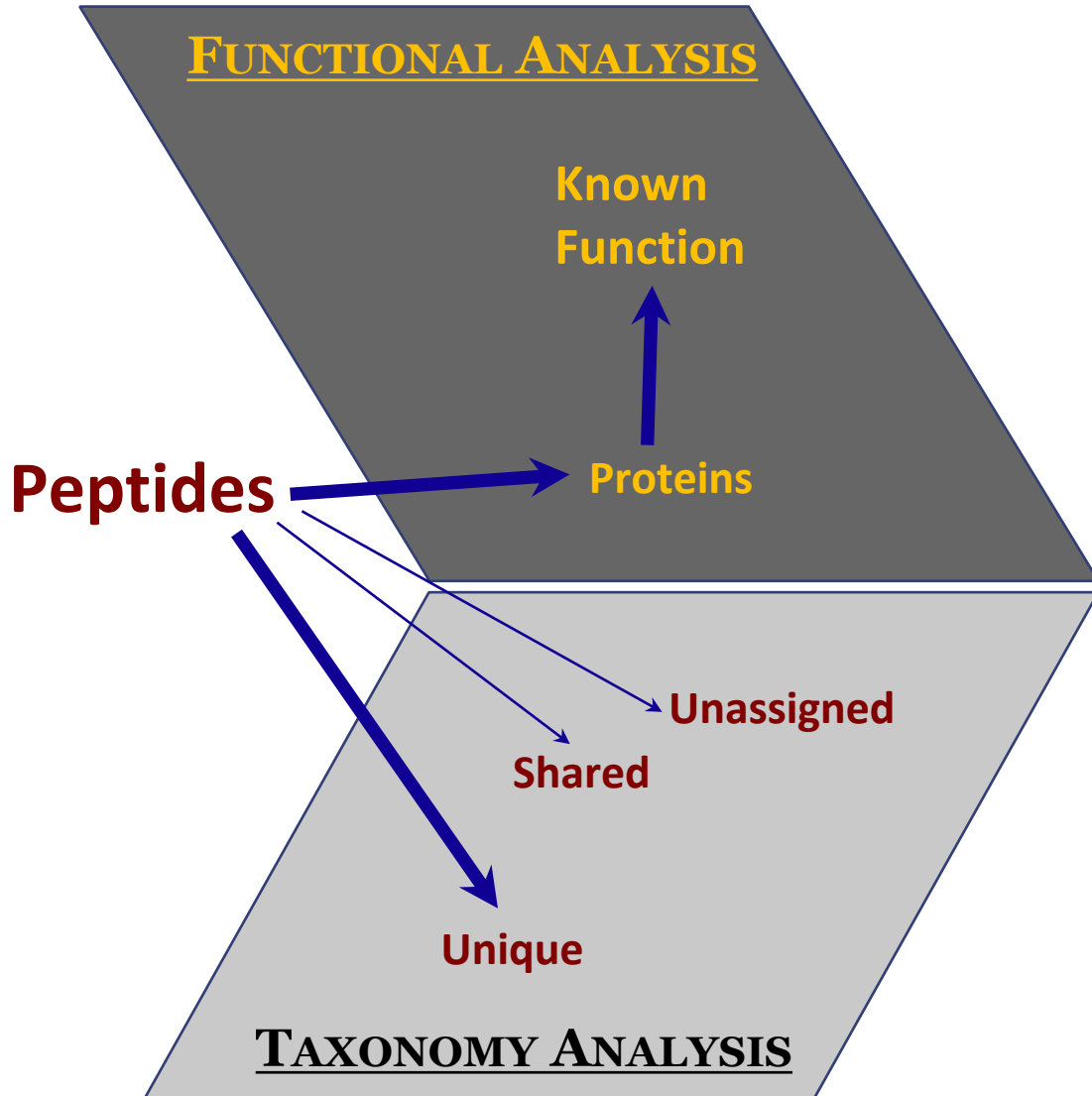


# Unipept

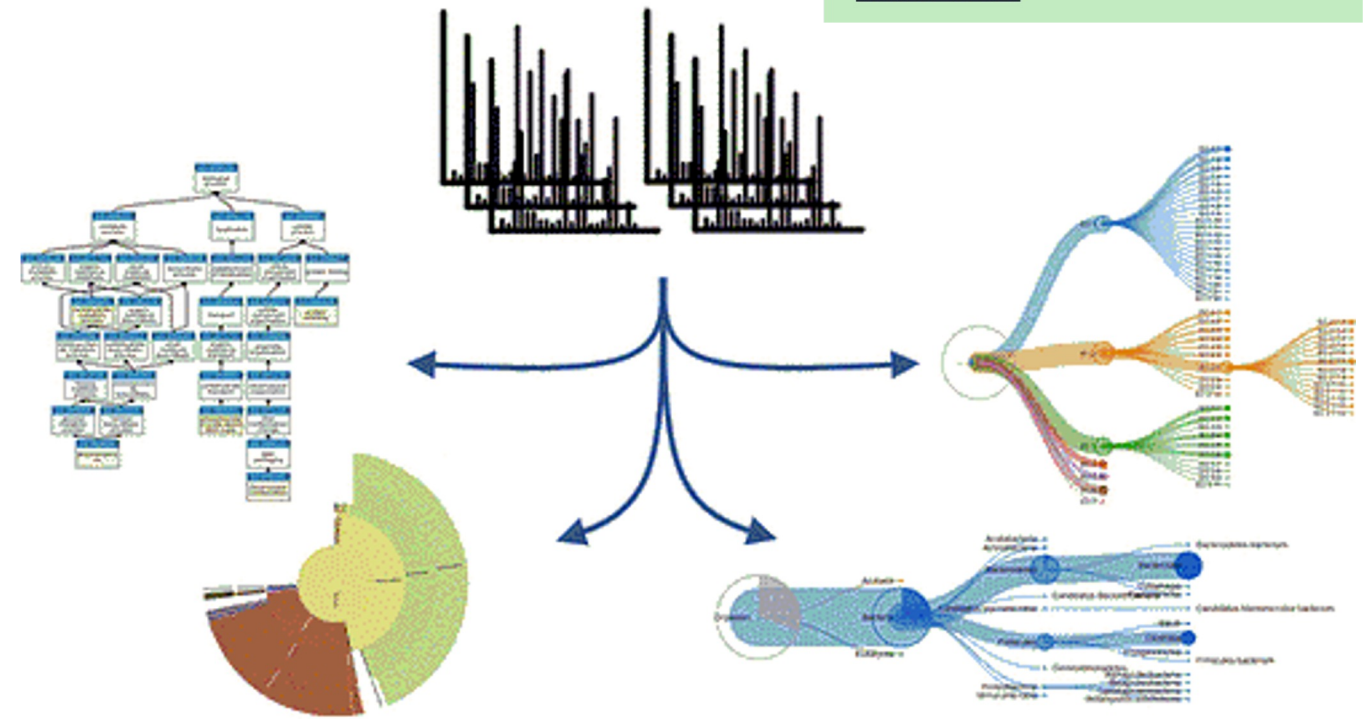
B. Mesuere et al. / Journal of Proteomics 171 (2018) 11–22



# Unipept

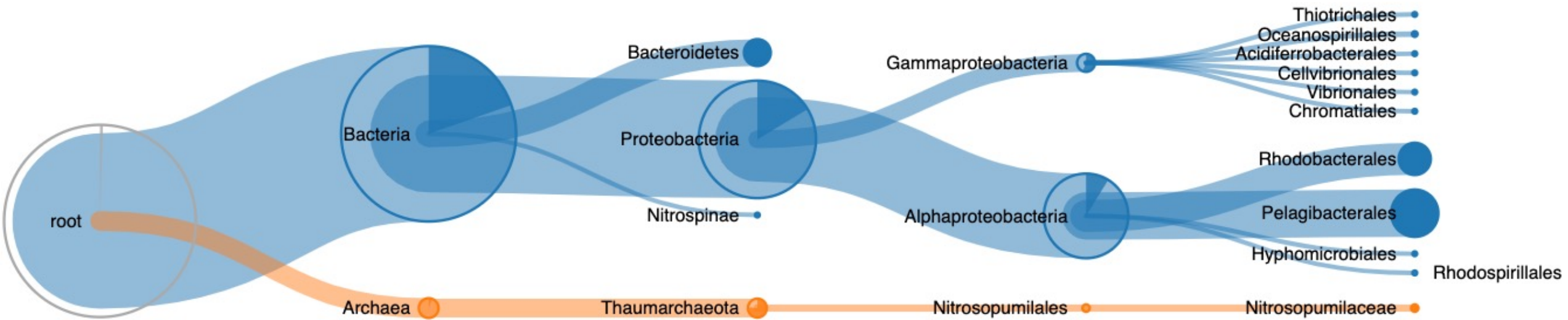


- 18: Unipept pept2lca on data 15 tsv
- 17: Unipept pept2lca on data 15 Taxonomy json
- 16: Unipept pept2prot on data 15 tsv



Taxonomic and Functional analysis with Unipept

<https://unipept.ugent.be/publications>



<https://usegalaxy.eu/u/pratikjagtap/h/metaproteomicsgtnimsc2022-completed>



# Who is there?

Get a taxonomy report from PSM report and Unipept pept2lca table

peptide	superkingdom	...	genus
AADGHTMHFDVITGEK	Archaea		Nitrosopumilus
AALESFTGNVTSALK	Bacteria		Polaribacter

PSM#	Proteins	Sequence	...
1	EGGEDMFVHKSDV...	<b>EGGEDMFVHK</b>	
2	GKRVAAAVGTVPE...	<b>VAAAVGTVPEQEWLK</b>	

## SQLite Relational Data base

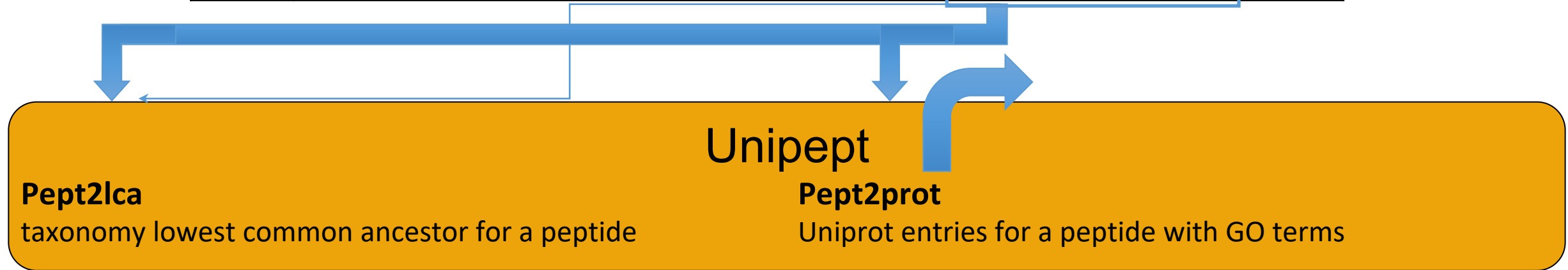
SQL query joins PSM and LCA to report number of PSMs and Peptides per genus

genus	PSMs	DISTINCT PEPTIDES
Planktomarina	161	20
Nitrosopumilus	122	27

# Who is there? What are they doing?

How do we get taxonomy and function of a microbiome from a list of peptides?

PSM#	Proteins	Sequence	...
1	EGGEDMFVHKSDVDGFINEGDK	EGGEDMFVHK	
2	GKRVAAAVGTVPPEQEWLK, KVAAAVGTVPPEQEWLK, RVAAAVGTV...	VAAAVGTVPPEQEWLK	



peptide	superkingdom	...	genus
AADGHTMHFDVITGEK	Archaea		Nitrosopumilus
AALESFTGNVTSALK	Bacteria		Polaribacter

peptide	uniprot_id	...	go_references
AAEKLAAQAR	W5T6F9		GO:0016021
AAEKLAAQAR	A0A0Q6ZKK0		GO:0005524 GO:0016887 GO:0015833

# Questions

- **How can I match metaproteomic mass spectrometry data to peptide sequences derived from shotgun metagenomic data?**
- **How can I perform taxonomy analysis and visualize metaproteomics data?**
- **How can I perform functional analysis on this metaproteomics data?**

<https://training.galaxyproject.org/training-material//topics/proteomics/tutorials/metaproteomics/tutorial.html>

# Hands On Session



<https://training.galaxyproject.org/training-material//topics/proteomics/tutorials/metaproteomics/tutorial.html>

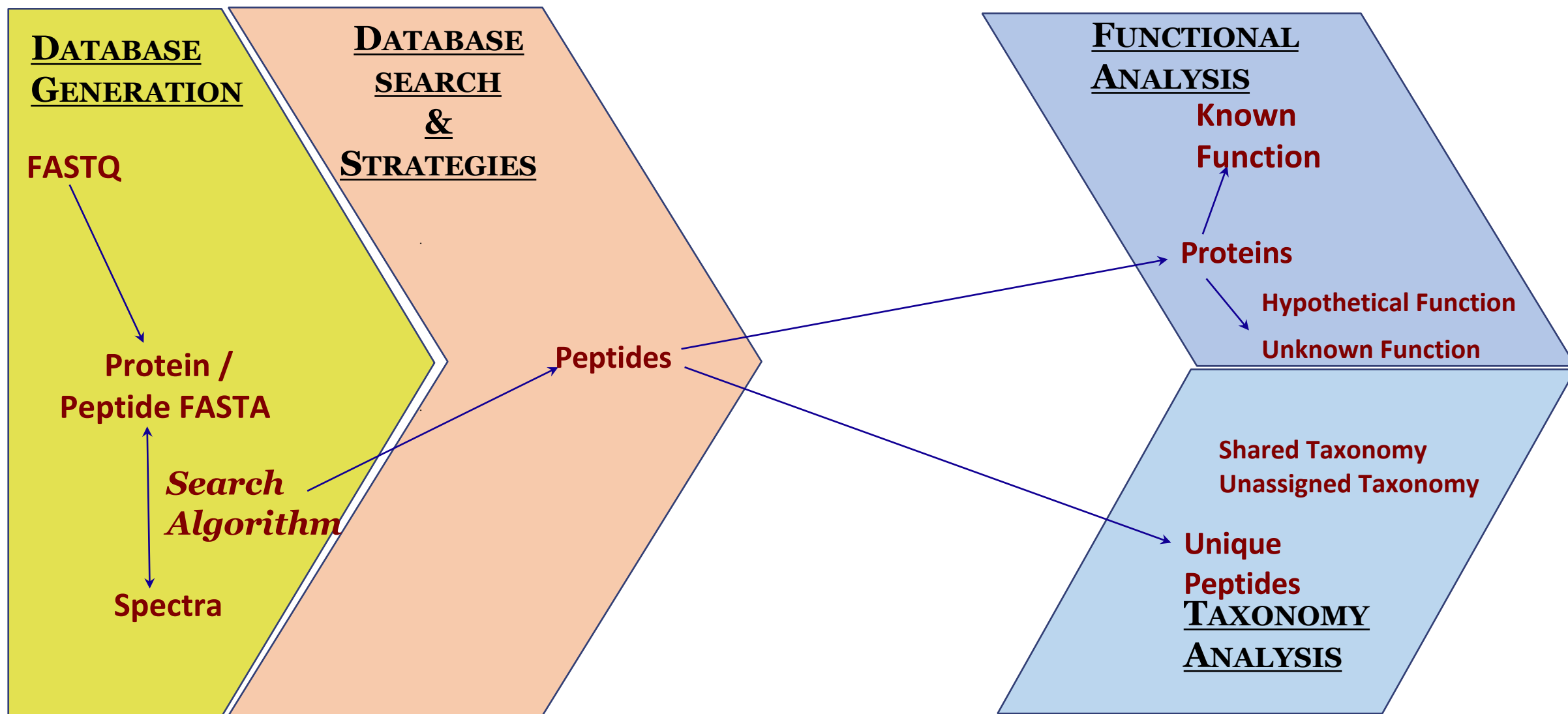


● Download and start workflow

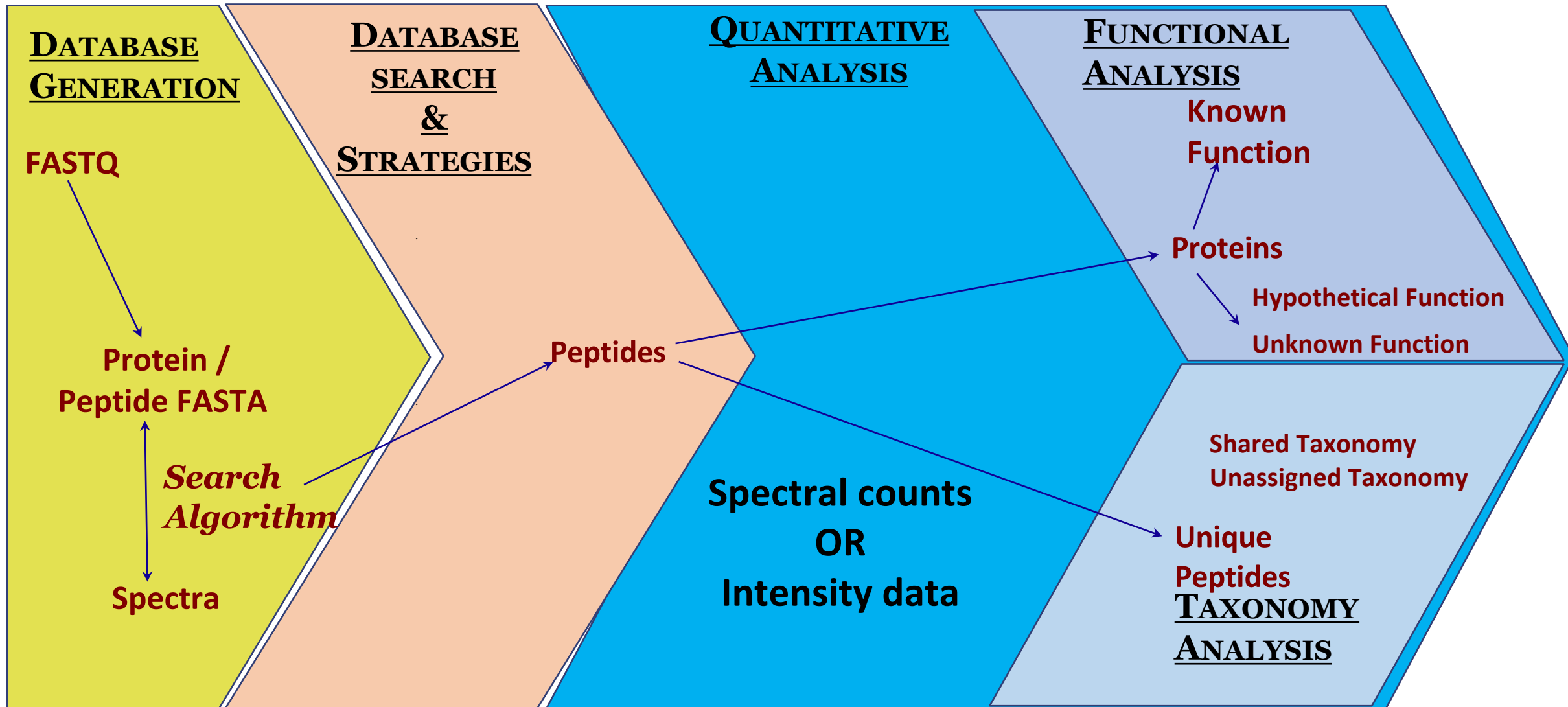
● Observe outputs

● Tool Basics

# METAPROTEOMICS WORKFLOW

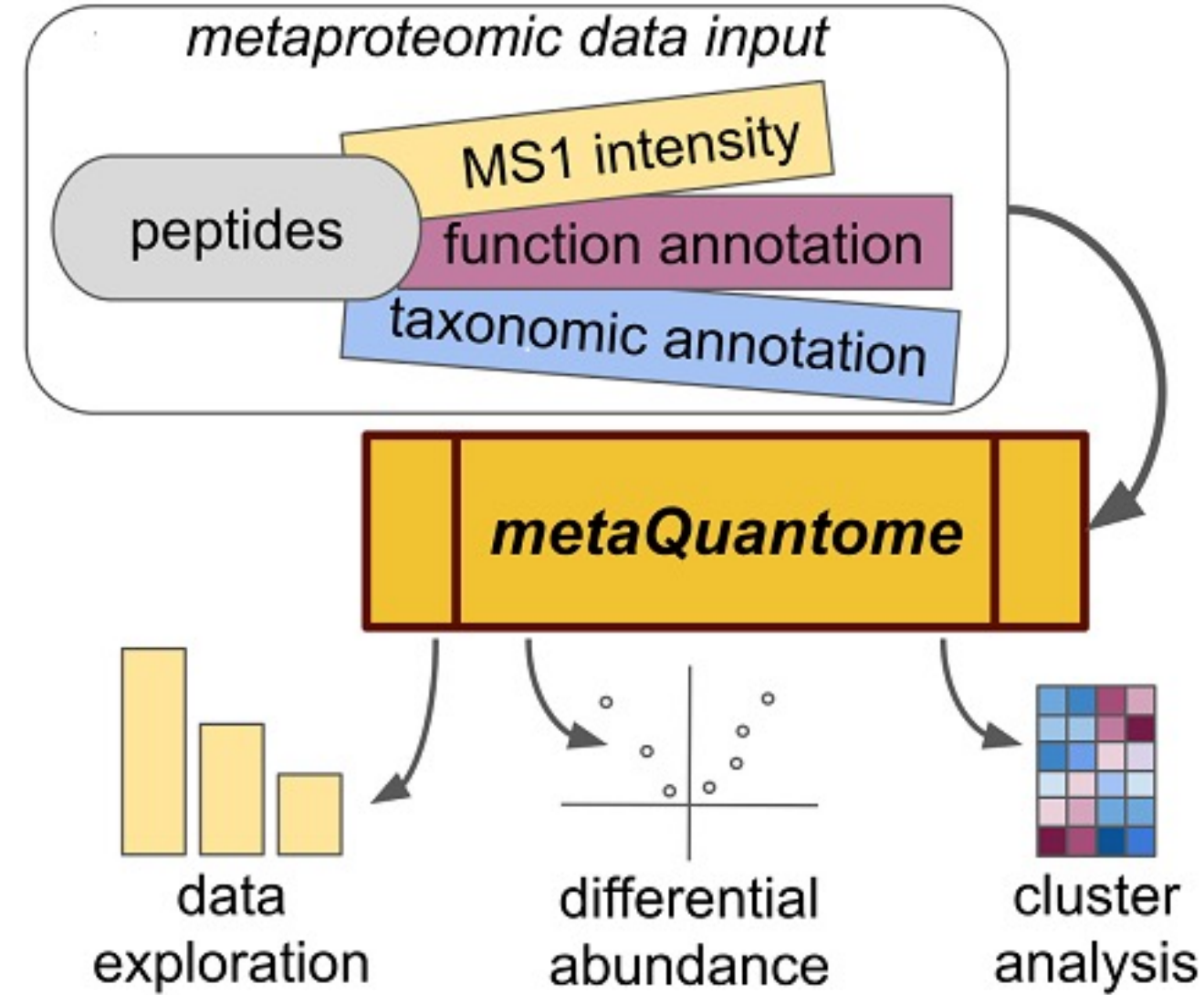


# METAPROTEOMICS WORKFLOW



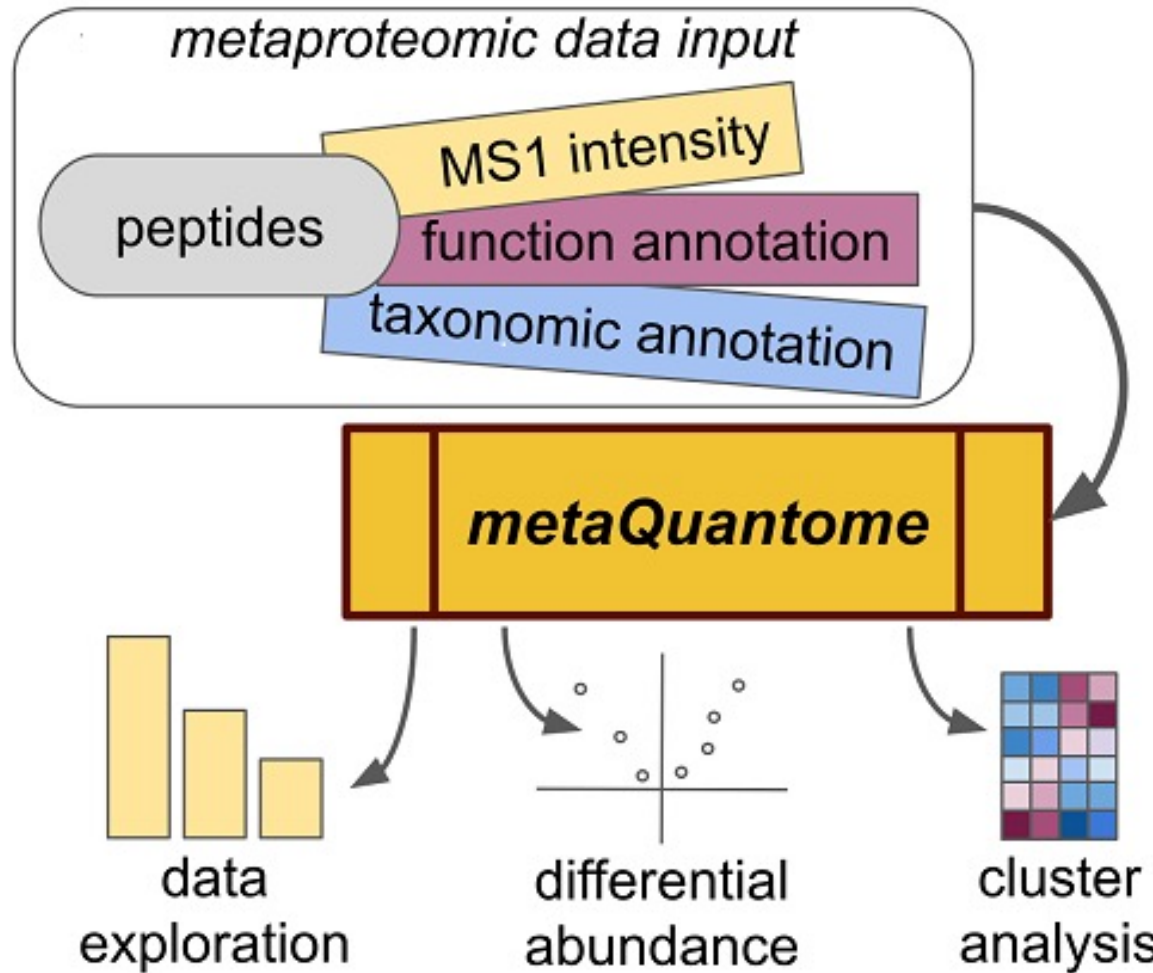
# metaQuantome

*Caleb Easterly*



**metaQuantome enables quantitative analysis of the taxonomic and functional state of a microbiome.**

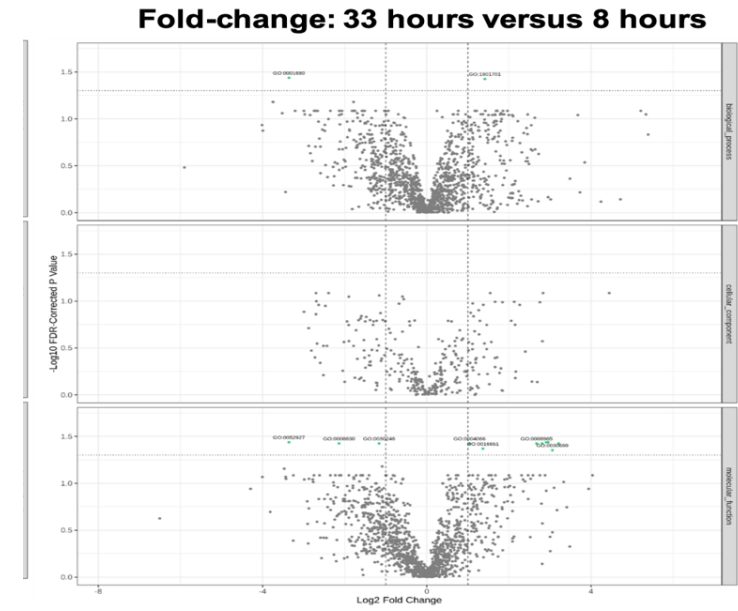
# metaQuantome



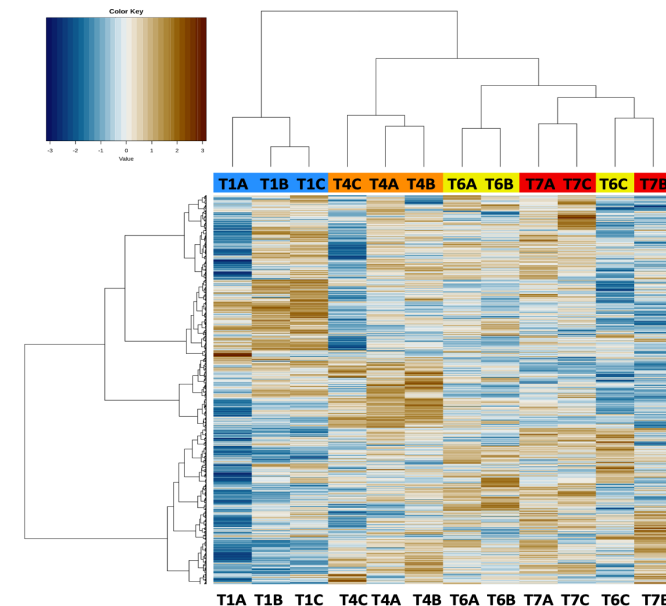
Easterly et al Mol Cell Proteomics. 2019 pii: mcp.RA118.001240. doi: 10.1074/mcp.RA118.001240.



## FUNCTION: VOLCANO PLOTS

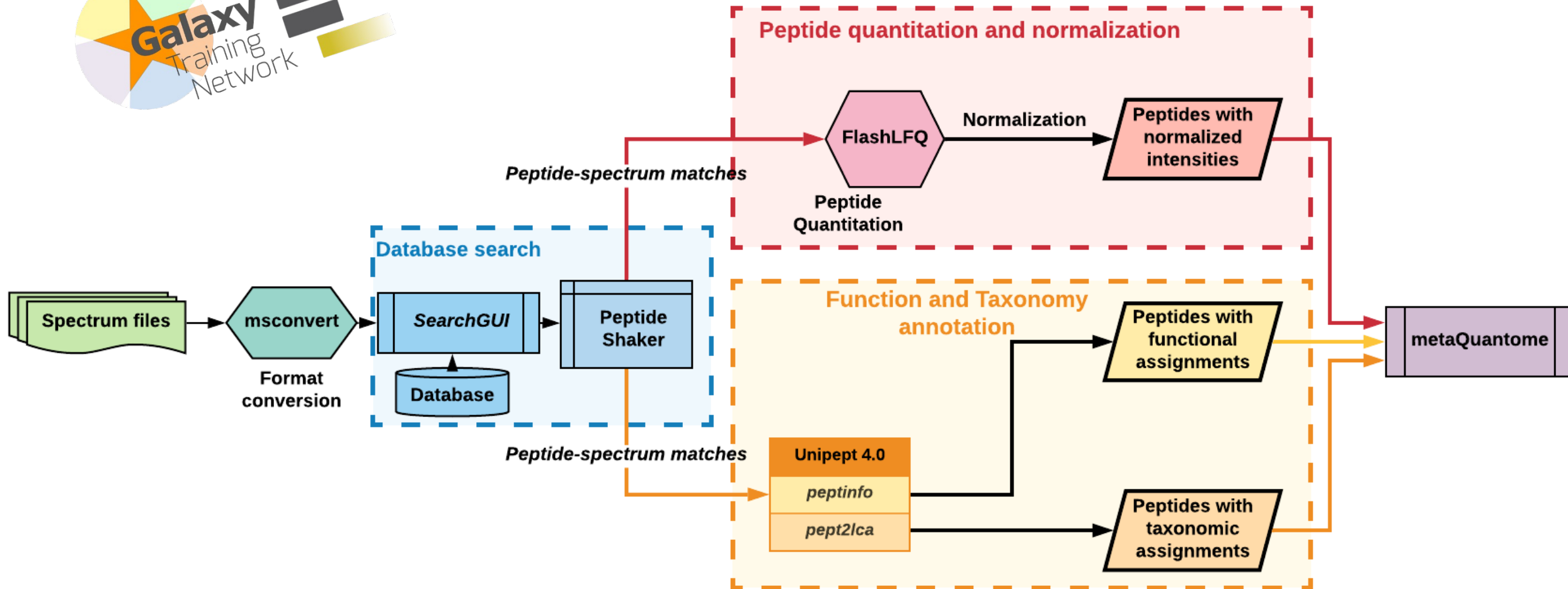


## FUNCTION: HEATMAP





# METAQUANTOME ON GALAXY TRAINING NETWORK



- <https://training.galaxyproject.org/training-material/topics/proteomics/tutorials/metaquantome-data-creation/tutorial.html>
- <https://training.galaxyproject.org/training-material/topics/proteomics/tutorials/metaquantome-function/tutorial.html>
- <https://training.galaxyproject.org/training-material/topics/proteomics/tutorials/metaquantome-taxonomy/tutorial.html>

# CASE STUDY: CELLULOSE DEGRADATION IN A BIOGAS REACTOR

Biogas-plant (60°C)  
Fredrikstad, Norway



Food waste  
Manure



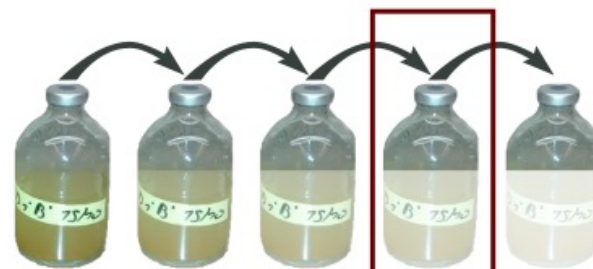
Lab-scale  
reactor (55°C)



Food waste  
Manure



Anaerobic bottles (65°C)



Serial dilution

Cellulose



- 0h
- 8h T1
- 13h
- 18h
- 23h T4
- 28h
- 33h T6
- 38h T7
- 43h



Mass  
Spectrometer

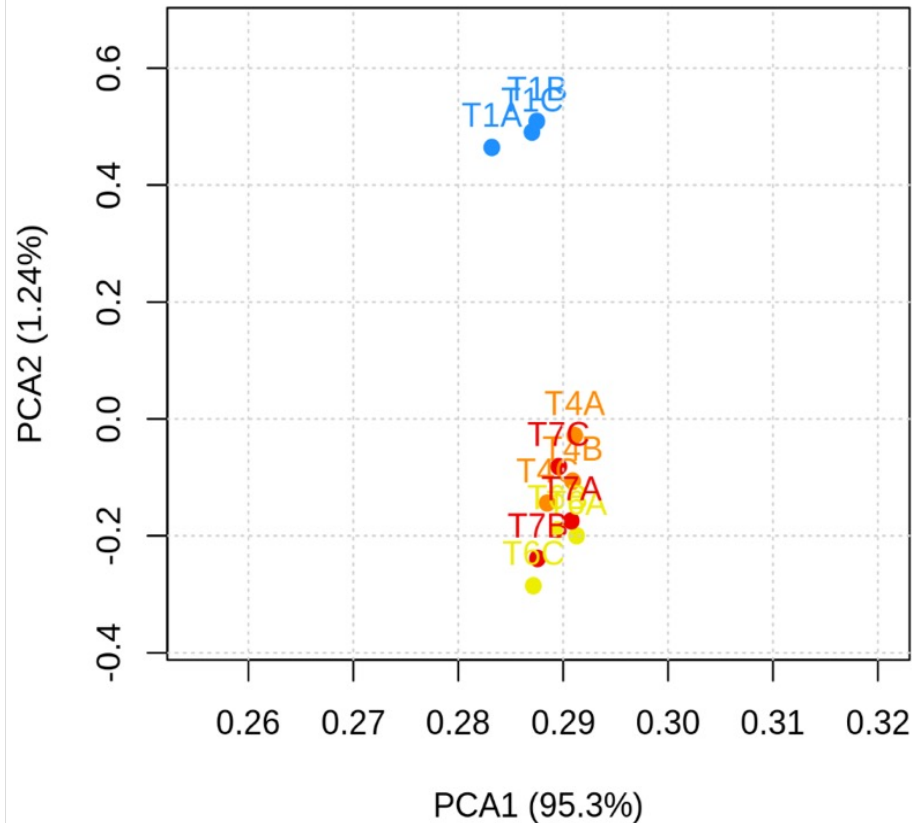


Magnus Arntzen  
NMBU, Norway

# metaQuantome ANALYSIS: PRINCIPAL COMPONENT ANALYSIS

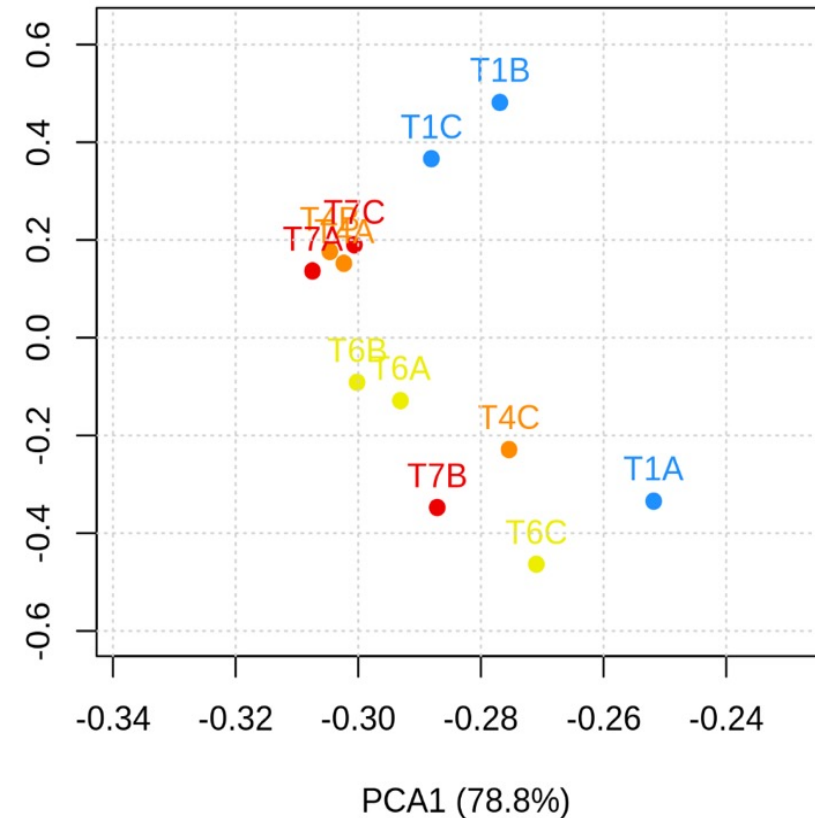
## FUNCTION: PCA PLOT

Cluster Separation: 0.93352



## TAXONOMY: PCA PLOT

Cluster Separation: 0.19056

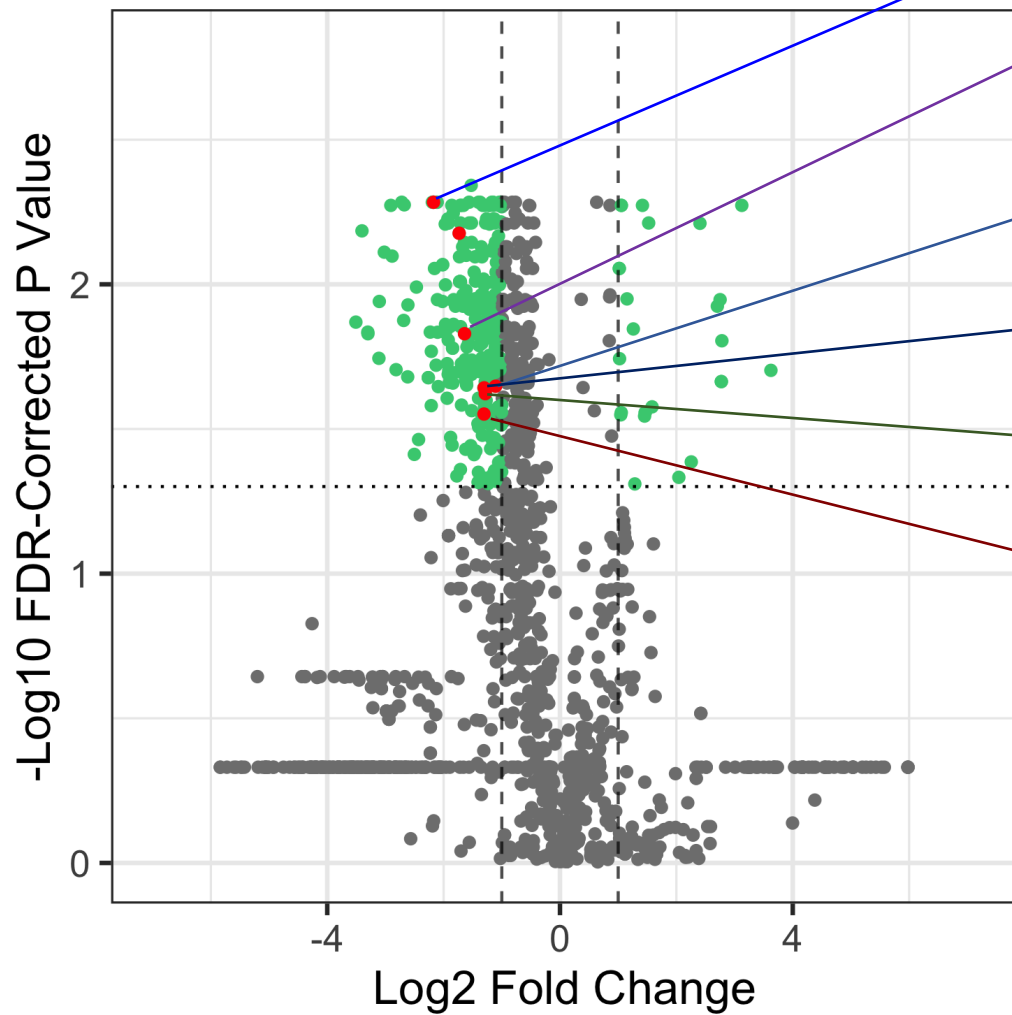


**Functional abundance values separate time point T1 (8 hr) from other time points thus highlighting the importance of understanding functional state of the microbiome.**

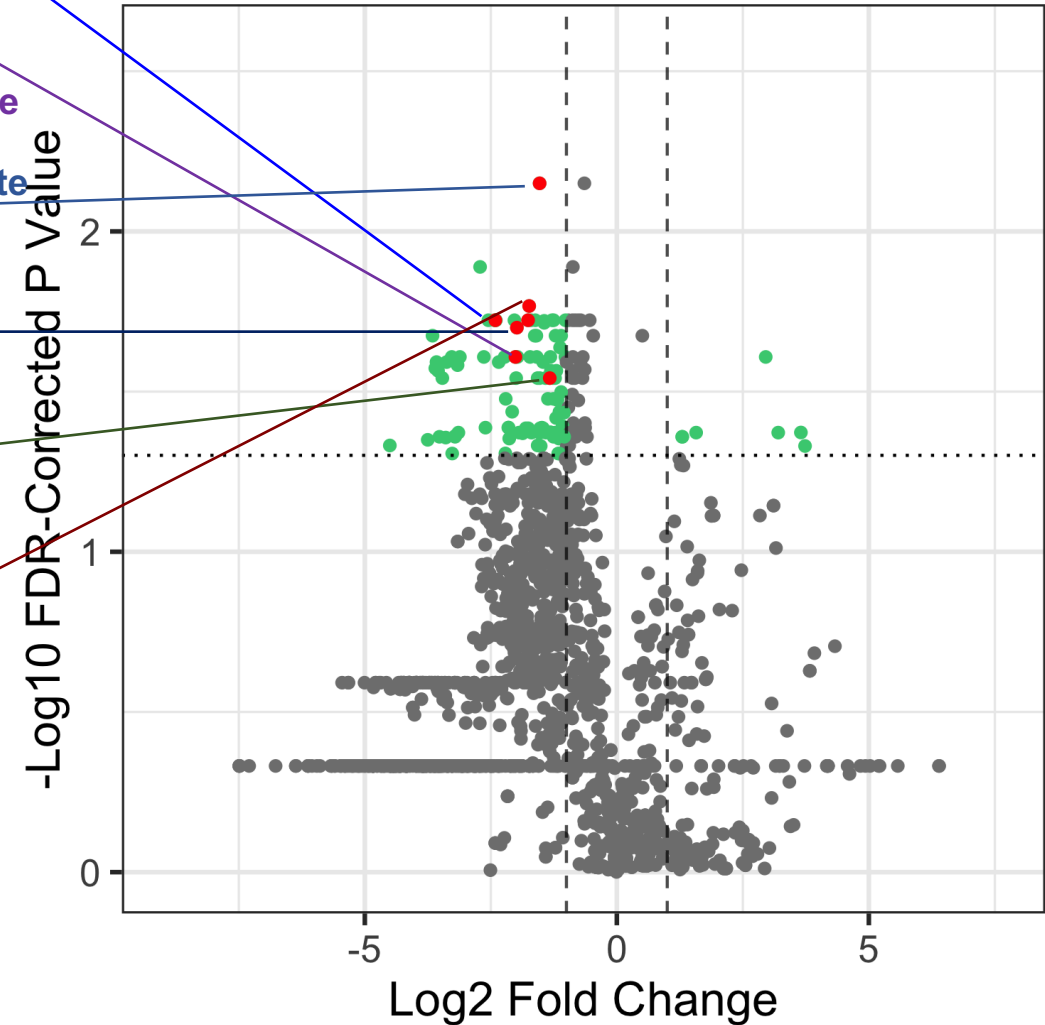
# metaQuantome Analysis

# VOLCANO PLOTS

T6 over T4



T7 over T4



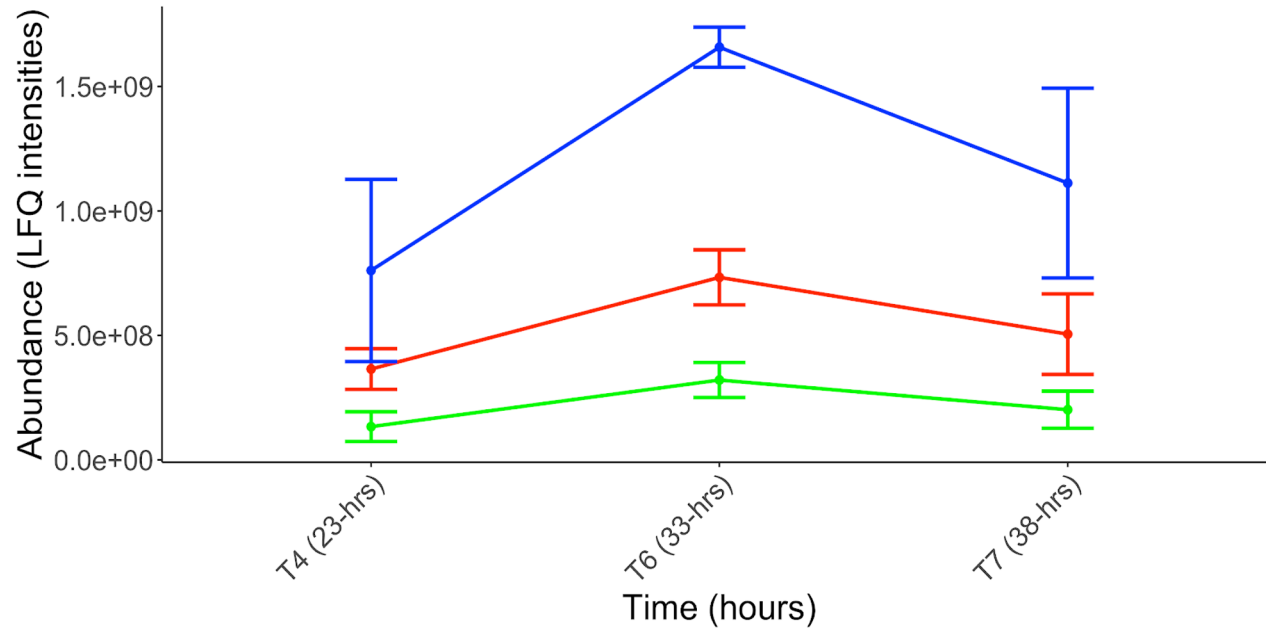
- Glucosyl transferase
- Cellulose 1,4-beta-cellobiosidase
- Carbohydrate binding
- Cellulose binding
- Amino peptidase
- Cellulase

**Gene Ontology terms were found to be differentially expressed in both timepoints T6 and T7 as compared to T4.**

# metaQuantome Analysis

## FUNCTIONS EXPRESSED BY A TAXON

*Hungateiclostridium*



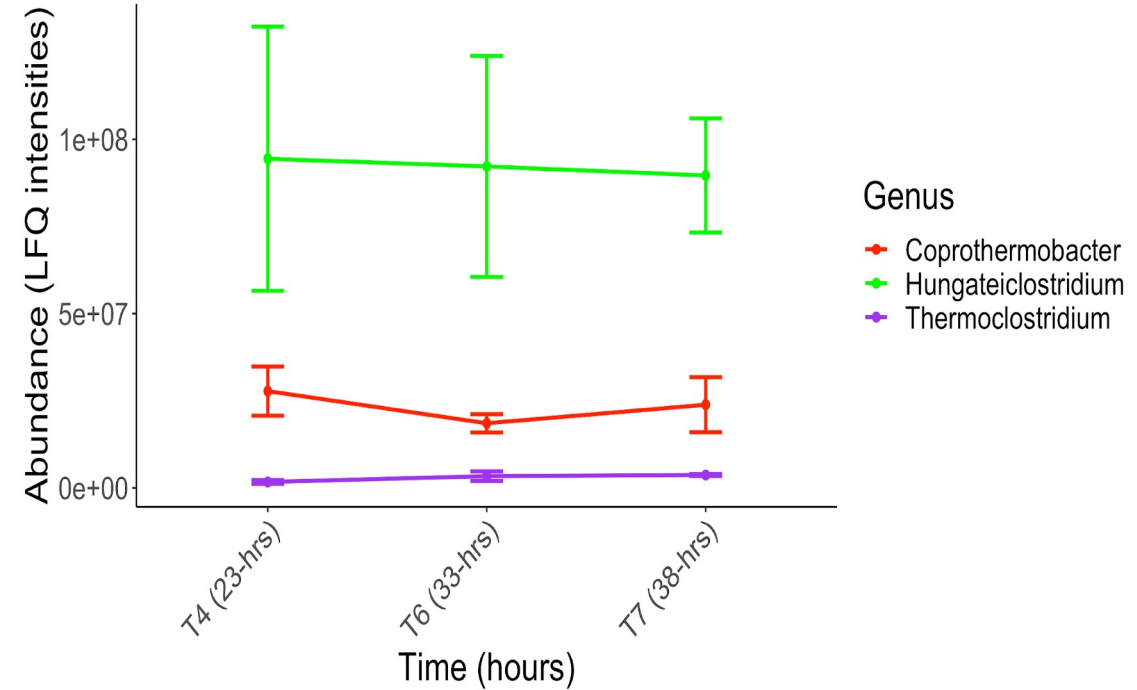
### Gene Ontology Terms

- Cellulose binding
- Cellulase activity
- Cellulose 1,4-beta-cellobiosidase activity

**Functions Associated With Cellulose Degradation in *Hungateiclostridium***

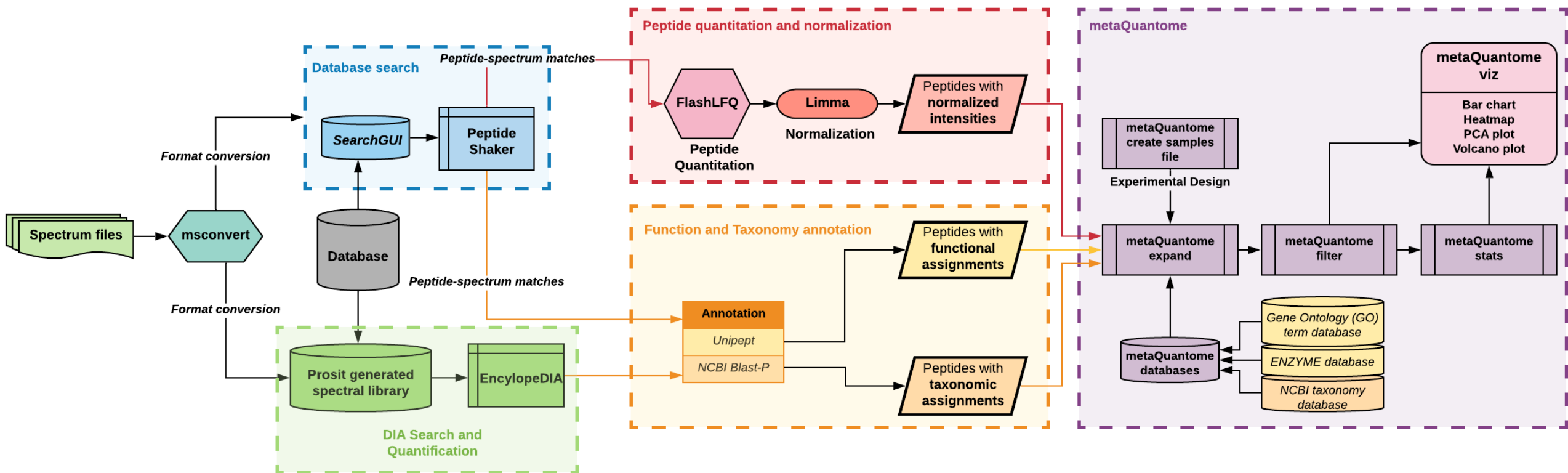
## TAXONOMIC CONTRIBUTION TO A FUNCTION

### GLYCOSIDE HYDROLASE



**Taxa associated with Glycosyl hydrolases and transferases**

# METAQUANTOME WORKFLOW



# Metaproteomics Publications

PROTEOMICS

Jagtap P *et al* (2012) Deep metaproteomic analysis of **human salivary supernatant**. *Proteomics*. 12(7):992-1001. doi: 10.1002/ pmic.201100503.

Microbiome

Rudney JD *et al* (2015) Protein relative abundance patterns associated with **sucrose-induced dysbiosis** are conserved across taxonomically diverse oral microcosm **biofilm models of dental caries**. *Microbiome*. 3:69. doi: 10.1186/s40168-015-0136-z.

SCIENTIFIC  
REPORTS  
nature research

Afiuni-Zadeh S *et al* (2018) Evaluating the potential of residual Pap test fluid as a resource for the metaproteomic analysis of the **cervical-vaginal microbiome**. *Sci Rep.* ;8(1):10868. doi: 10.1038/s41598-018-29092-4.



ATS Journals

AJRCMB

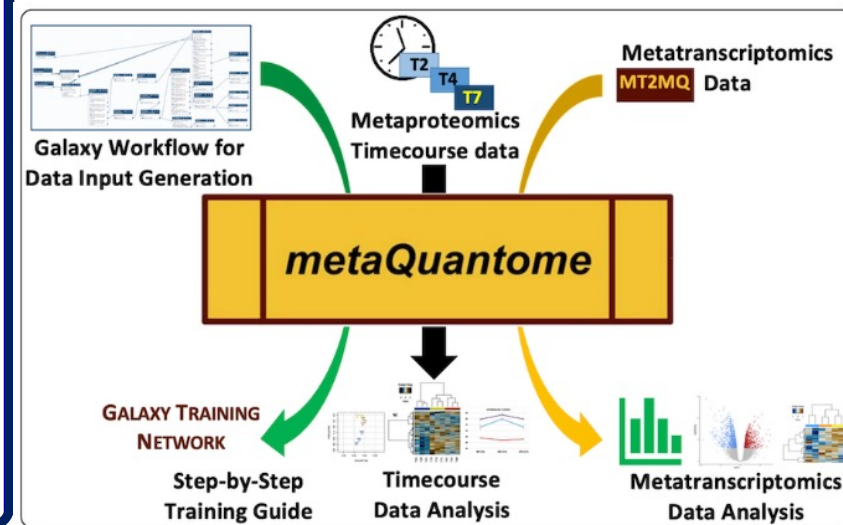
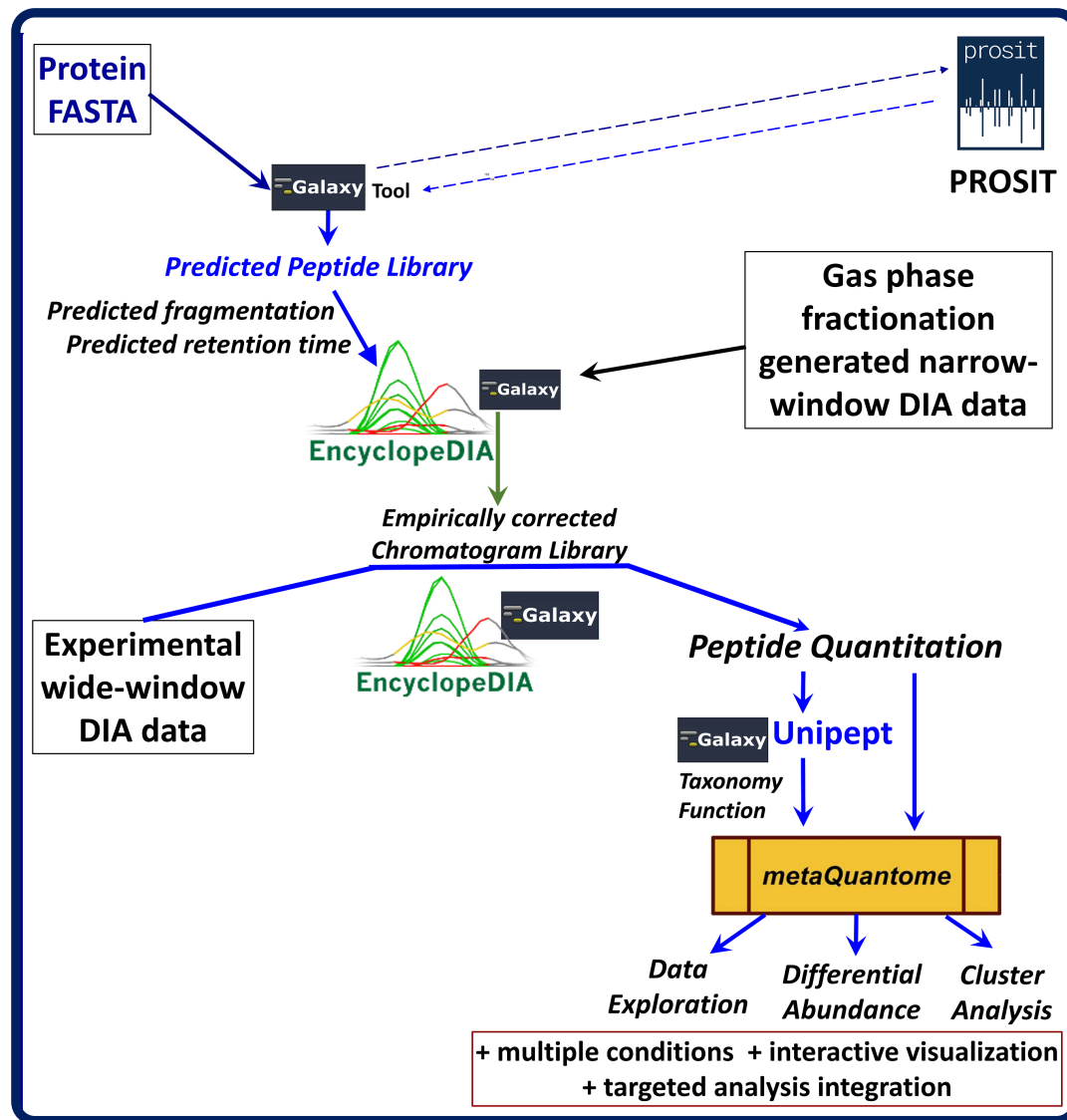
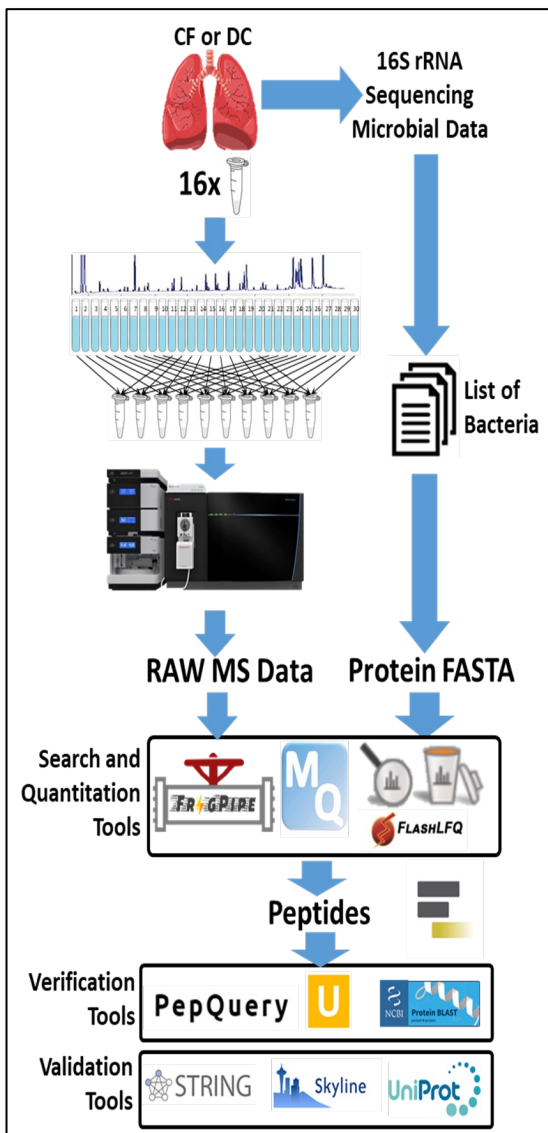
Jagtap PD *et al* (2018) **BAL Fluid Metaproteome** in Acute Respiratory Failure. *Am J Respir Cell Mol Biol*. 59(5):648-652. doi: 10.1165/rcmb.2018-0068LE.



Journal of  
proteome  
research

Thuy-Boun PS *et al* (2021). Metaproteomics Analysis of **SARS-CoV-2-Infected Patient Samples** Reveals Presence of Potential Coinfecting Microorganisms. *J Proteome Res*. 20(2):1451-1454. doi: 10.1021/acs.jproteome.0c00822

# WHAT'S ON THE HORIZON?



CHARACTERIZATION OF HOST & MICROBIAL PROTEINS FROM CLINICAL SAMPLES

QUANTITATION

MULTIOMICS



# A METAPROTEOMICS BIOINFORMATICS WORKFLOW TO STUDY HOST- MICROBE DYNAMICS IN CLINICAL SAMPLES

*Pratik Jagtap*

Galaxy for Proteomics (Galaxy-P) team

*University of Minnesota* 

*Co-authors:*

*Monica E. Kruk, Subina Mehta, Katherine Do, James E. Johnson, Reid Wagner, Chris H. Wendt, John B. O'Connor, Theresa Laguna and Timothy J. Griffin*



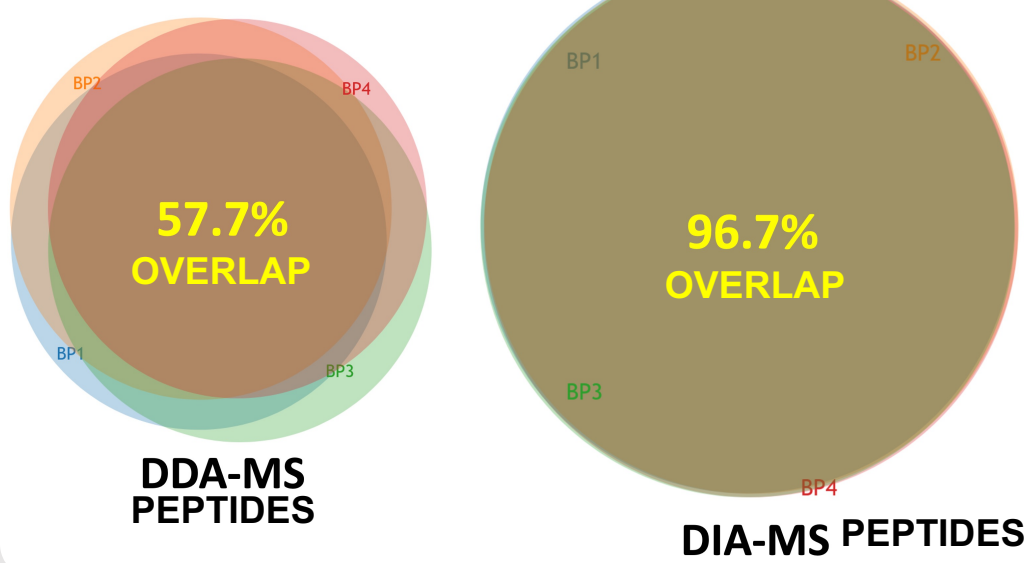
**IMSC 2022**  
**MAASTRICHT, THE NETHERLANDS**

27th August - 2nd September

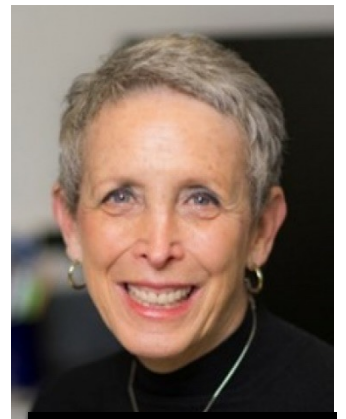
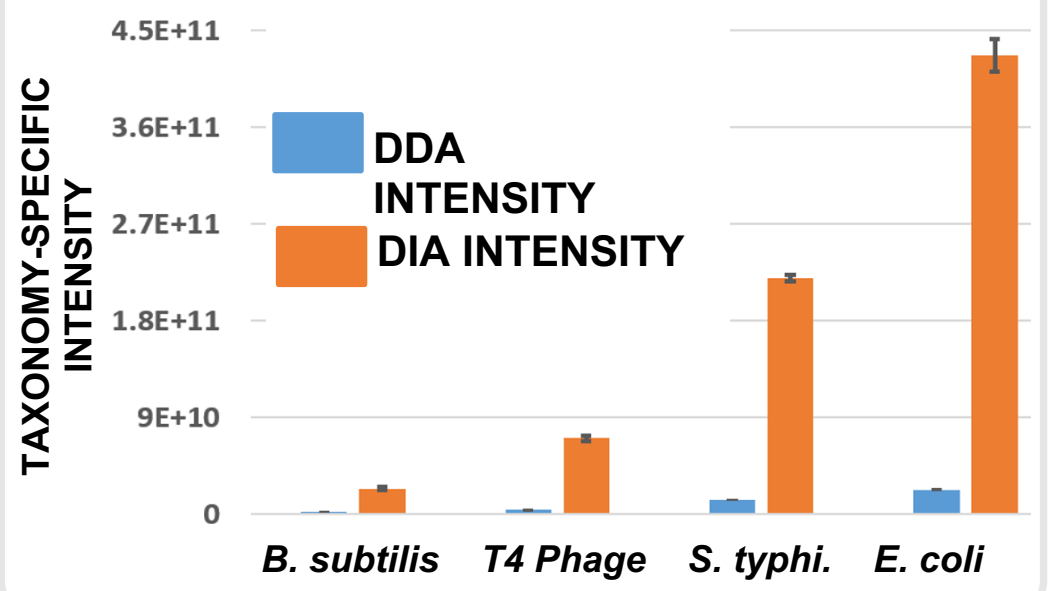


# ASMS 2021: DIA-MS OFFERS BENEFITS FOR QUANTITATIVE METAPROTEOMICS

## DIA-MS HAS BETTER REPRODUCIBILITY ACROSS REPLICATES



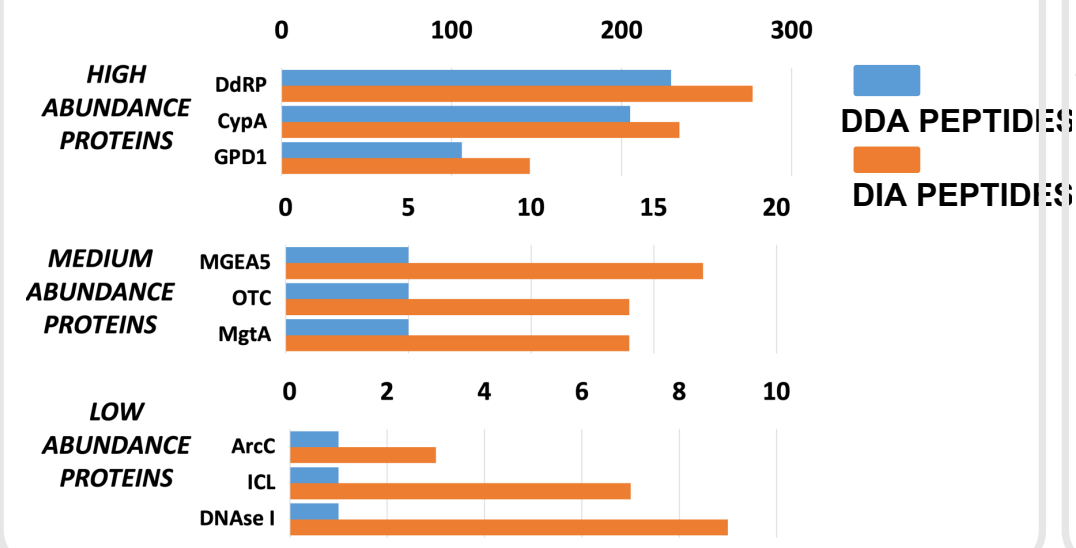
## DIA-MS OFFERS DEEPER TAXONOMY COVERAGE



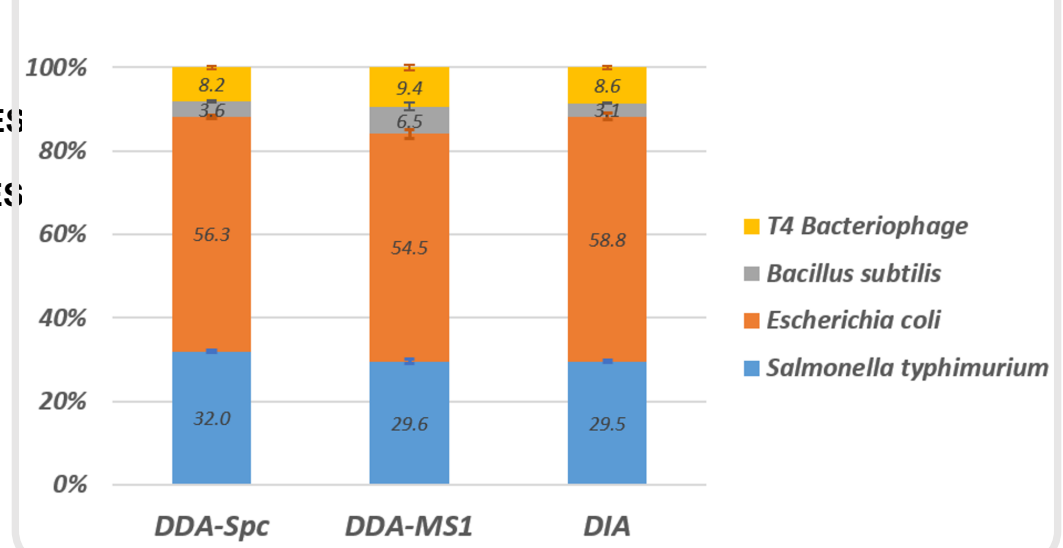
**Susan T. Weintraub**



## DIA-MS OFFERS DEEPER FUNCTIONAL COVERAGE



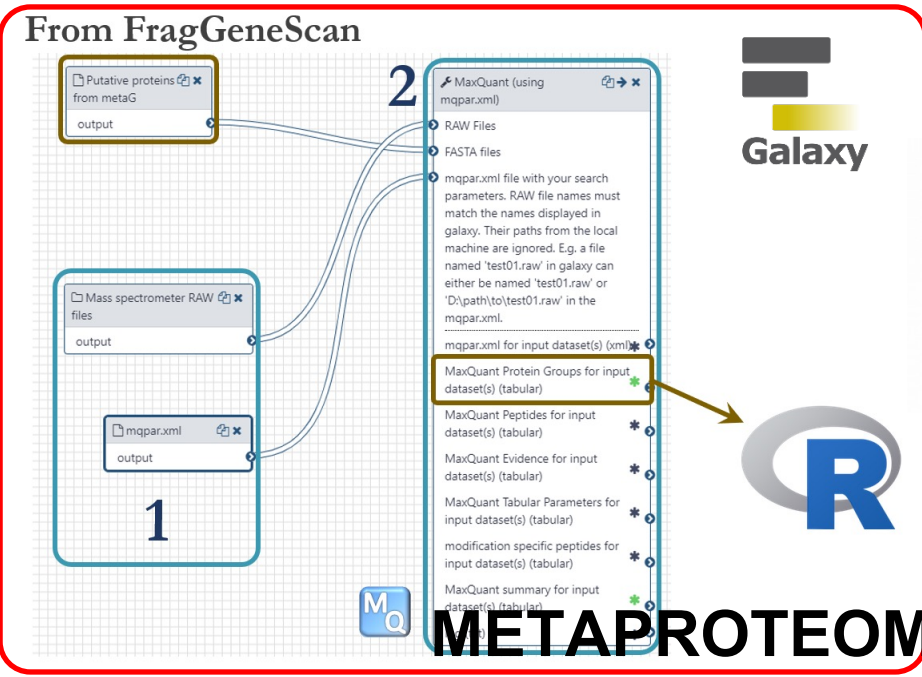
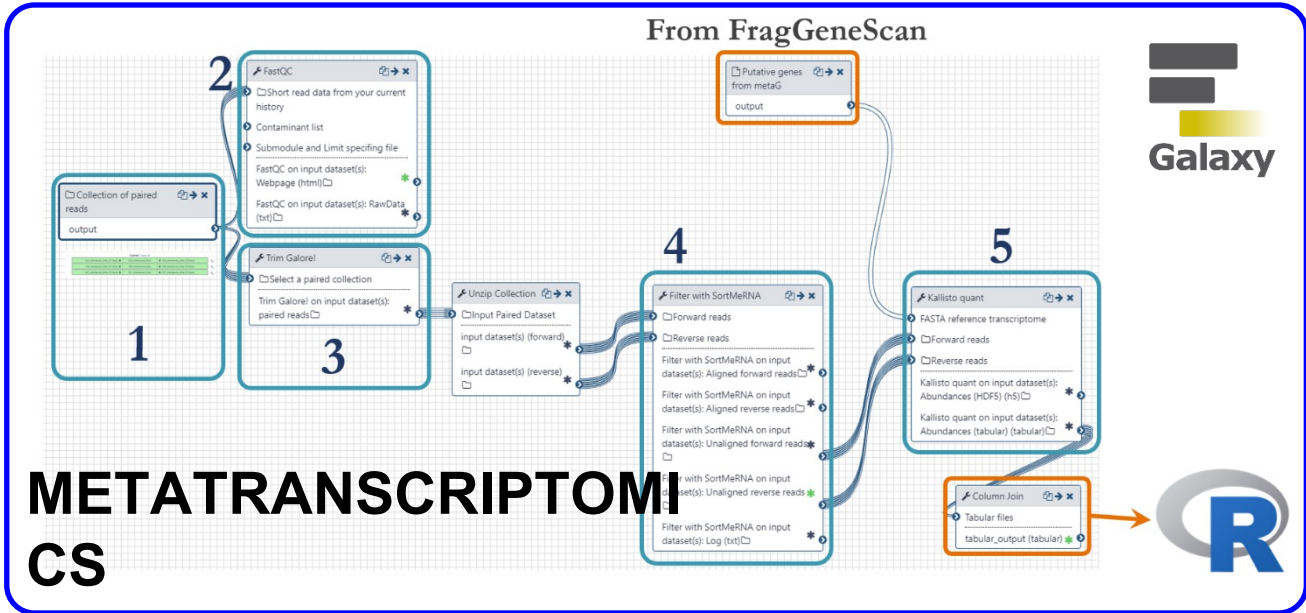
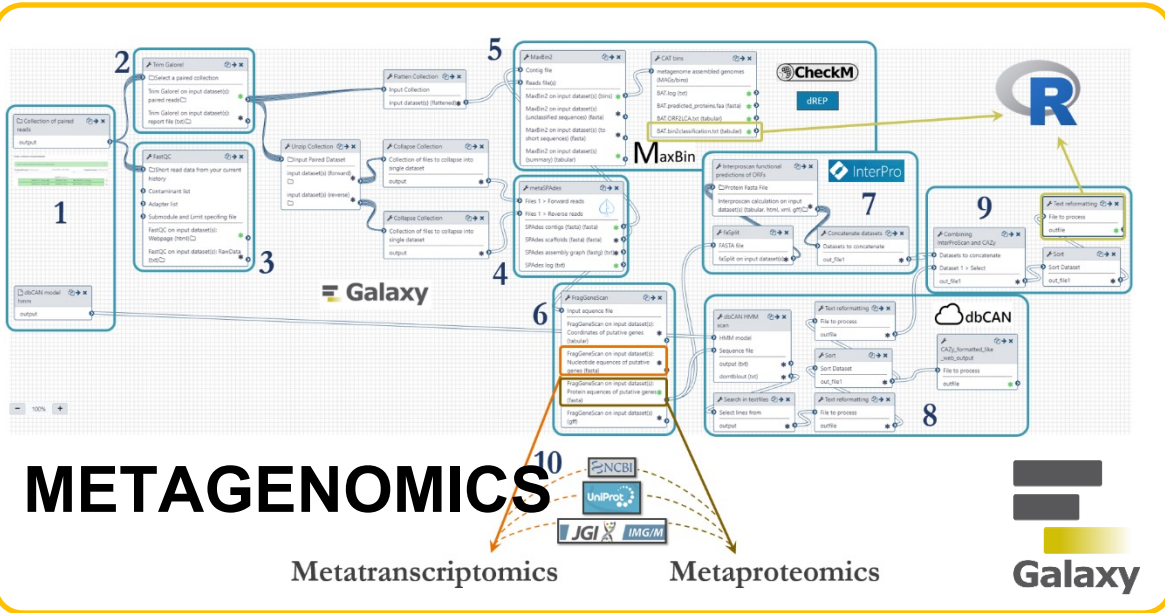
## WHAT LEVEL OF ACCURACY DOES DIA-MS OFFER?



**Brook L. Nunn**



# META-OMICS APPROACH BY NMBU TEAM



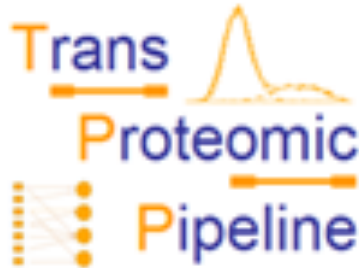
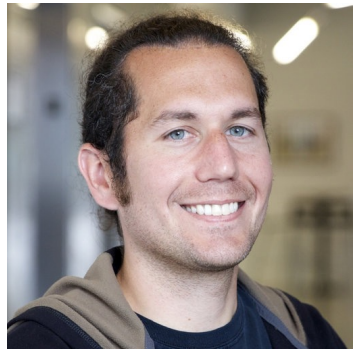
Francesco Delogu Magnus Arntzen

<https://galaxyproject.eu/posts/2020/04/14/integrative-meta-omics/>

# Cloud Computing Workshop (2022)

The iPRG will conduct a series of online video tutorials about the use of cloud computing resources for MS-based proteomics, focusing on Nextflow, the Trans-Proteomic Pipeline (TPP) and Galaxy Platform.

## September 2022



Michael Hoopmann - *ISB, Seattle, WA*

- Instructions on how to use TPP to analyze MS data.
- Answer questions from the participants

## October 2022



Melanie Foell - *Freiburg University, Germany*

- Instructions on how to use Galaxy to analyze MS data.
- Answer questions from the participants

## November 2022



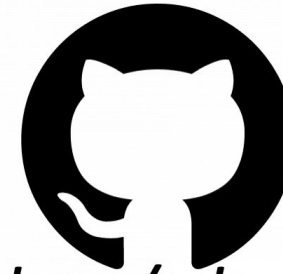
Yasset Perez-Riverol - *EBI, Hinxton, UK*

- Instructions on how to use Nextflow to analyze MS data.
- Answer questions from the participants

# TEN YEARS OF GALAXY-P !



[z.umn.edu/galaxypreferences](https://z.umn.edu/galaxypreferences)



<https://github.com/galaxyproteomics>



<https://training.galaxyproject.org>



<http://galaxy.org/workshops/>



# Accessing tools and Workflows

## METAGENOMICS:

Toolshed: [z.umn.edu/metagenomics\\_toolshed](http://z.umn.edu/metagenomics_toolshed)

Galaxy Training Network: <https://training.galaxyproject.org/training-material/topics/metagenomics/>



## METATRANSCRIPTOMICS:

Workflow: <http://z.umn.edu/MTWF2020>

Galaxy Training Network: <http://bit.ly/gtn-metatranscriptomics>



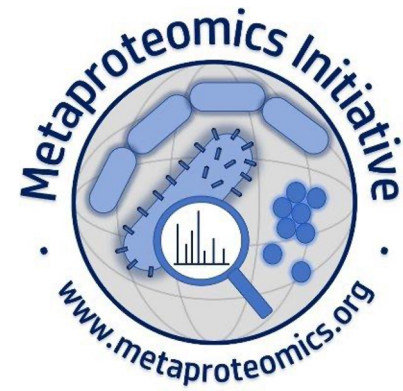
## METAPROTEOMICS:

Workflow: [z.umn.edu/MPWF2020](http://z.umn.edu/MPWF2020)

Galaxy Training Network: <http://z.umn.edu/gtn-metaproteomics>

Also available on : <https://proteomics.usegalaxy.eu/> and  
Metaproteomics Gateway: [z.umn.edu/metaproteomicsgateway](http://z.umn.edu/metaproteomicsgateway)

[galaxyp.org/contact](http://galaxyp.org/contact)



<https://gallantries.github.io/video-library/videos/proteomics/metaproteomics/tutorial/>

# ACKNOWLEDGMENTS

University of Minnesota

**Timothy Griffin**  
Subina Mehta  
Monica Kruk  
Katherine Do  
Danielle Weise  
Andrew Rajczewski



Minnesota Supercomputing Institute

**James Johnson**  
**Reid Wagner**

Brian Searle  
*Ohio State University*

Harald Barsnes  
Marc Vaudel  
*University of Bergen, Norway*

Phil Pope  
Magnus Øverlie Arntzen  
*NMBU, Oslo, Norway*

Brook Nunn  
*U of Washington*

**Andrew McCardle**  
*Imperial College, London, UK*

Saskia Hiltmann

Thilo Muth  
*Robert Koch Institut, Germany*

Biologists / collaborators

Chris Wendt  
Teresa Laguna  
Maneesh Bhargava  
Ryan Hunter  
David Largaespada

Peter Thuy-Boun  
Dennis Wolan  
*Scripps Institute*

Lennart Martens  
Bart Mesuere  
*UGhent, Belgium*

Sanjeeva Srivastava  
*IIT, Mumbai*

Lloyd Smith (Co-I)  
Michael Shortreed  
*UW-Madison*

Alessandro Tanca  
*Porto Conte Ricerche, Italy*

Anamika Krishanpal  
*Persistent Systems Limited India*

Jeremy Fisher  
*Indiana University*

Valdemir Carvalho  
*Fleury Group, Brazil*

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

Matt Chambers  
*Nashville, TN*

Funding



**NATIONAL CANCER INSTITUTE**  
Informatics Technology for  
Cancer Research



Maria Doyle  
*Melbourne, Australia*

[galaxyp.org](http://galaxyp.org)

[twitter.com/usegalaxy](https://twitter.com/usegalaxy)