Introduction to Metaproteomics

https://gcc2022.sched.com/event/xaCl/introduction-to-metaproteomics

Galaxy Community Conference

July 17-20, 2022

GCC2022 MINNEAPOLIS

COME TOGETHER #UseGalaxy2022

Minnesota Supercomputing Institute

University of Minnesota

Driven to Discover®

GalaxyP

Galaxy for Proteomics Project

Learn more at galaxyp.org

z.umn.edu/itcrgalaxyvideo

Workshop instructors & acknowledgements

- **Instructors**
  - Tim Griffin
  - Pratik Jagtap
  - Subina Mehta
  - James Johnson

- **Other contributors**
  - Andrew Rajczewski
  - Reid Wagner
  - Katherine Do
  - usegalaxy.eu team
  - Galaxy community

https://gcc2022.sched.com/event/xaCl/introduction-to-metaproteomics

Agenda and schedule

1:20 PM – 1:30 PM: Introduction to Mass Spectrometry

1:30 PM - 1:50 PM: Multi-omics and Galaxy Workflows for Metaproteomics Research

1:50 PM - 3:10 PM: Taxonomy and functions of microbiome from environmental samples

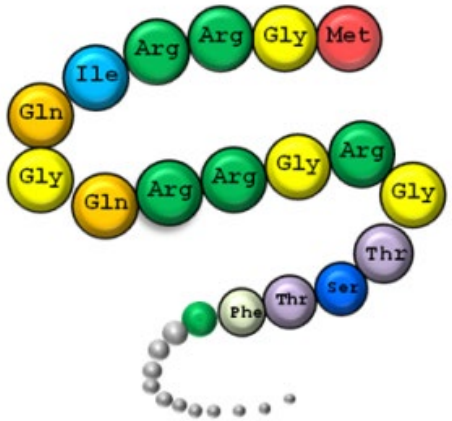
3:10 PM - 3:50 PM: Summary and Questions and Answers

Please ask questions as they come up!

Workshop Topics

Mass spectrometry-based proteomic informatics

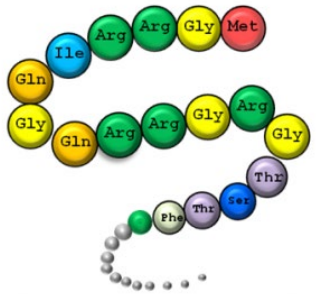
Metaproteomics (multi-omics)



Protein sequence analysis

GalaxyP
galaxyp.org

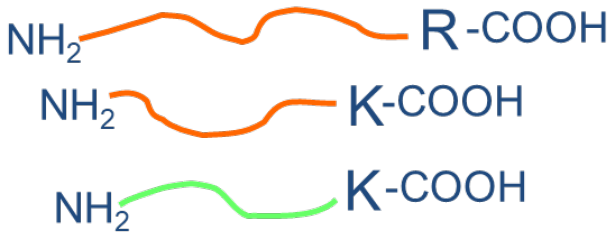
Shotgun proteomics: breaking proteins into peptides



1. Protein, MW = 10,000 +

digest into peptides (enzymatic, chemical)

2. Peptides, MW < 4,000



Trypsin

N-TH**K**.NCPHIVVGTPG**R**.IPD-C

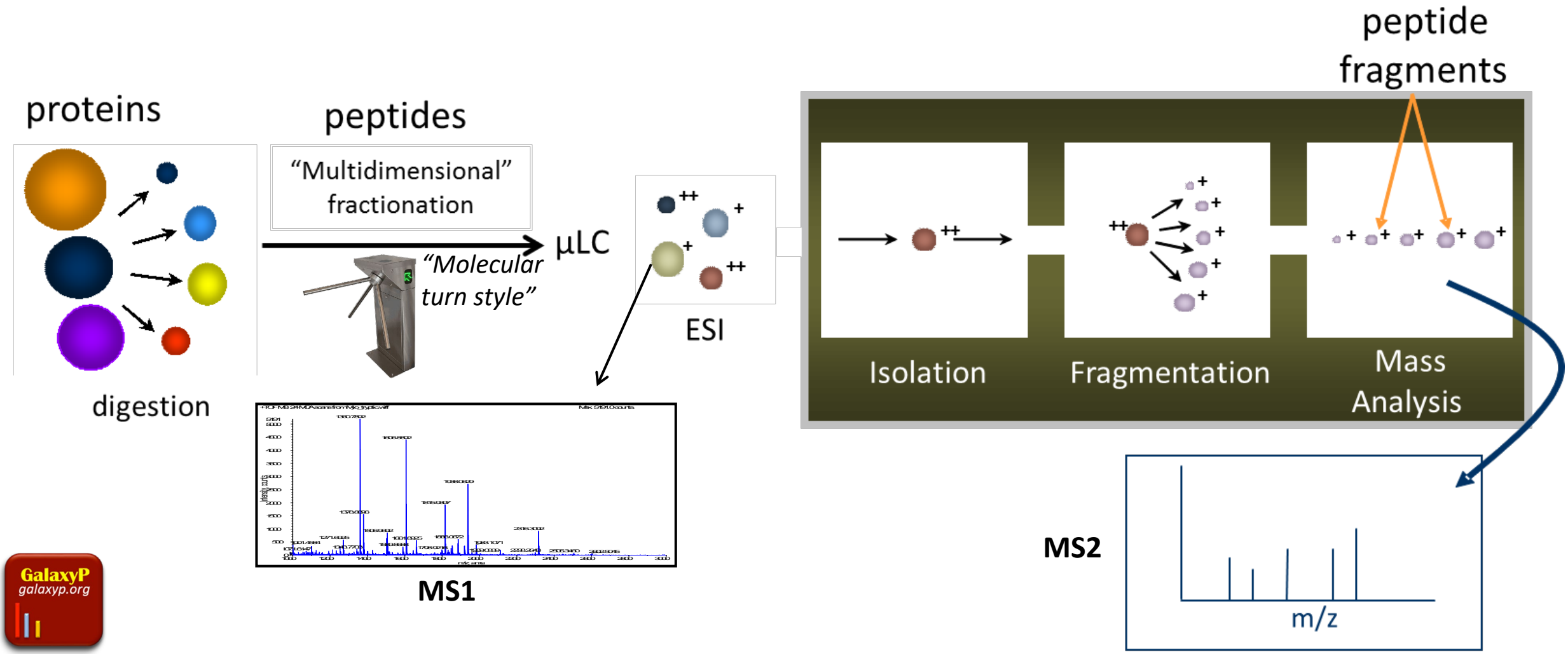
cleaves c-teminal side of arginine (R) and lysine (K)

3. Peptide fragments

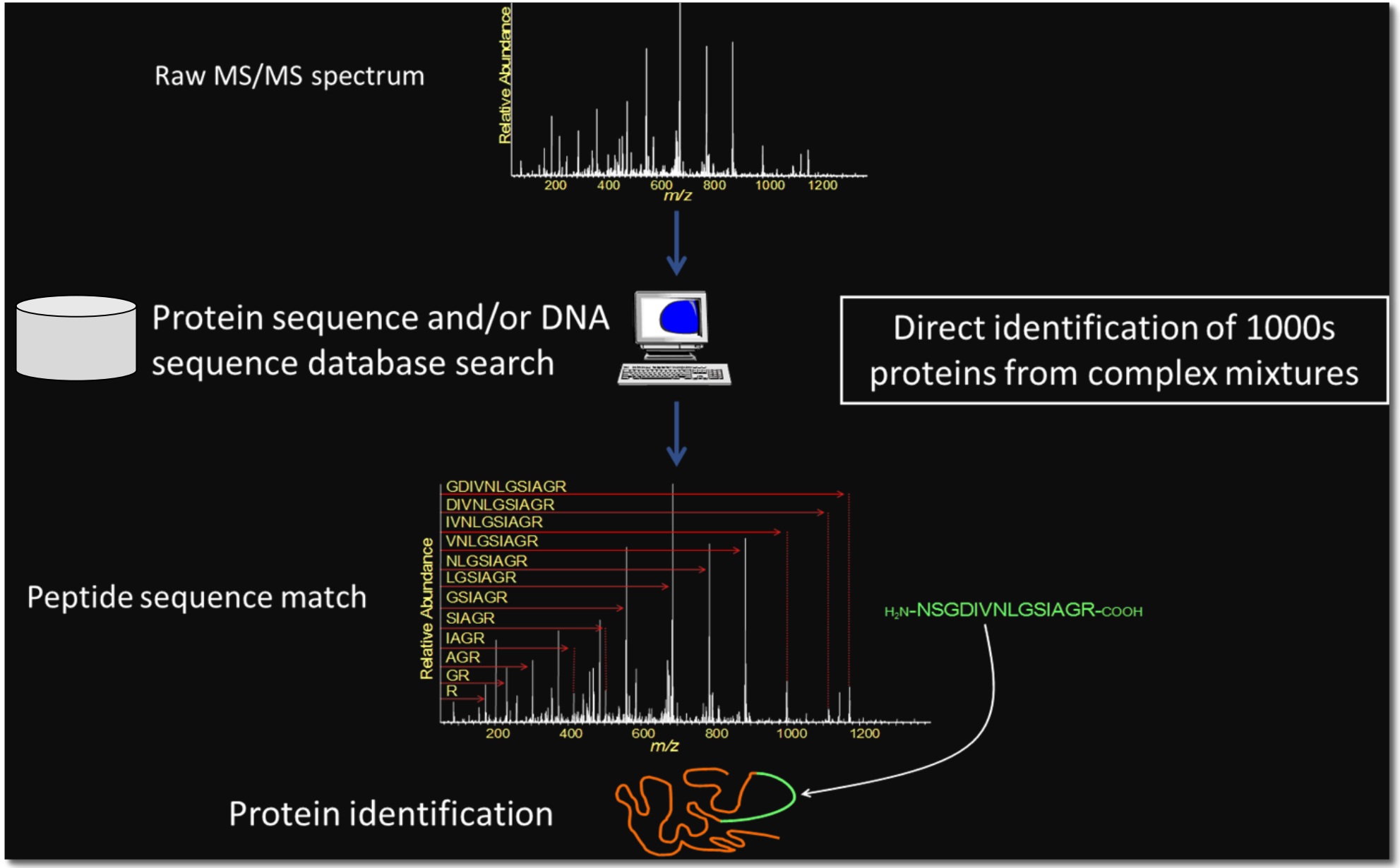
GalaxyP
galaxyp.org

Mass Spectrometry-based proteomics: A primer

Peptide fractionation coupled to tandem mass spectrometry (MS/MS)

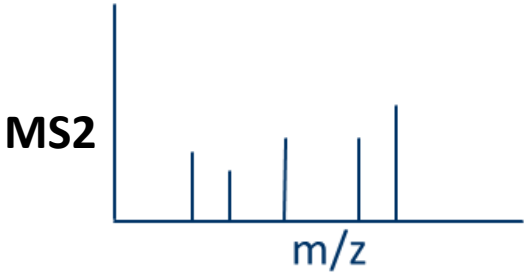


Matching amino acid sequences to MS/MS data



GalaxyP
galaxyp.org

Inferring protein identity for peptide sequences

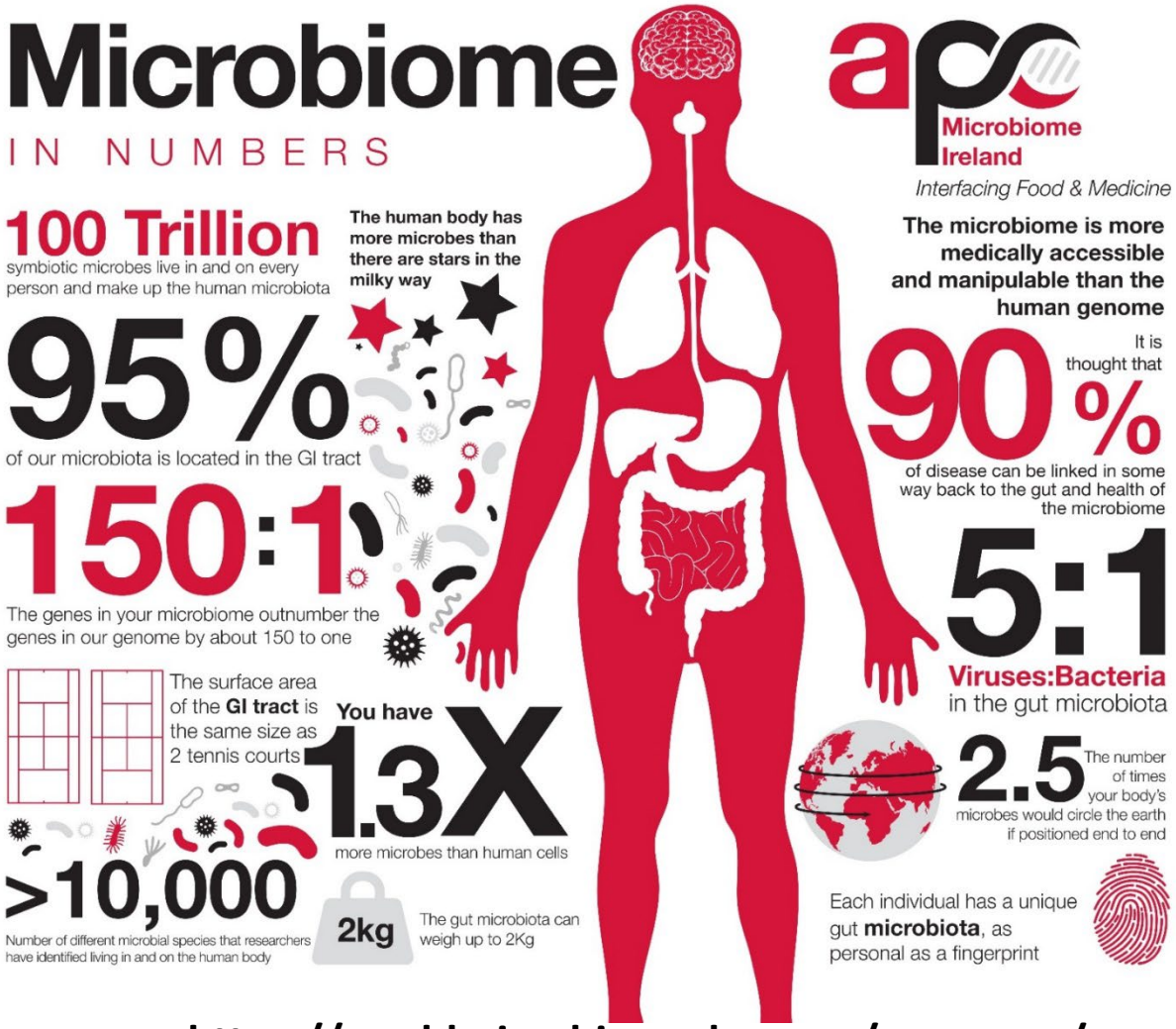


CAQCHTVEK

Cytochrome C

NH_2GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGFTYTDANKNKGITW
KEETLMEYLENPKKYIPGTKMIFAGIKKKTEREDLIAYLKKATNECOOH

MICROBIOME RESEARCH



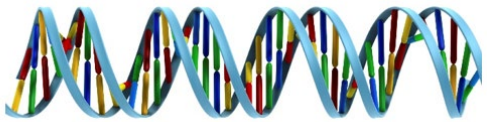
https://worldmicrobiomeday.com/resources/

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

MICROBIOME

https://www.biospectrumasia.com

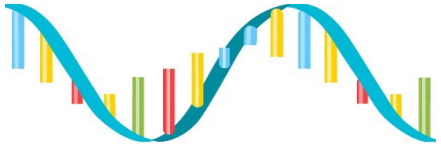
Potential to unravel the **mechanistic details of microbial interactions with host / environment.**



METAGENOMICS

TAXONOMY

function



METATRANSCRIPTOMICS

TAXONOMY

function

METAPROTEOMICS

TAXONOMY

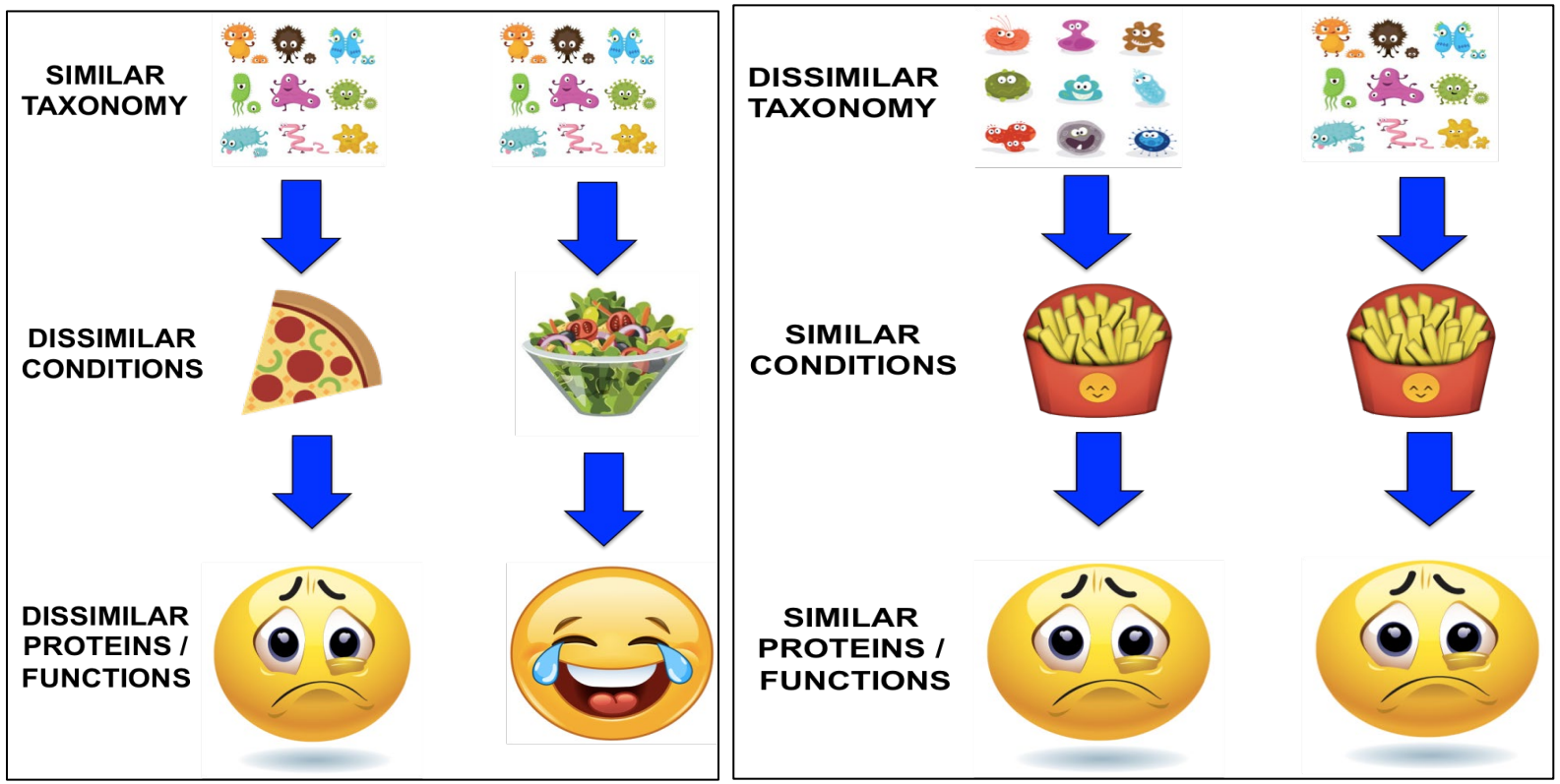
FUNCTION

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.

Metaproteomic Challenges

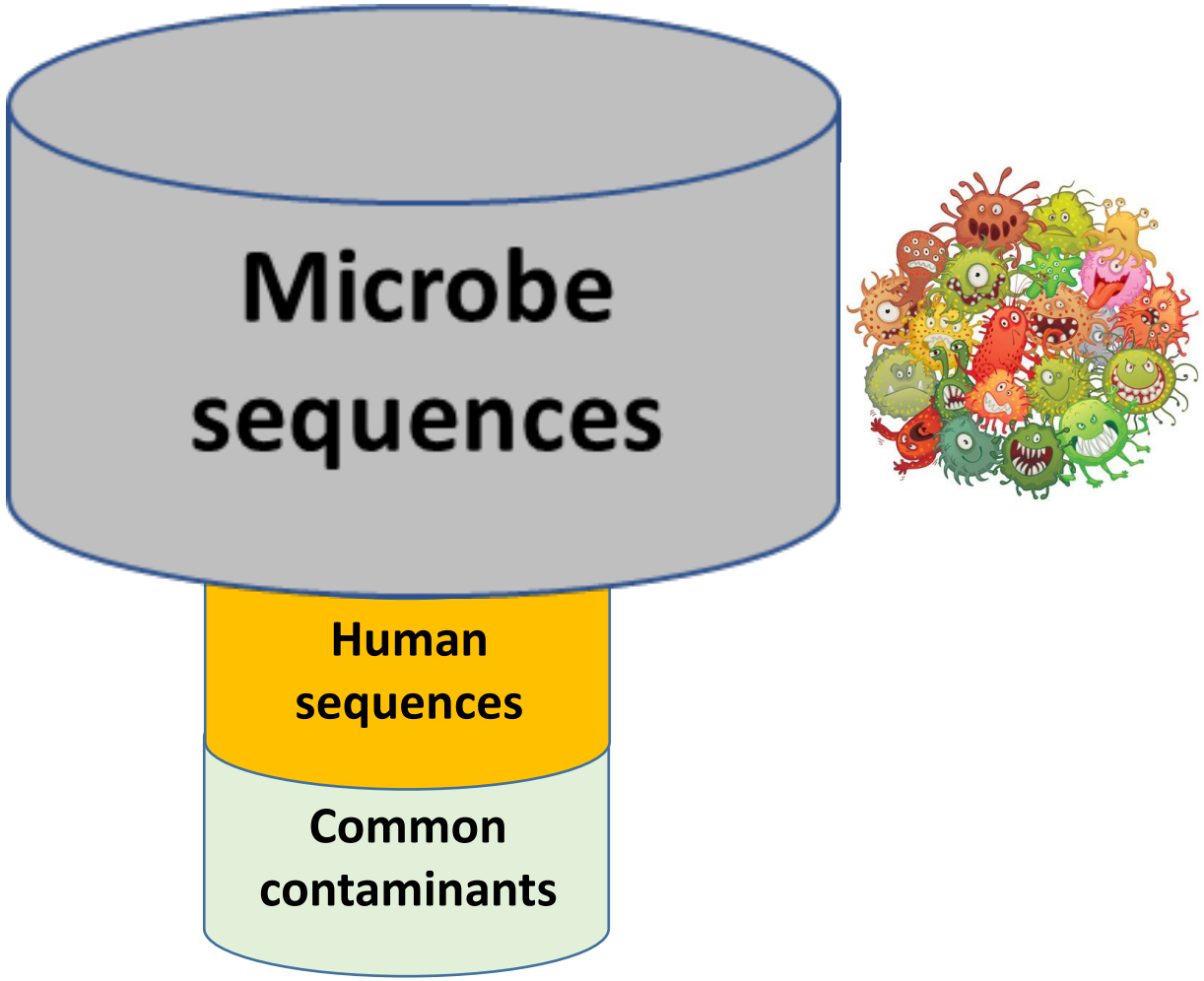
Single-organism proteomics	Metaproteomics

Search Database		
- Size	Small to medium size (10 K to 100K sequences)	Large (1 million and above)
- Complexity	Single + Contaminants	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

The importance of the protein sequence database



GalaxyP
galaxyp.org

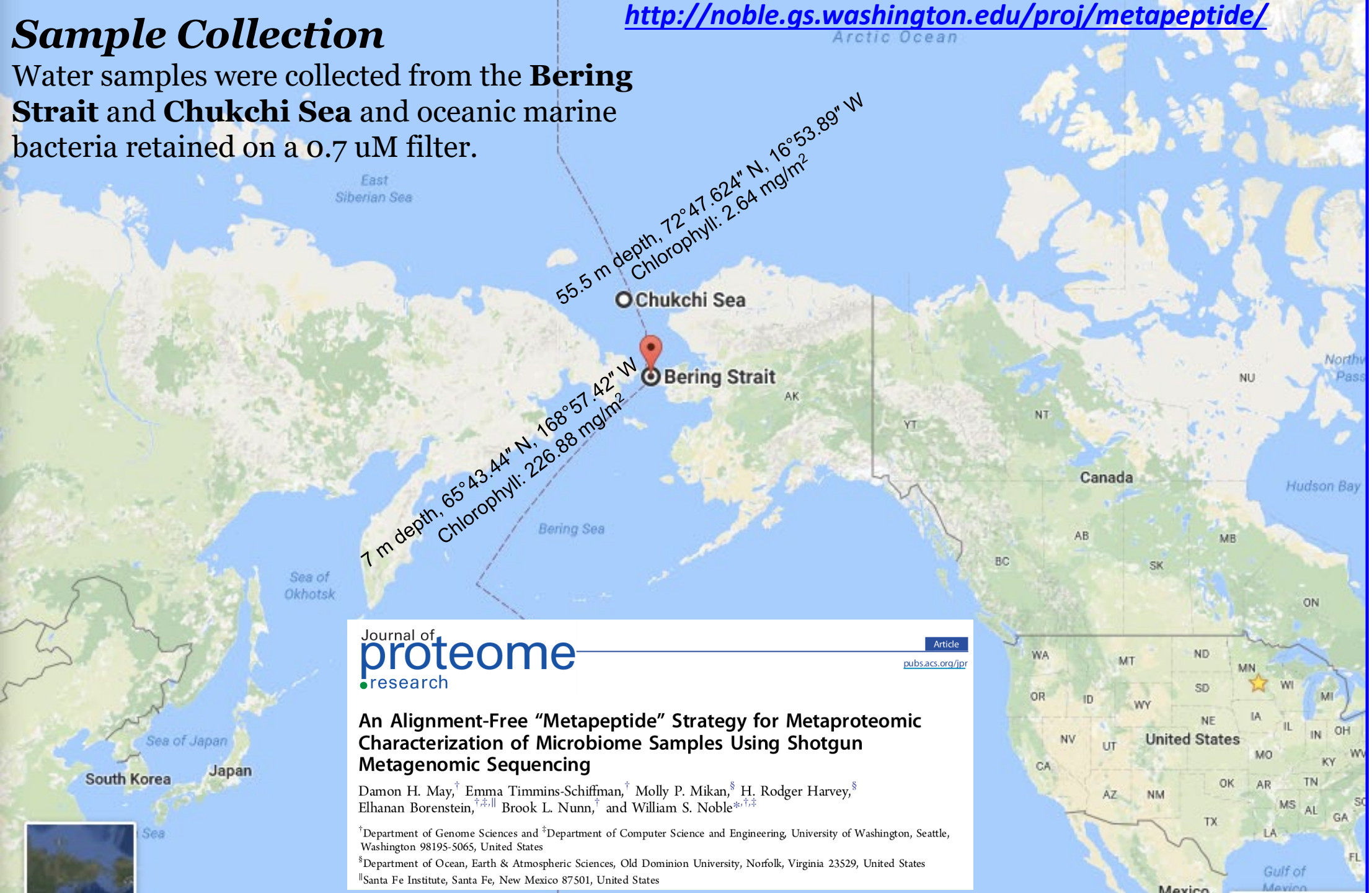
Sample Collection

Water samples were collected from the **Bering Strait** and **Chukchi Sea** and oceanic marine bacteria retained on a 0.7 uM filter.

http://noble.gs.washington.edu/proj/metapeptide/

Metagenome: Illumina HiSeq

Mass Spectrometry: Q-Exactive-HF



Journal of proteome research

Article

pubs.acs.org/jpr

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

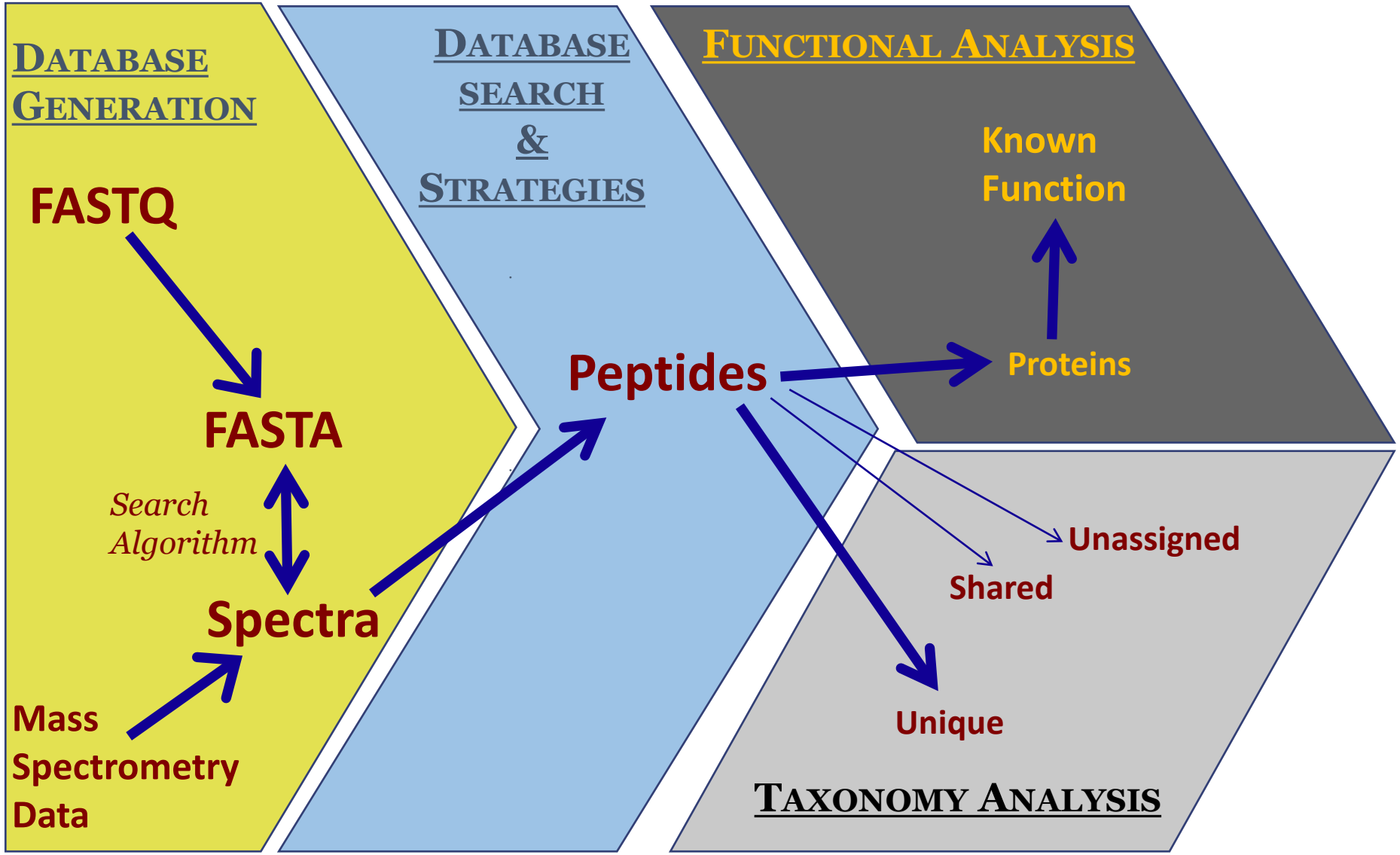
Damon H. May,[†] Emma Timmins-Schiffman,[†] Molly P. Mikan,[§] H. Rodger Harvey,[§] Elhanan Borenstein,[†,‡,∥] Brook L. Nunn,[†] and William S. Noble*[,†,‡]

[†]Department of Genome Sciences and [‡]Department of Computer Science and Engineering, University of Washington, Seattle, Washington 98195-5065, United States

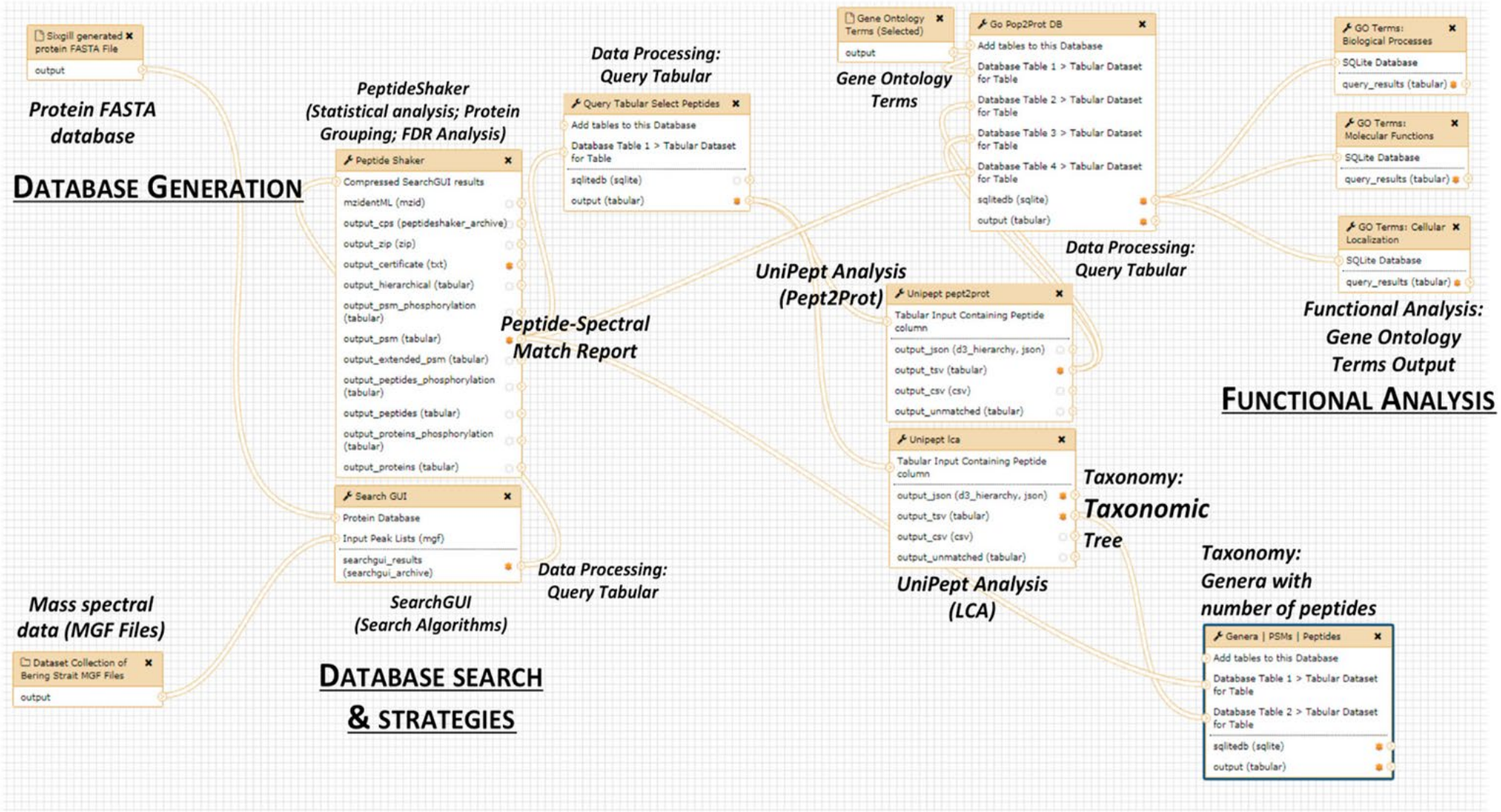
[§]Department of Ocean, Earth & Atmospheric Sciences, Old Dominion University, Norfolk, Virginia 23529, United States

[∥]Santa Fe Institute, Santa Fe, New Mexico 87501, United States

Metaproteomics Workflow



Solution: Galaxy Bioinformatics Platform



Software tools can be used in a sequential manner to generate analytical workflows that can be reused, shared and creatively modified.

Blank *et al* Proteomes 2018, 6(1), 7; https://doi.org/10.3390/proteomes6010007

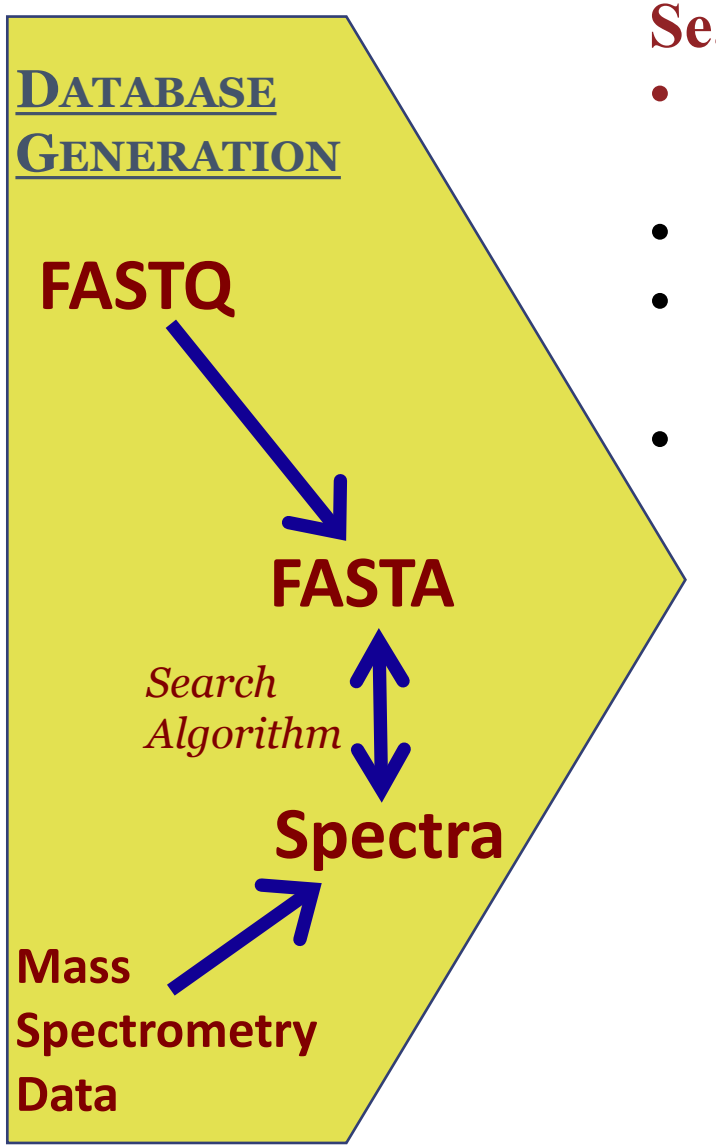
Hands On Session

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

https://usegalaxy.eu

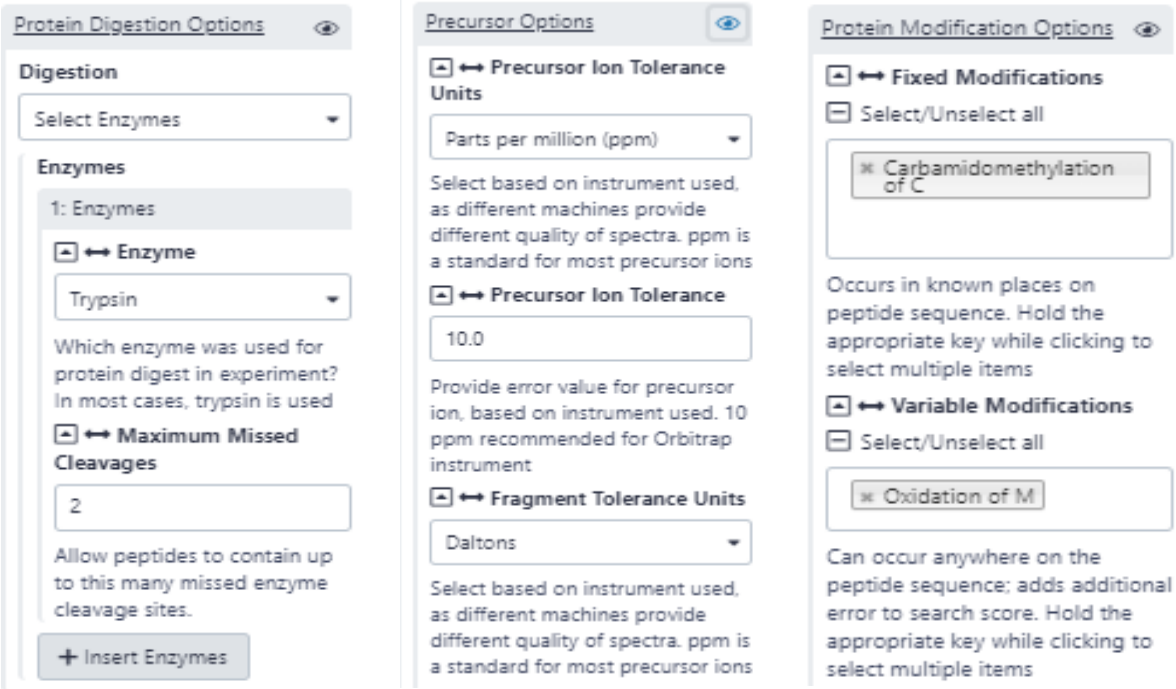
- Download and start workflow
- Observe outputs
- Tool Basics

SearchGUI

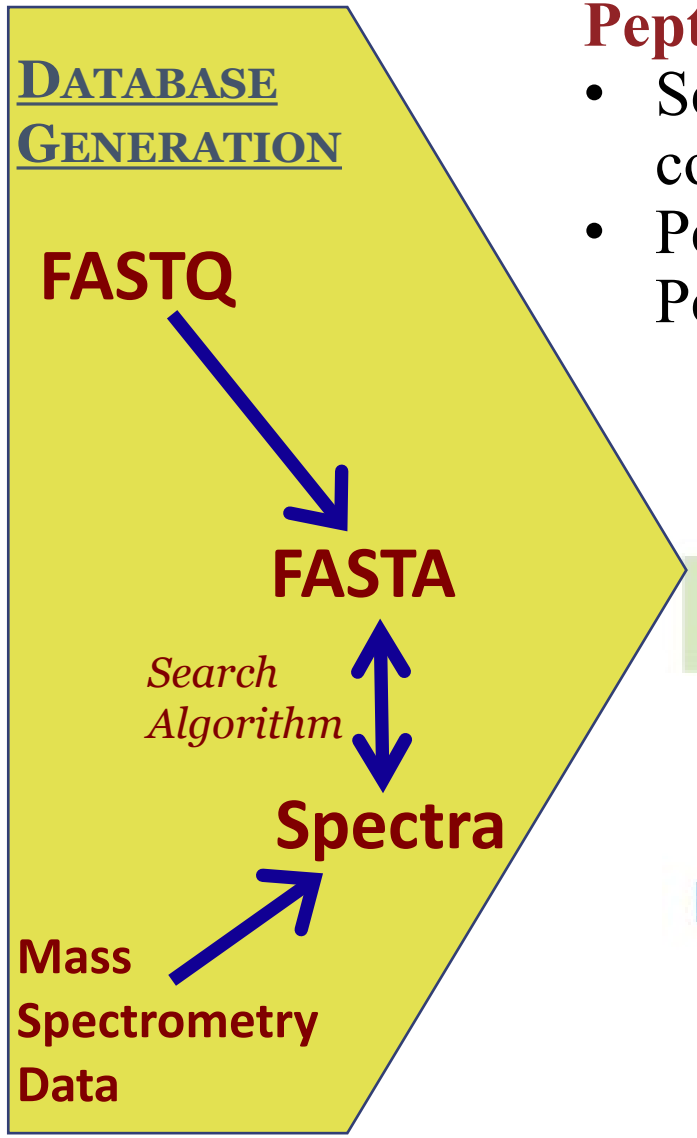


SearchGUI matches MS/MS spectra to peptide sequences

- 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

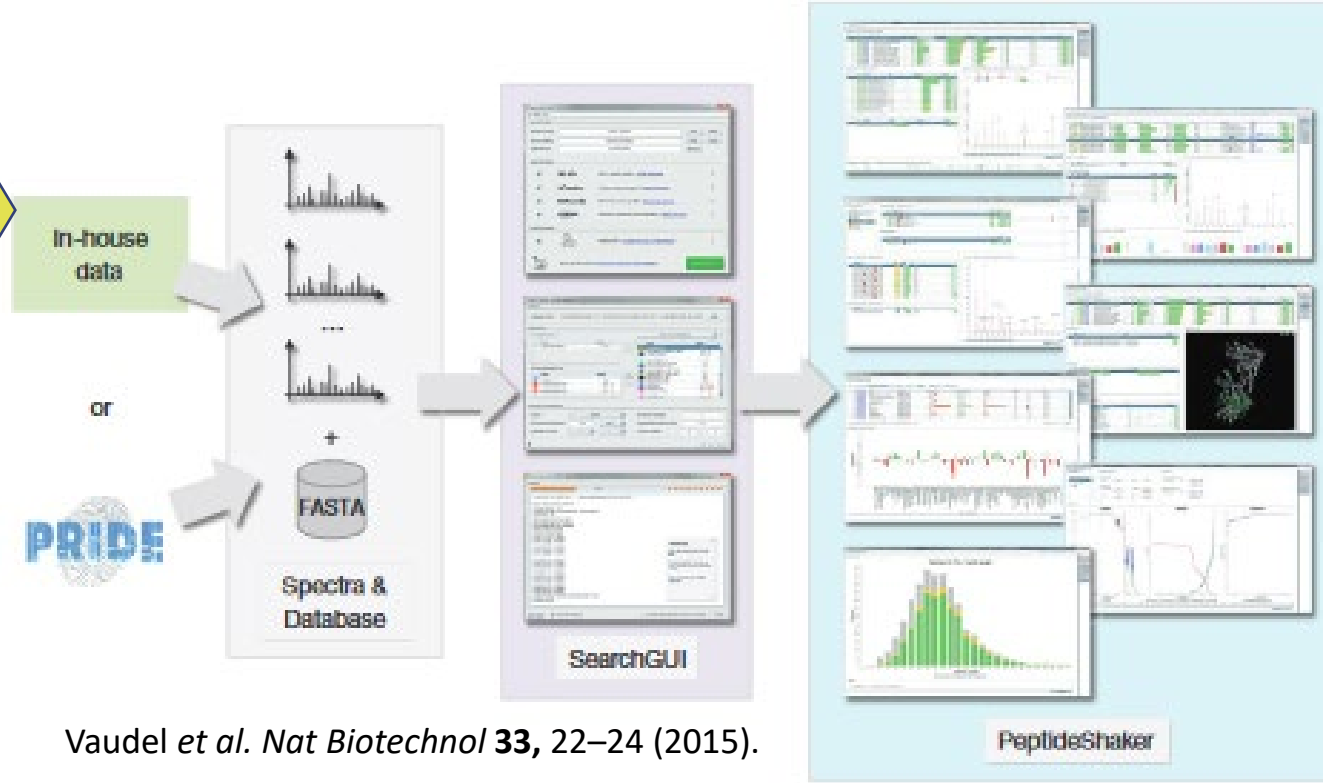


PeptideShaker



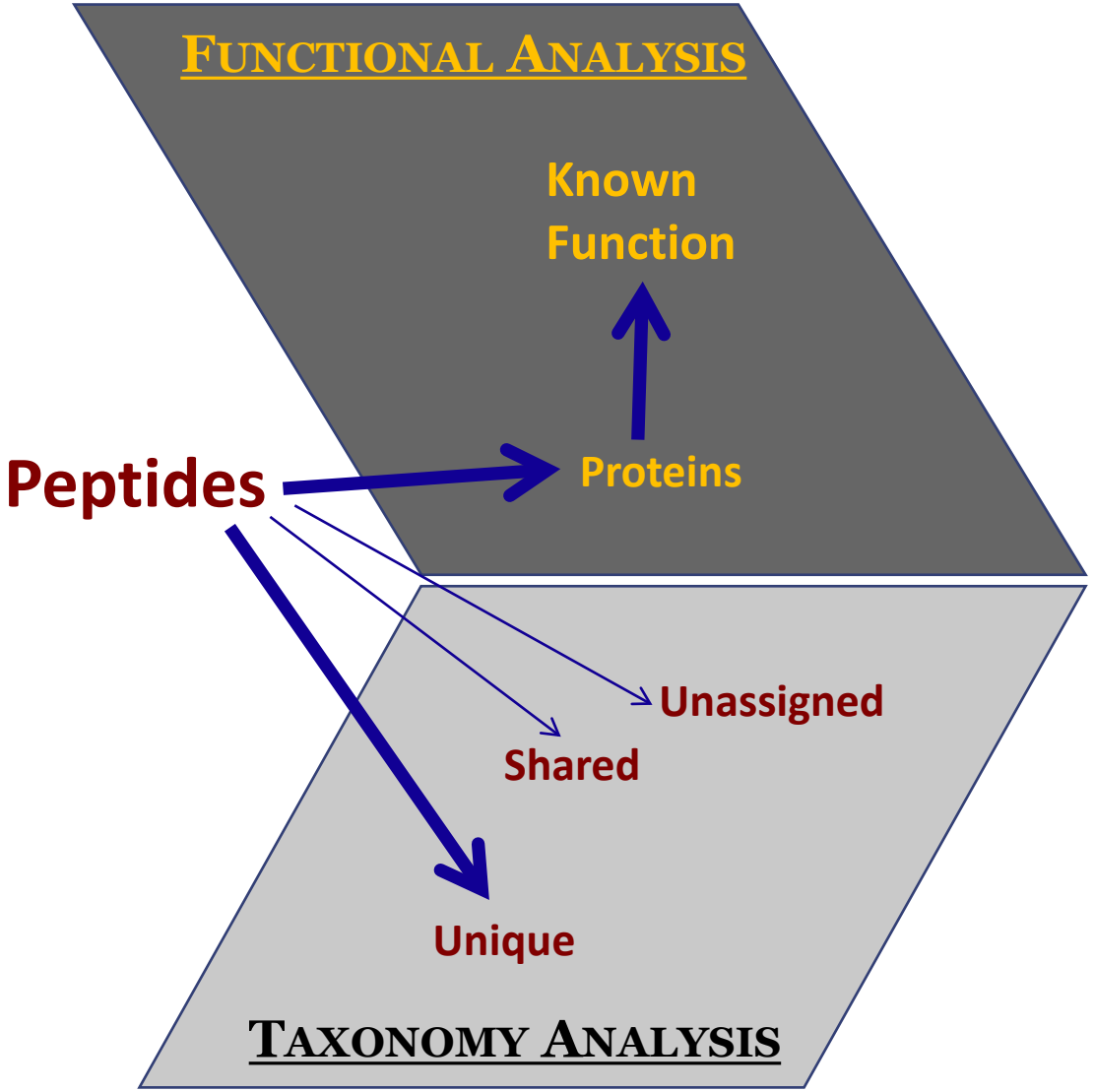
PeptideShaker filters SearchGUI results.

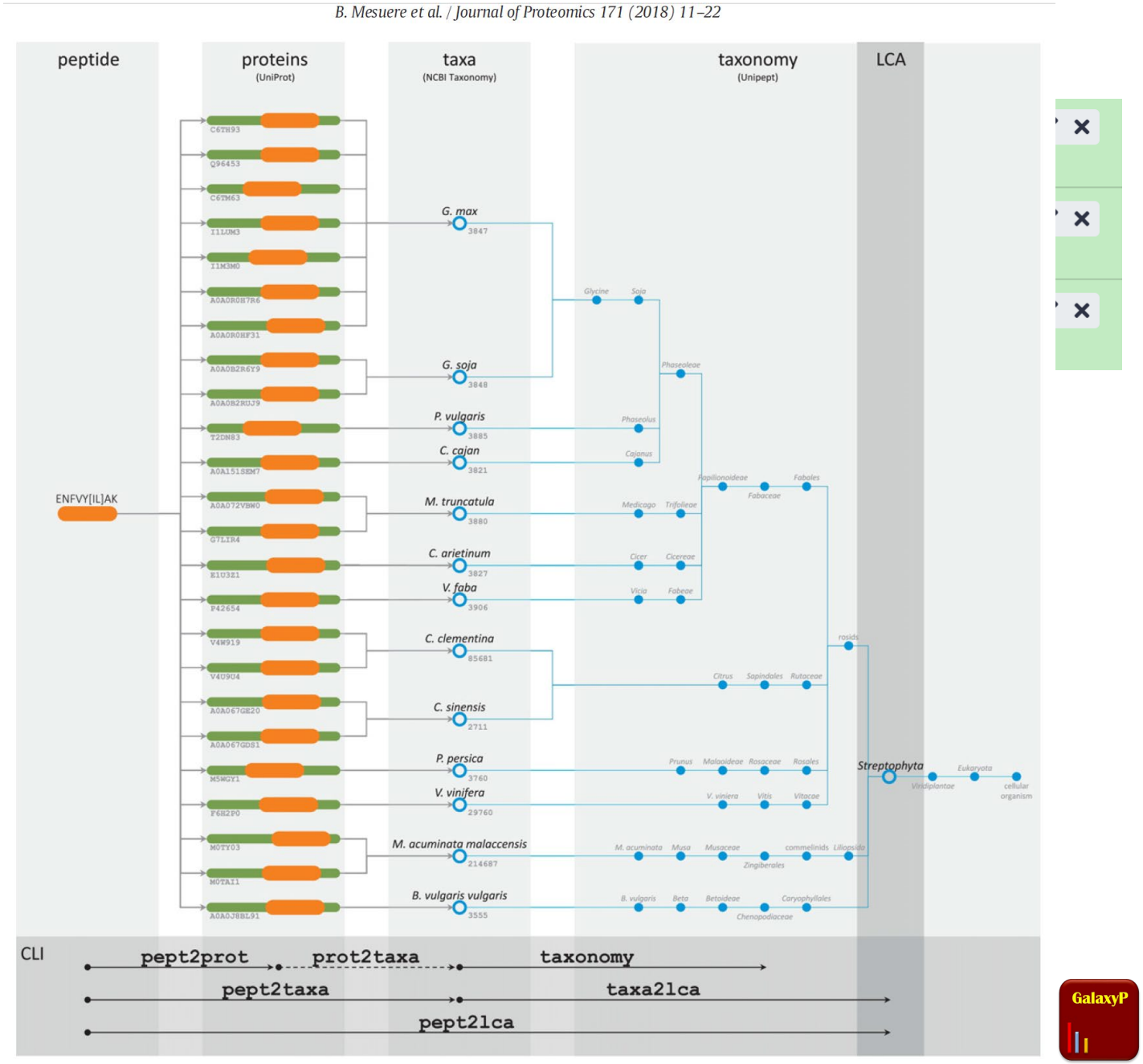
- Search GUI results are filtered by FDR to yield most confident peptide spectral matches (PSMs)
- Peptide Shaker generates outputs such as Protein Report, Peptide Report and mzIdentML files for subsequent analysis.



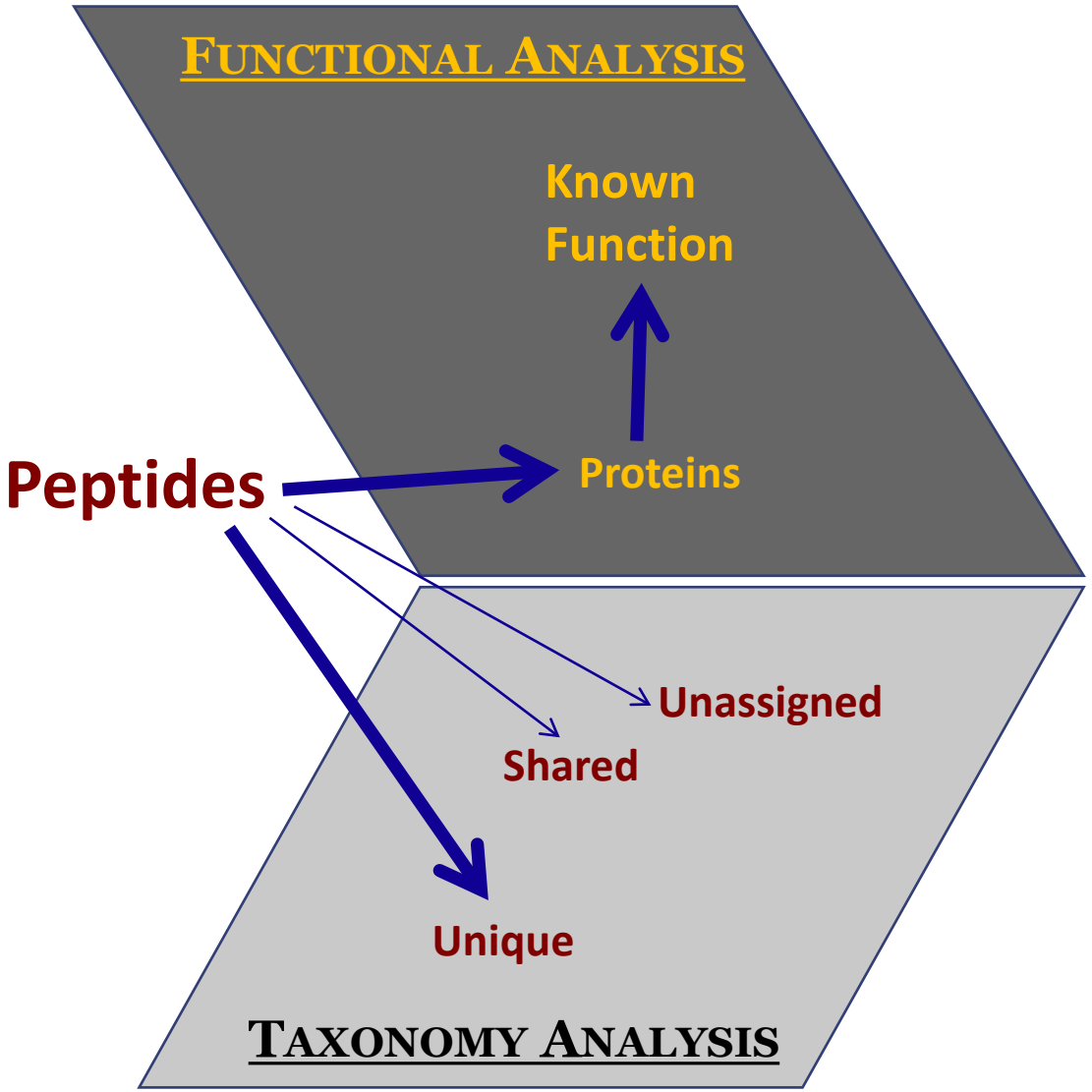
Vaudel *et al. Nat Biotechnol* **33,** 22–24 (2015).

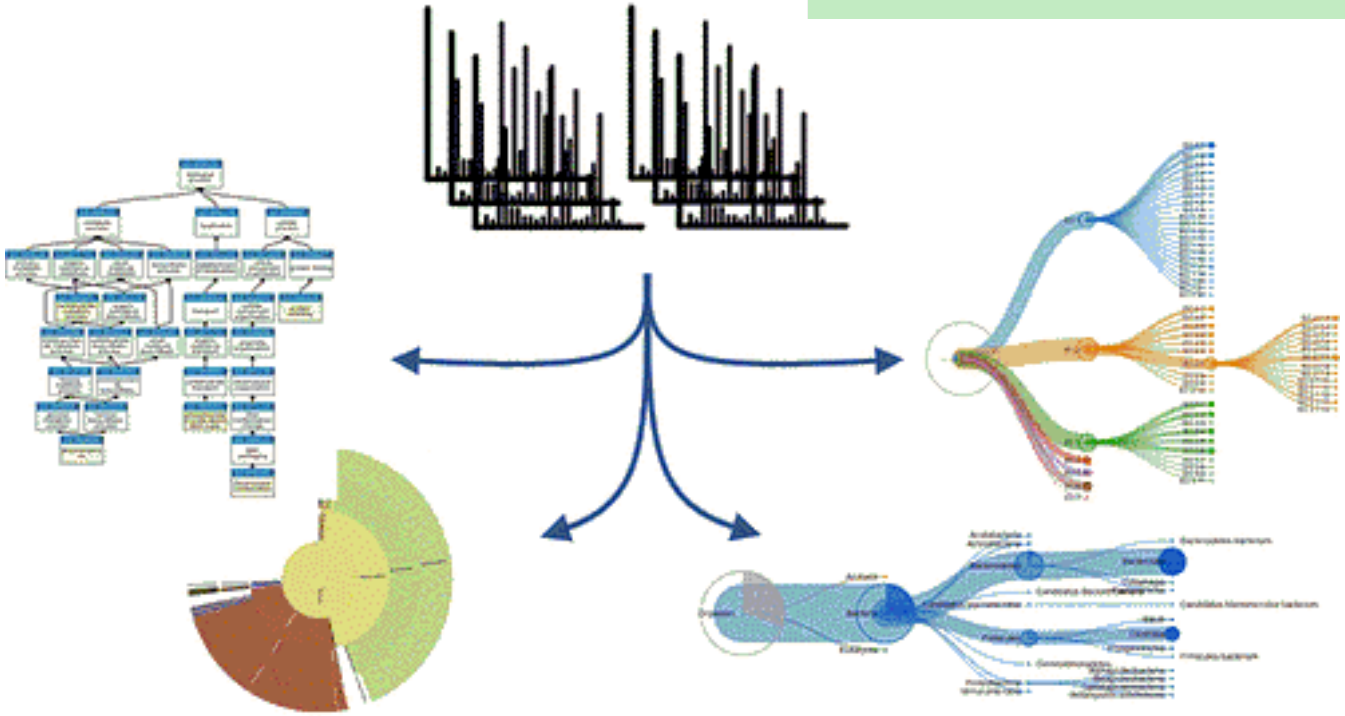
UNIPEPT


FUNCTIONAL ANALYSIS
Known Function
Proteins
Peptides
Unassigned
Shared
Unique
TAXONOMY ANALYSIS


B. Mesuere et al. / Journal of Proteomics 171 (2018) 11–22
peptide
proteins
(UniProt)
taxa
(NCBI Taxonomy)
taxonomy
(Unipept)
LCA
ENFVY[IL]AK
G. max
G. soja
P. vulgaris
C. cajan
M. truncatula
C. arietinum
V. faba
C. clementina
C. sinensis
P. persica
V. vinifera
M. acuminata malaccensis
B. vulgaris vulgaris
Streptophyta
CLI
pept2prot
prot2taxa
taxonomy
pept2taxa
taxa2lca
pept2lca
GalaxyP

Unipept





Taxonomic and Functional analysis with Unipept

https://unipept.ugent.be/publications

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…	**EGGEDMFVHK**	
2	GKRVAAAVGTVPE…	**VAAAVGTVPEQEWLK**	

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	GKRVAAAVGTVPEQEWLK, KVAAAVGTVPEQEWLK, RVAAAVGTV…	VAAAVGTVPEQEWLK	

Unipept

Pept2lca
taxonomy lowest common ancestor for a peptide

Pept2prot
Uniprot entries for a peptide with GO terms

peptide	superkingdom	…	genus
AADGHTMHFDVITGEK	Archaea		Nitrosopumilus
AALESFTGNVTSALK	Bacteria		Polaribacter

peptide	uniprot_id	…	go_references
AAEKLSAAQAR	W5T6F9		GO:0016021
AAEKLSAAQAR	A0A0Q6ZKK0		GO:0005524 GO:0016887 GO:0015833

Questions

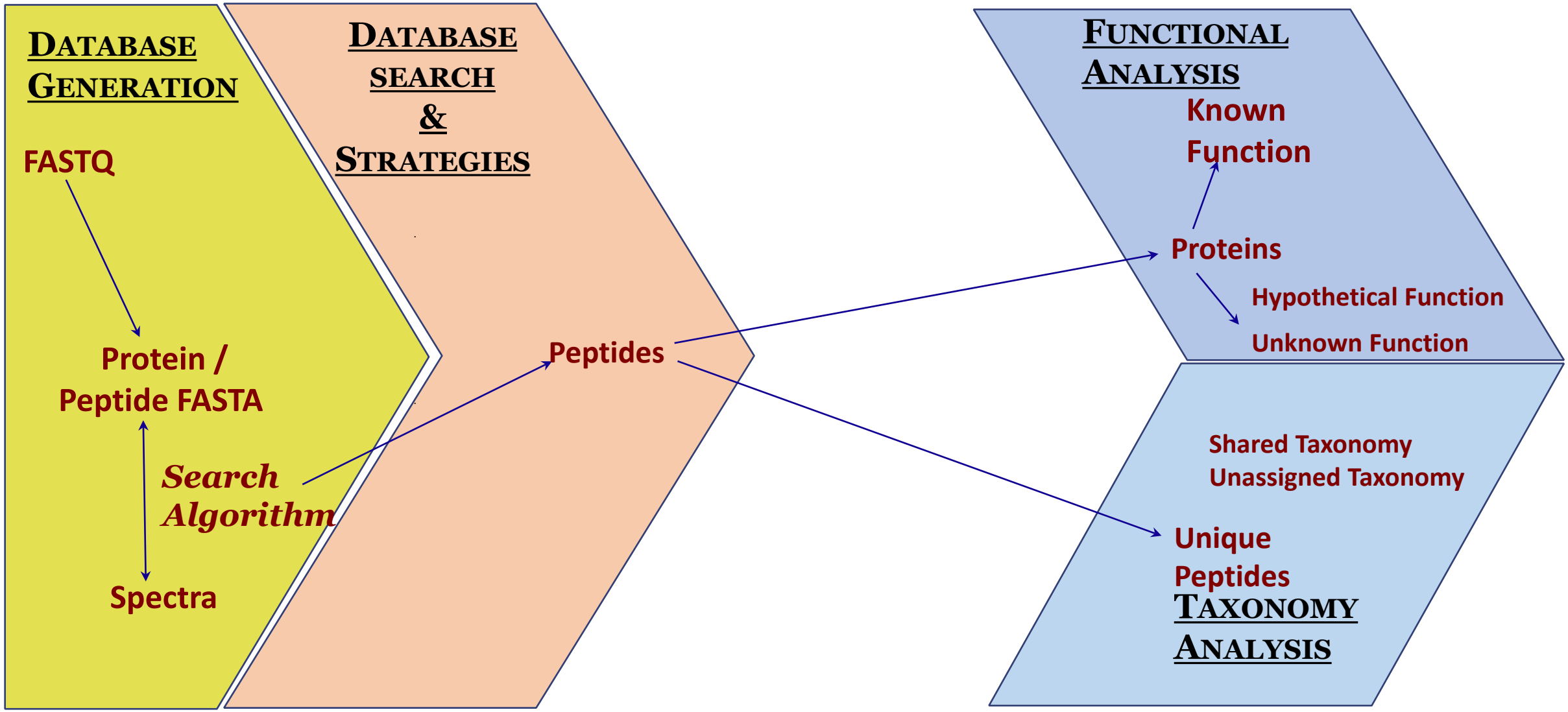
- How can I match metaproteomic mass spectrometry data to peptide sequences derived from shotgun metagenomic data?

- How can I perform taxonomy analysis and visualize metaproteomics data?

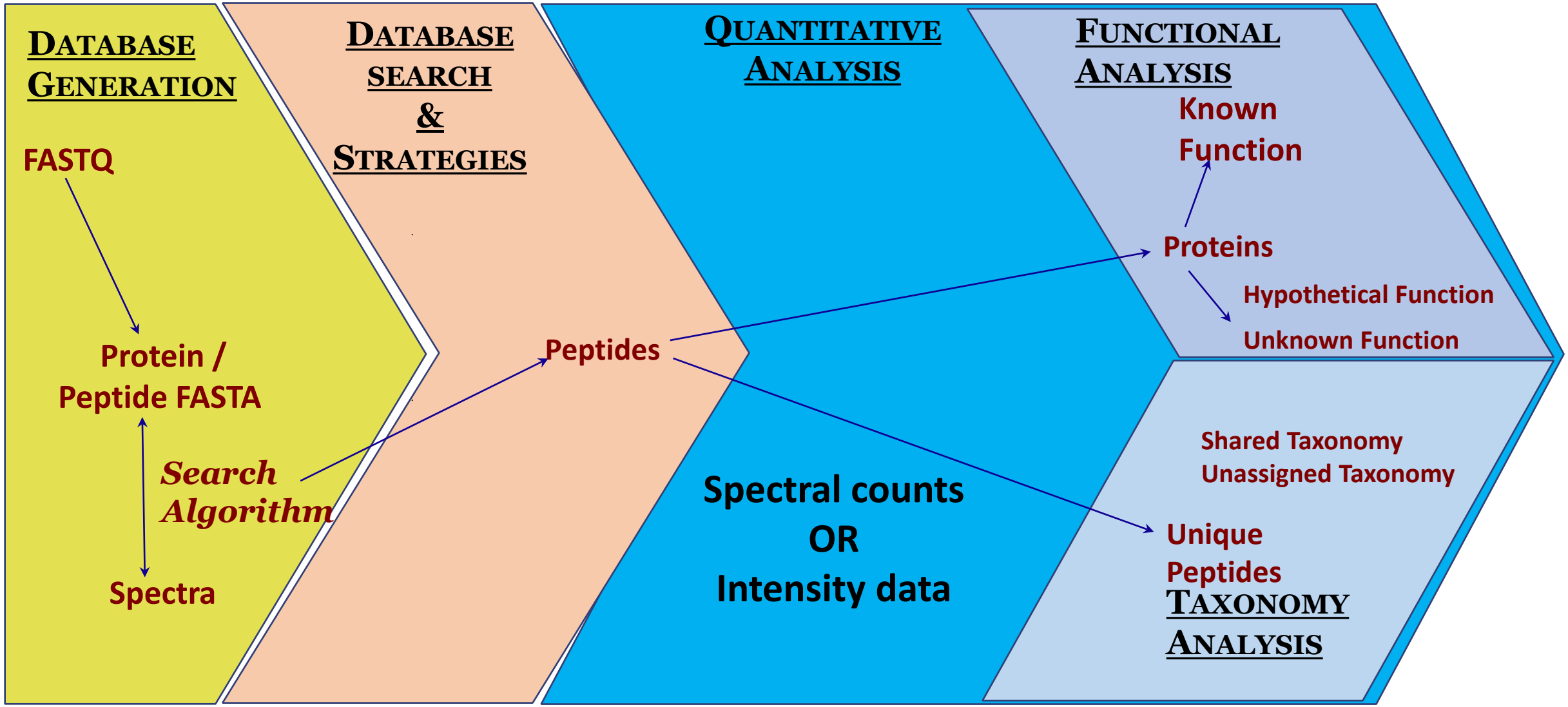
- How can I perform functional analysis on this metaproteomics data?

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

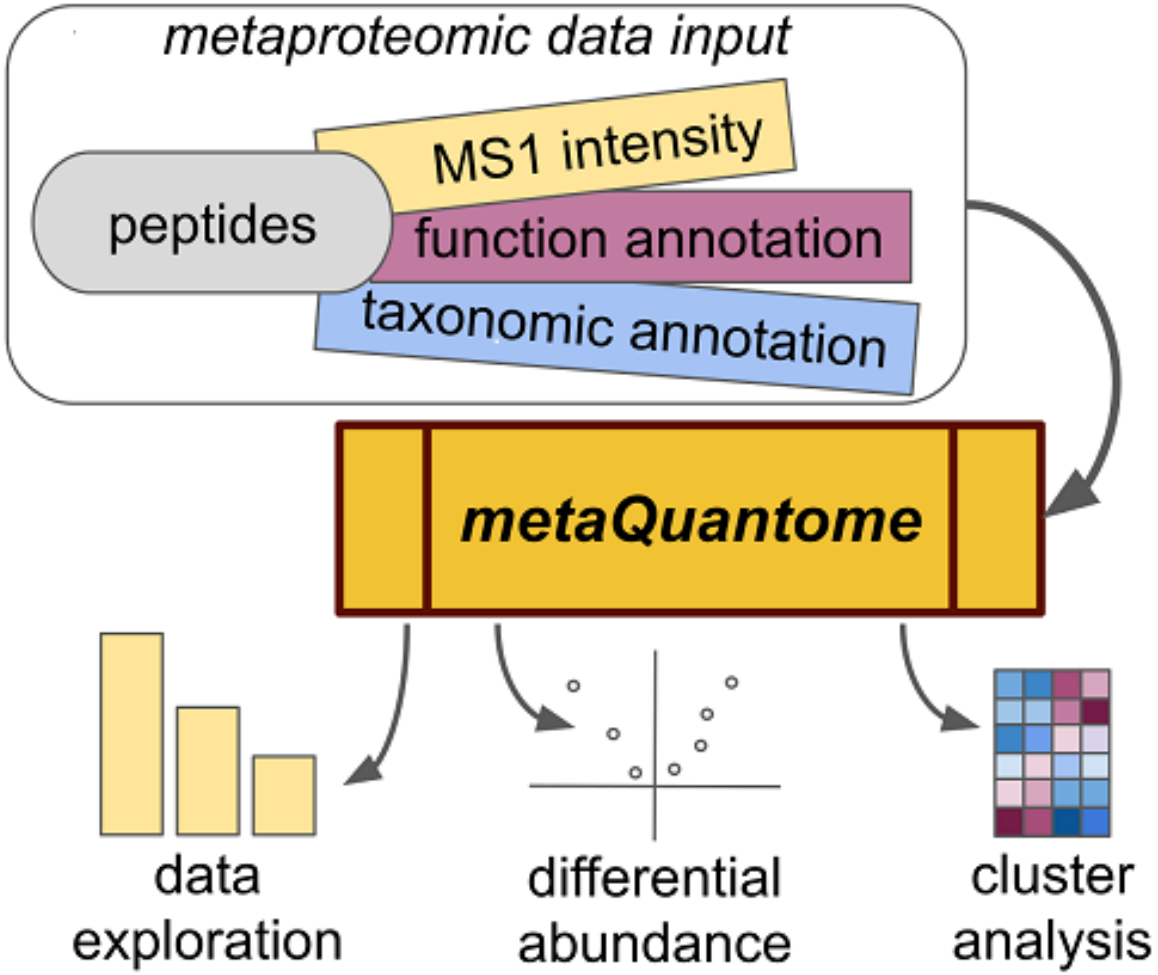
Metaproteomics Workflow



Metaproteomics Workflow



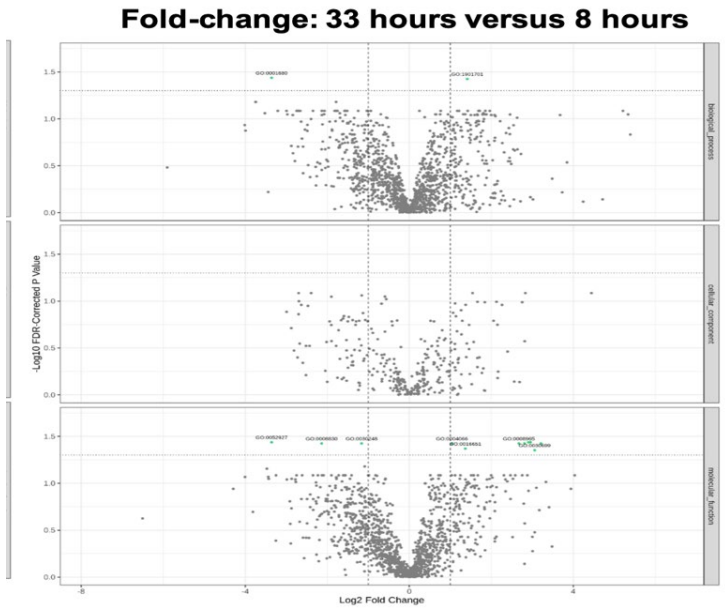
metaQuantome



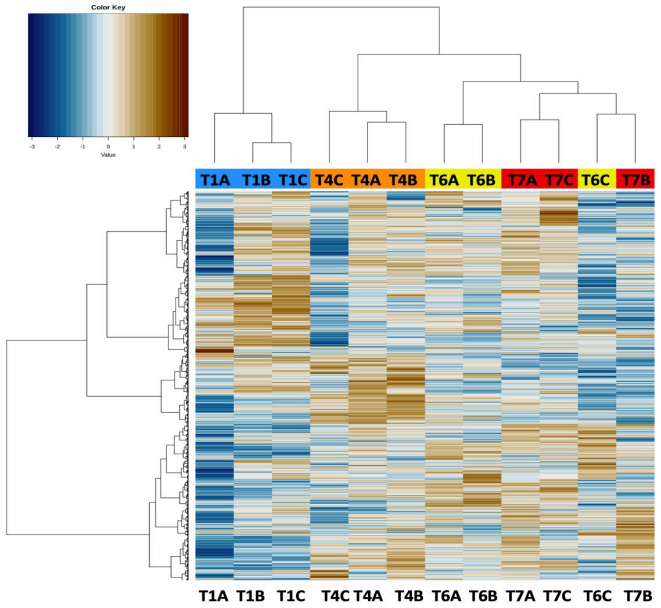
Easterly et al Mol Cell Proteomics. 2019 pii: mcp.RA118.001240. doi: 10.1074/mcp.RA118.001240.

Galaxy

FUNCTION: VOLCANO PLOTS

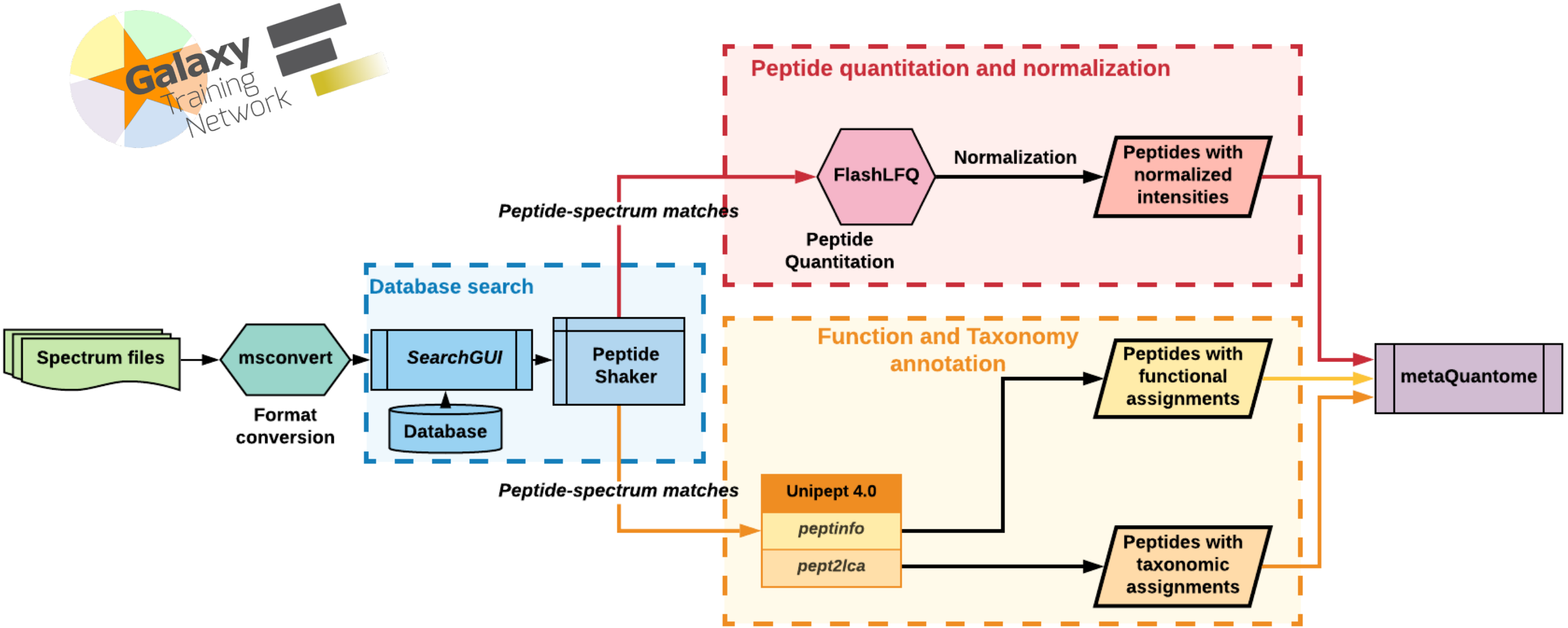


FUNCTION: HEATMAP



GalaxyP

metaQuantome on Galaxy Training Network



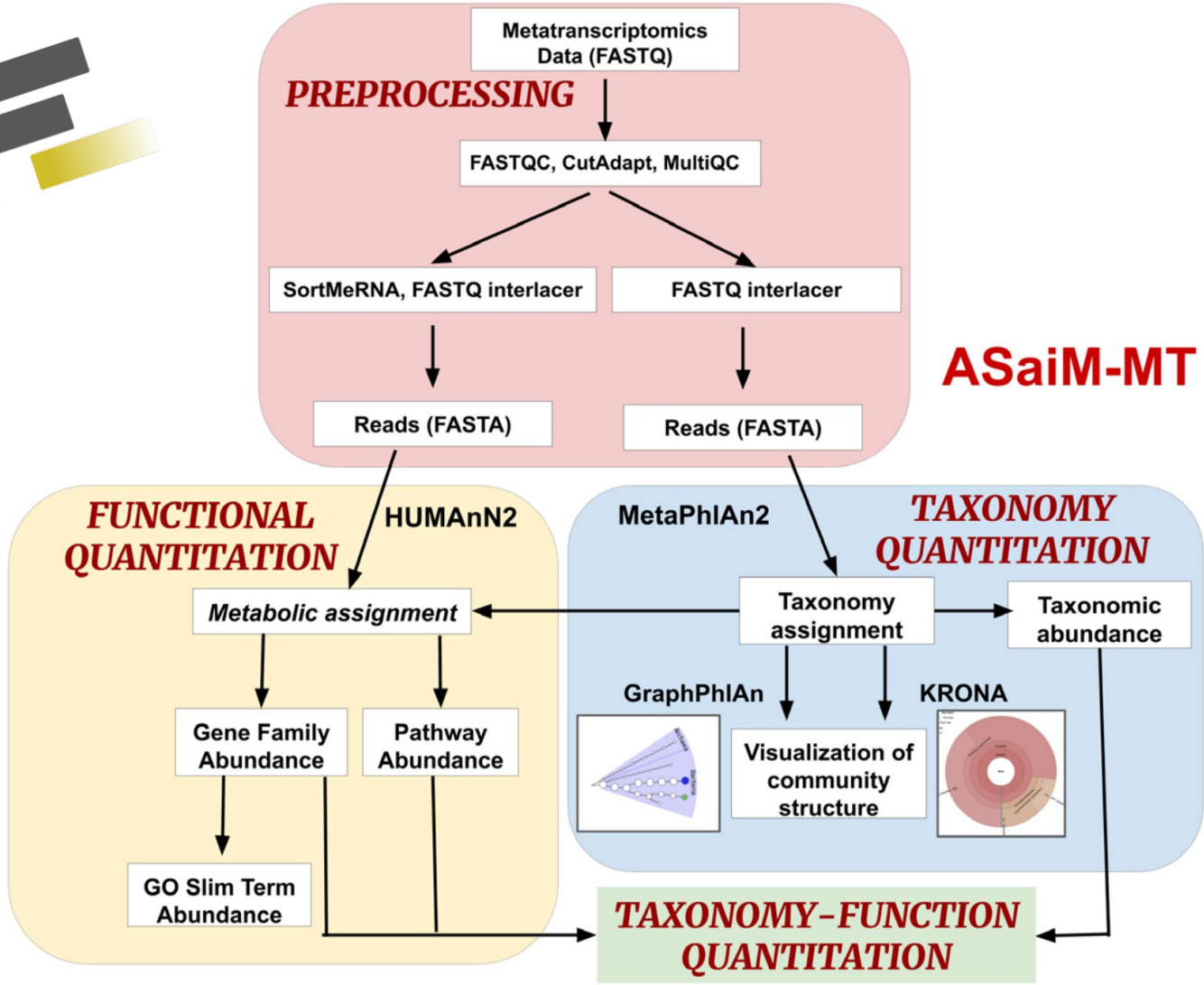
- https://training.galaxyproject.org/training-material/topics/proteomics/tutorials/metaquantome-data-creation/tutorial.html
- https://training.galaxyproject.org/training-material/topics/proteomics/tutorials/metaquantome-function/tutorial.html
- https://training.galaxyproject.org/training-material/topics/proteomics/tutorials/metaquantome-taxonomy/tutorial.html

Metatranscriptomics on Galaxy Training Network (ASaiM-MT)

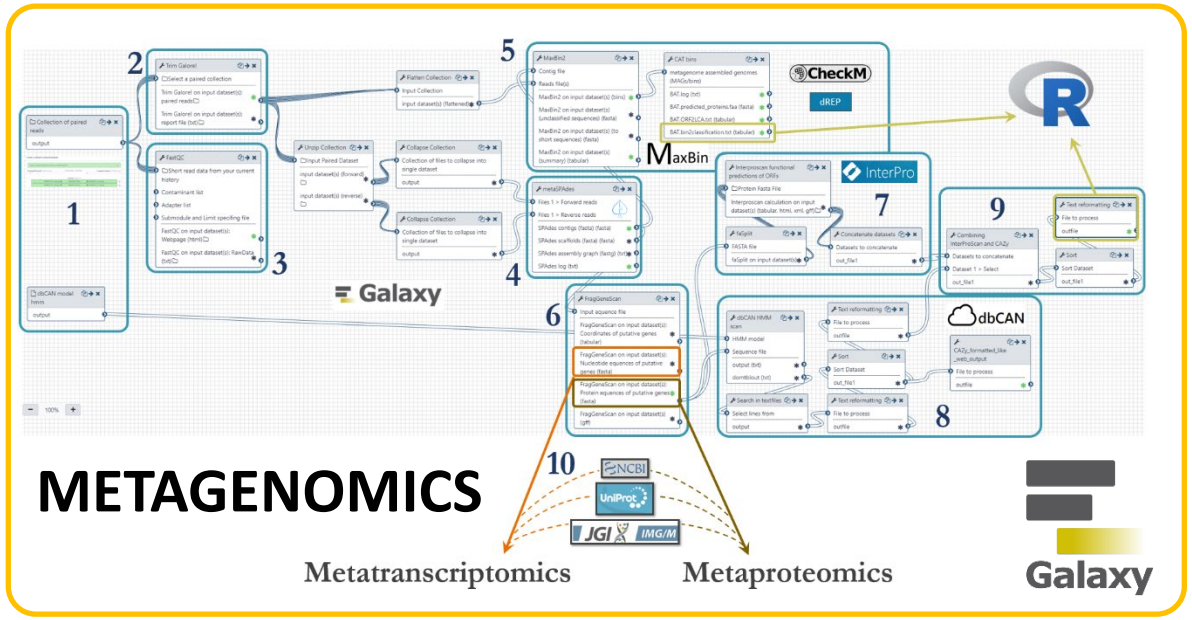
https://training.galaxyproject.org/training-material//topics/metagenomics/tutorials/metatranscriptomics/tutorial.html

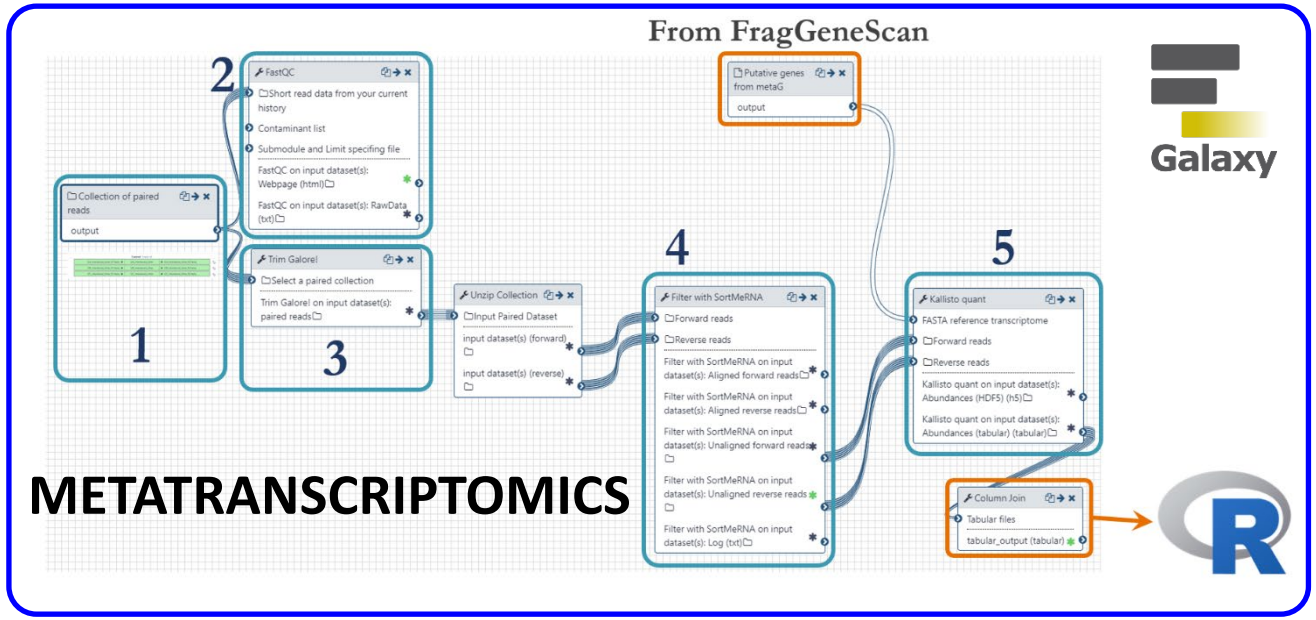
ASaiM

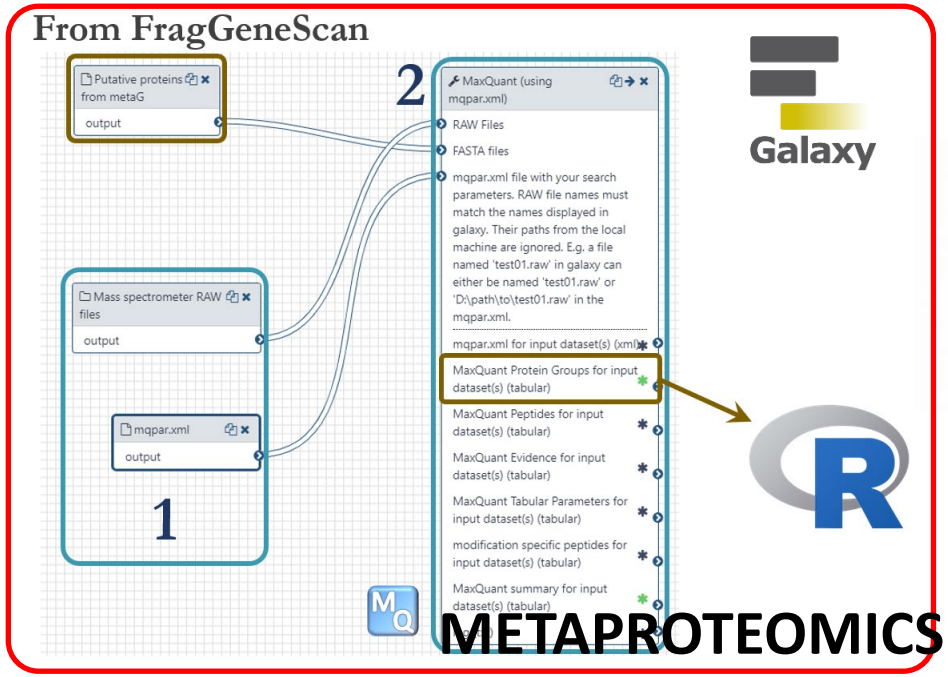
Batut et al Gigascience. 2018 7(6)
doi: 10.1093/gigascience/giy057



META-OMICS APPROACH BY NMBU TEAM







Francesco Delogu

Magnus Arntzen

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

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.

ACS Publications Most Trusted. Most Cited. Most Read.

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

Accessing tools and Workflows

METAGENOMICS:
Toolshed: *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*
Galaxy Training Network: *http://z.umn.edu/gtn-metaproteomics*

Also available on : https://proteomics.usegalaxy.eu/ **and**
Metaproteomics Gateway: z.umn.edu/metaproteomicsgateway

galaxyp.org/contact

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

Andrew McCardle
Imperial College, London, UK

Brook Nunn
U of Washington

Saskia Hiltemann

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

Anamika Krishanpal
Persistent Systems Limited India

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

Alessandro Tanca
Porto Conte Ricerche, Italy

Jeremy Fisher
Indiana University

Valdemir Carvalho
Fleury Group, Brazil

Bjoern Gruening
Bérénice Batut
University of Freiburg, Freiburg, Germany

Maria Doyle
Melbourne , Australia

Matt Chambers
Nashville, TN

Funding

NATIONAL CANCER INSTITUTE
Informatics Technology for Cancer Research

galaxyp.org

twitter.com/usegalaxyp