

Introduction to Metaproteomics

Learn more at galaxyp.org
z.umn.edu/itcrgalaxyvideo

Workshop acknowledgements

- **Instructors**

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



- **Other contributors**

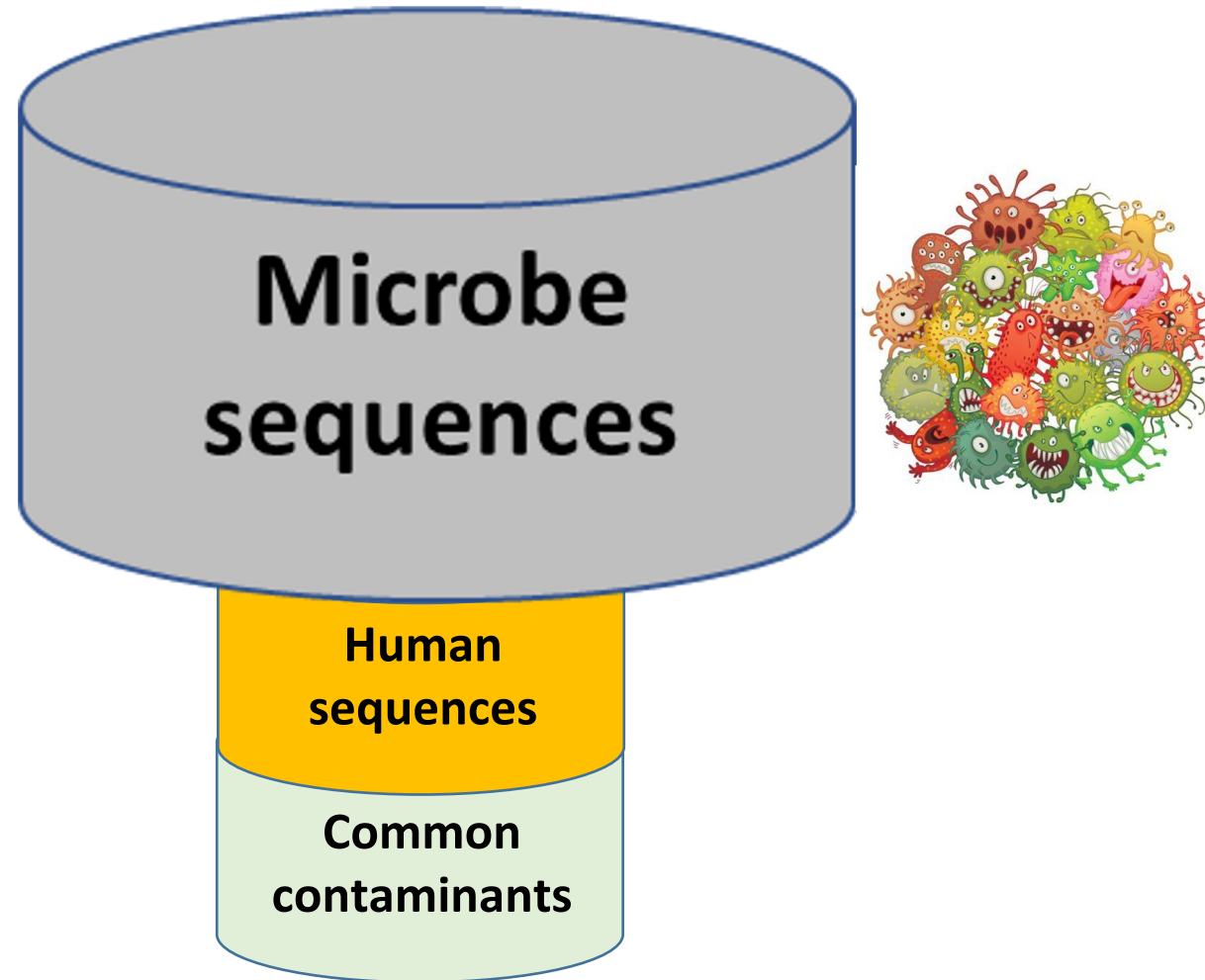
- Andrew Rajczewski
- Reid Wagner
- Katherine Do
- usegalaxy.eu team
- Galaxy community



NATIONAL CANCER INSTITUTE
Informatics Technology for
Cancer Research

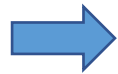


The importance of the protein sequence database



HANDS-ON SESSION

Instructions



Please **Register** for creating an account with a valid email ID and Password at usegalaxy.eu.

Once Registered, click on TIAAS to join the GCC 2022 Galaxy session.
<https://usegalaxy.eu/join-training/quant-meta>

Go to Shared Data

Go to Shared Data

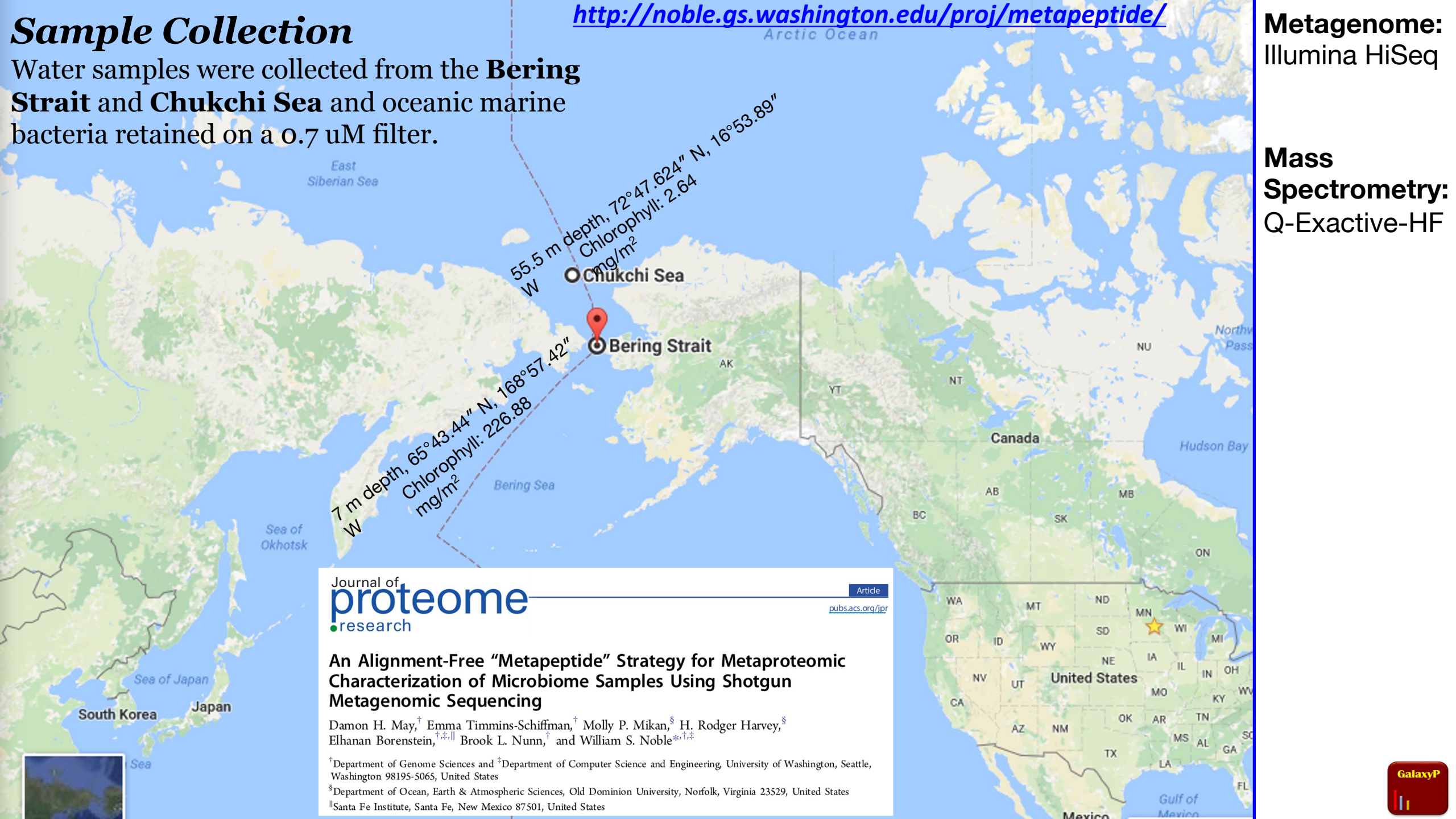
Run the workflow on active history



Sample Collection

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

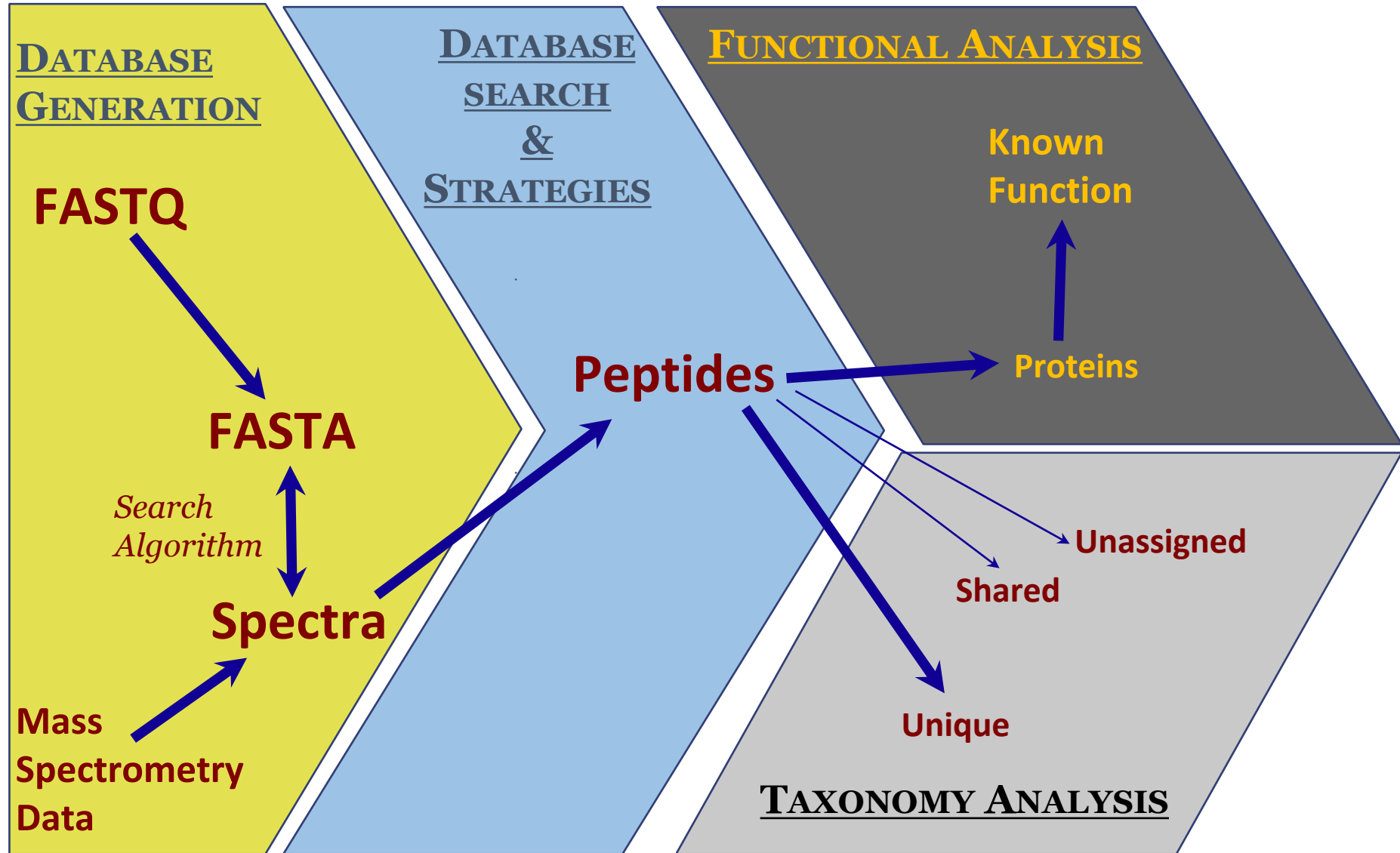
An Alignment-Free "Metapeptide" Strategy for Metaproteomic Characterization of Microbiome Samples Using Shotgun Metagenomic Sequencing

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

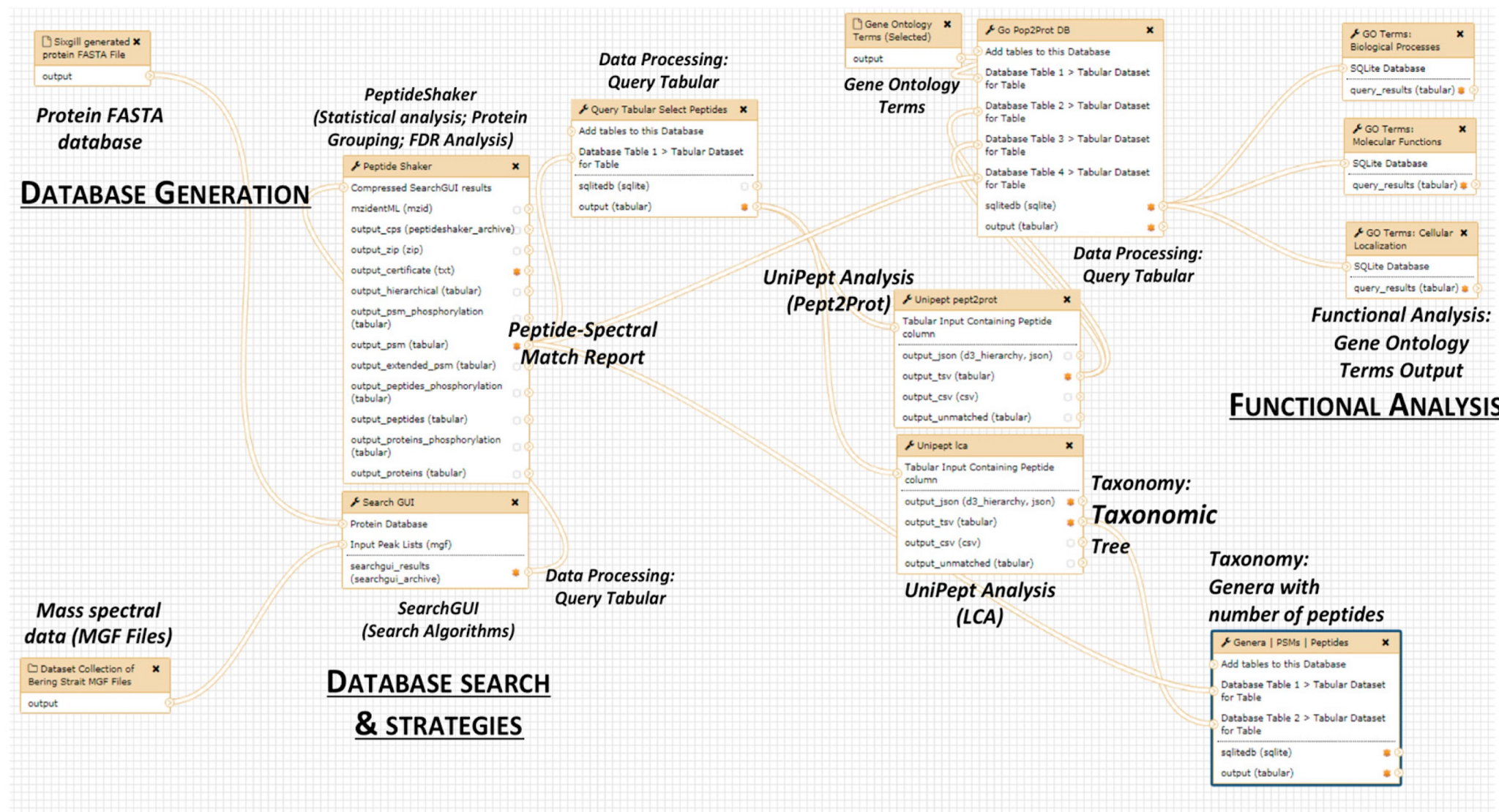
[†]Department of Genome Sciences and [§]Department of Computer Science and Engineering, University of Washington, Seattle, Washington 98195-5065, United States
[§]Department of Ocean, Earth & Atmospheric Sciences, Old Dominion University, Norfolk, Virginia 23529, United States
^{||}Santa Fe Institute, Santa Fe, New Mexico 87501, United States



Metaproteomics Workflow



Solution: Galaxy Bioinformatics Platform



FUNCTIONAL ANALYSIS

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

Hands On Session



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

<https://usegalaxy.eu/join-training/gcc2022>



<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

DATABASE GENERATION

FASTQ

FASTA

Search Algorithm

Spectra

Mass Spectrometry Data

The screenshot displays three panels of search parameters for SearchGUI:

- Protein Digestion Options:** Includes a 'Digestion' section with a 'Select Enzymes' dropdown, an 'Enzymes' section with a '1: Enzymes' header and an 'Enzyme' dropdown set to 'Trypsin', and a 'Maximum Missed Cleavages' section with a text input set to '2'. A note states: 'Allow peptides to contain up to this many missed enzyme cleavage sites.' There is an '+ Insert Enzymes' button at the bottom.
- Precursor Options:** Includes a 'Precursor Ion Tolerance Units' section with a dropdown set to 'Parts per million (ppm)', a text input for '10.0', and a 'Fragment Tolerance Units' section with a dropdown set to 'Daltons'. A note states: 'Provide error value for precursor ion, based on instrument used. 10 ppm recommended for Orbitrap instrument.'
- Protein Modification Options:** Includes a 'Fixed Modifications' section with a dropdown set to 'Carbamidomethylation of C' and a 'Variable Modifications' section with a dropdown set to 'Oxidation of M'. A note states: 'Occurs in known places on peptide sequence. Hold the appropriate key while clicking to select multiple items.'

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

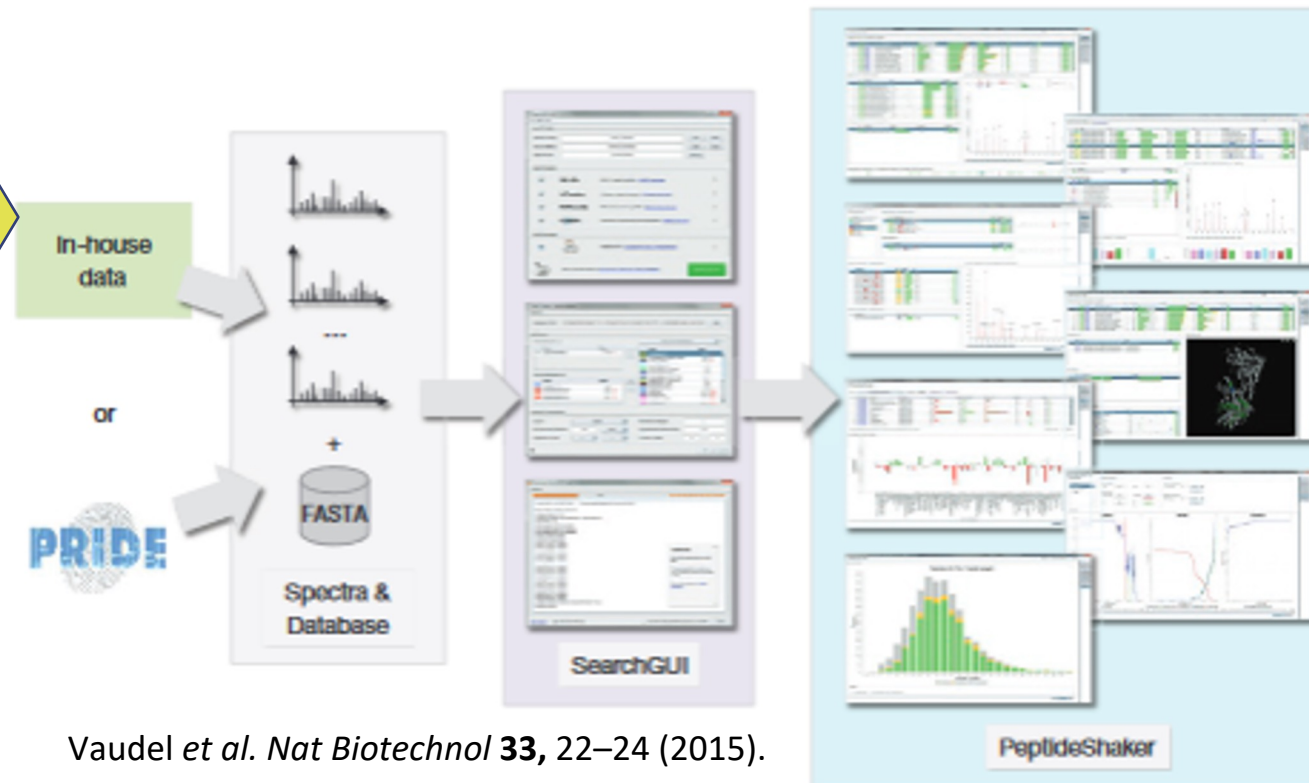
FASTQ

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

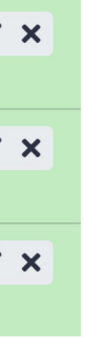
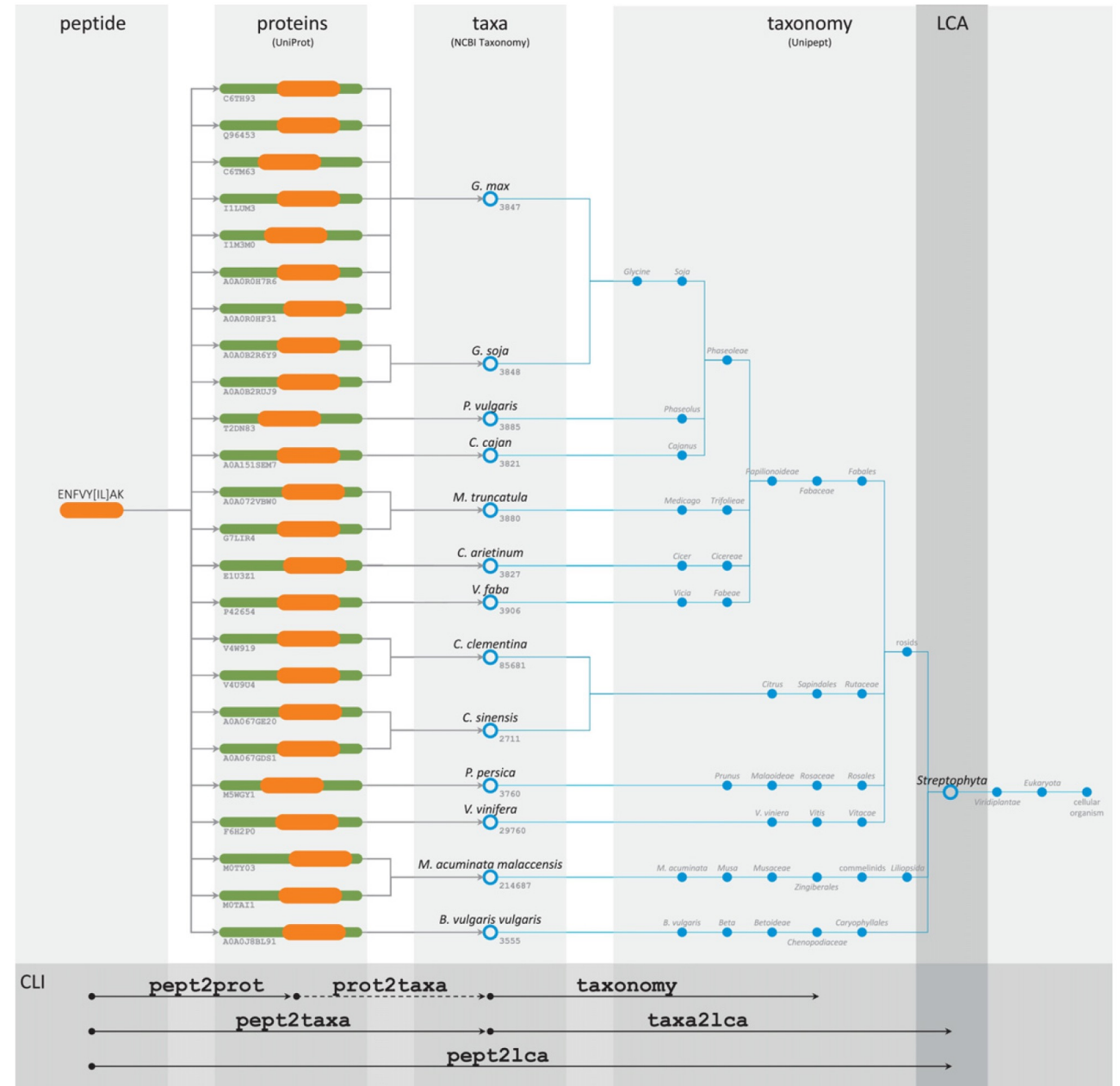
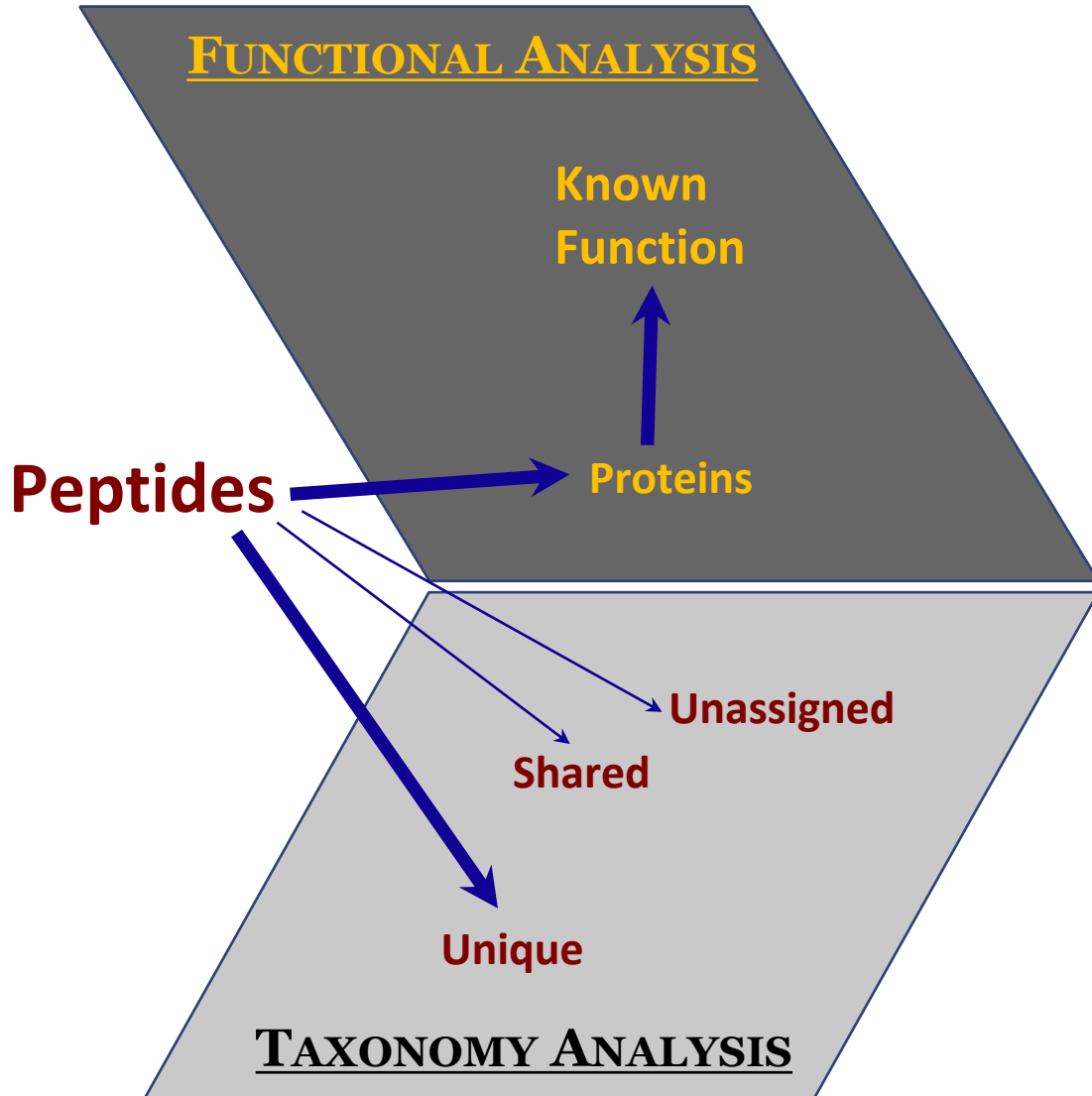


10: Search GUI on data 7, data 6, and others

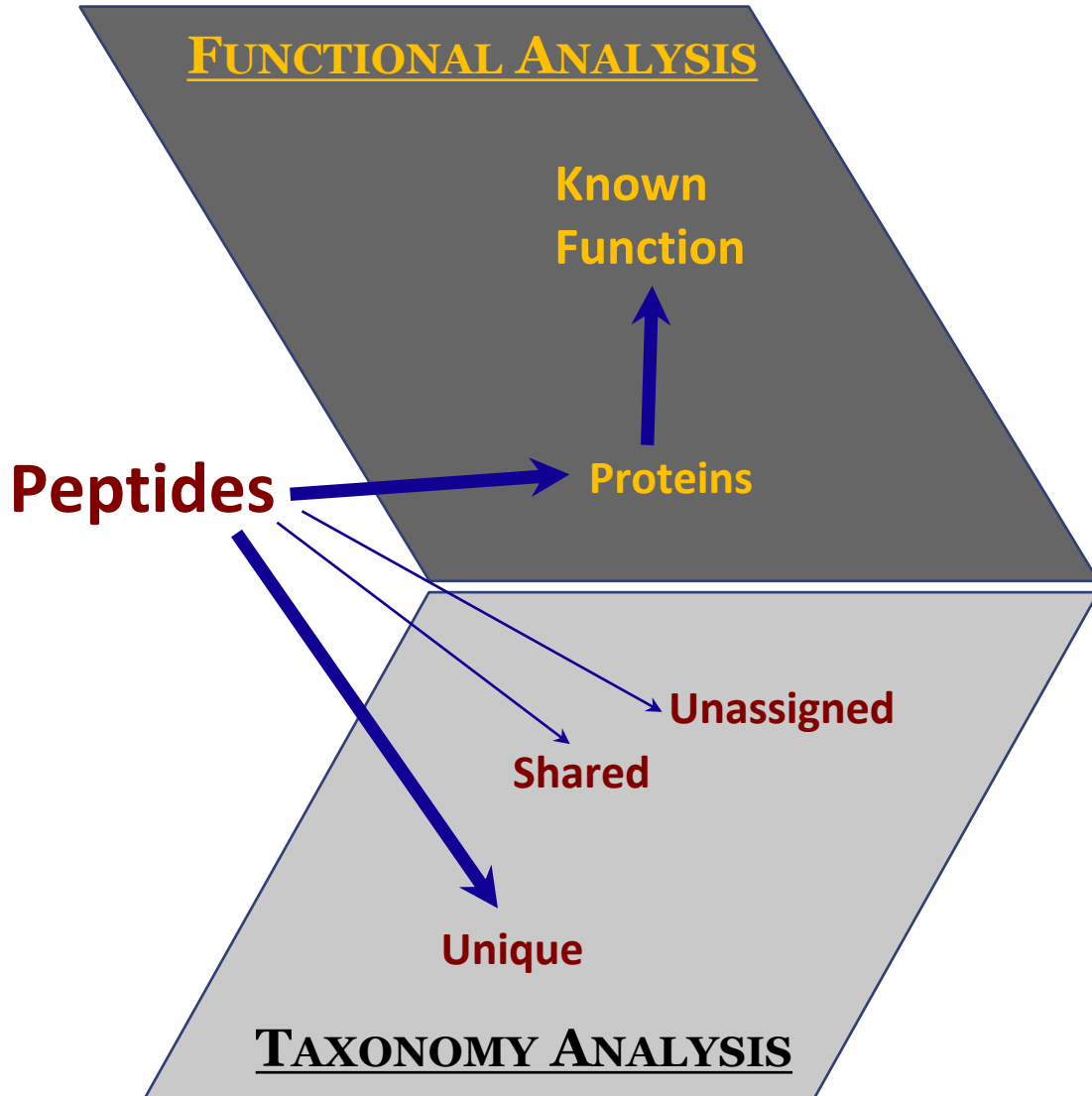


Unipept

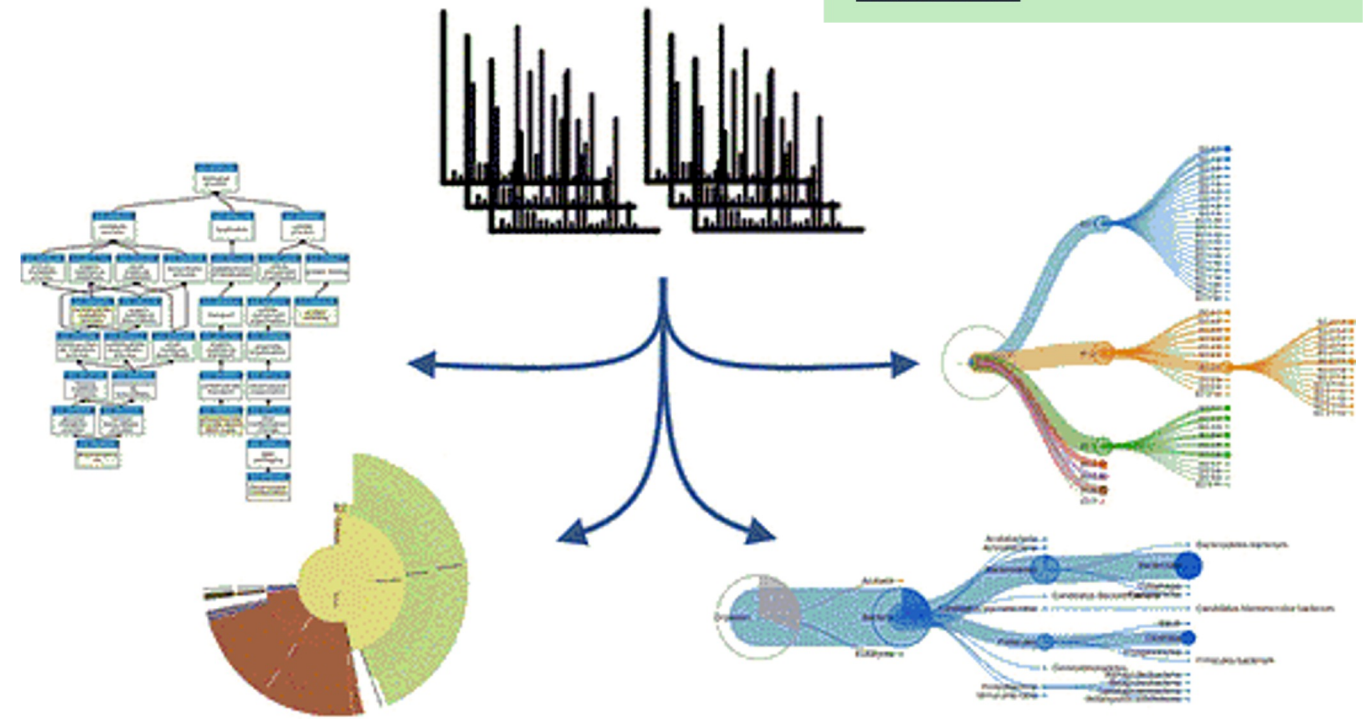
B. Mesuere et al. / Journal of Proteomics 171 (2018) 11–22



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>

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

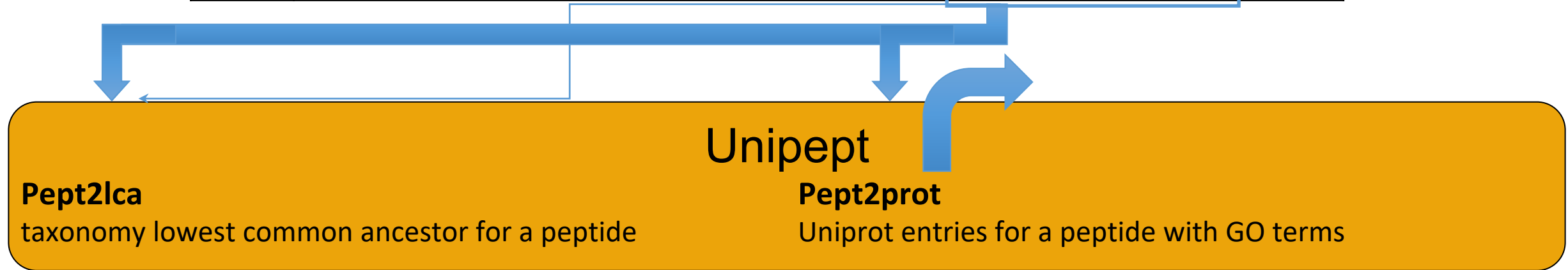
SQL query joins PSM and LCA to report number of PSMs and Peptides per genus

| 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 | 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>

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>



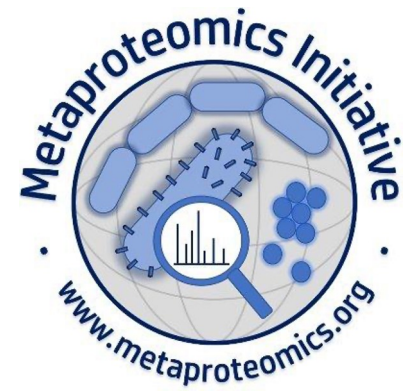
METAPROTEOMICS:

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

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

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

Matt Chambers
Nashville, TN

Funding



NATIONAL CANCER INSTITUTE
Informatics Technology for
Cancer Research



Maria Doyle
Melbourne, Australia

galaxyp.org

twitter.com/usegalaxy