Enabling Proteo-Informatic Analysis via Galaxy-P Platform

Pratik Jagtap

University of Minnesota

Google Scholar: z.umn.edu/pjgs
PubMed: z.umn.edu/pjagtapreferences

Email: pjagtap@umn.edu
Twitter: pratikomics
usegalaxyp

Learn more at ***galaxyp.org***

z.umn.edu/itcrgalaxyvideo

Enabling Proteo-Informatic Analysis via Galaxy-P Platform

- *MULTIOMICS RESEARCH*
- *MASS SPECTROMETRY DATA ANAYSIS*
- *GALAXY BIOINFORMATICS PLATFORM*
- *METAPROTEOMICS*

Central Dogma of Biology

DNA

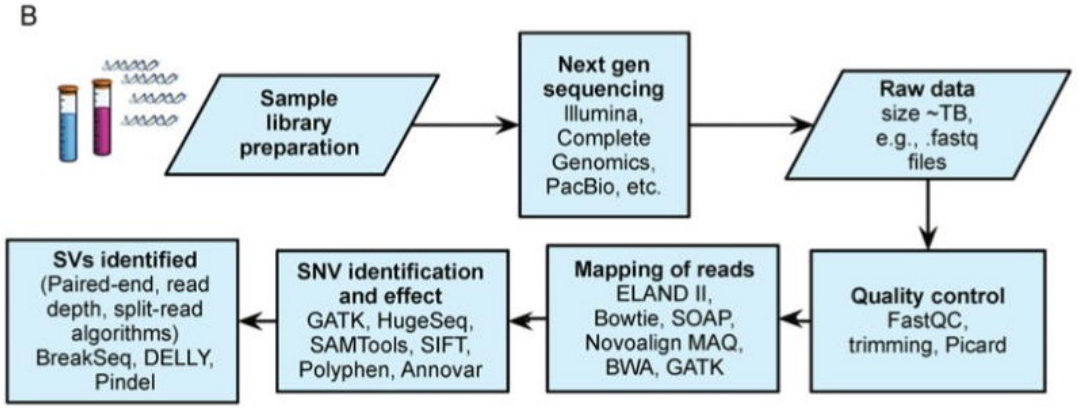
Transcription

RNA

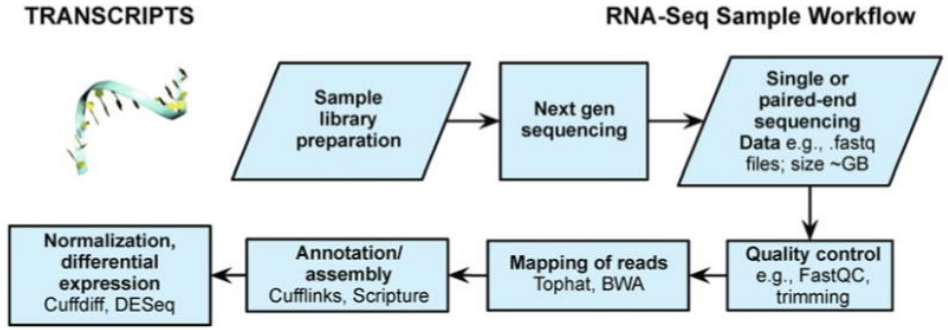
Translation

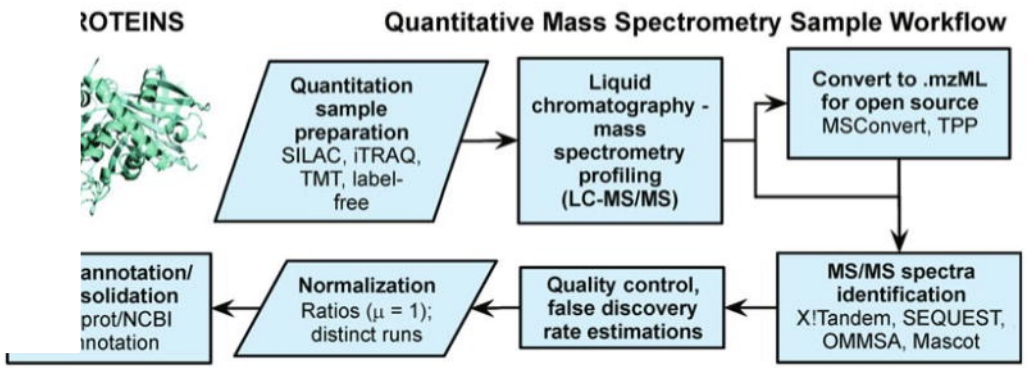
Protein

https://wildlifesnpits.wordpress.com/2015/11/18/transcriptomics-for-conservation/



Mias and Snyder, Quant Biol. (2013) 1(1): 71–90. doi: [10.1007/s40484-013-0005

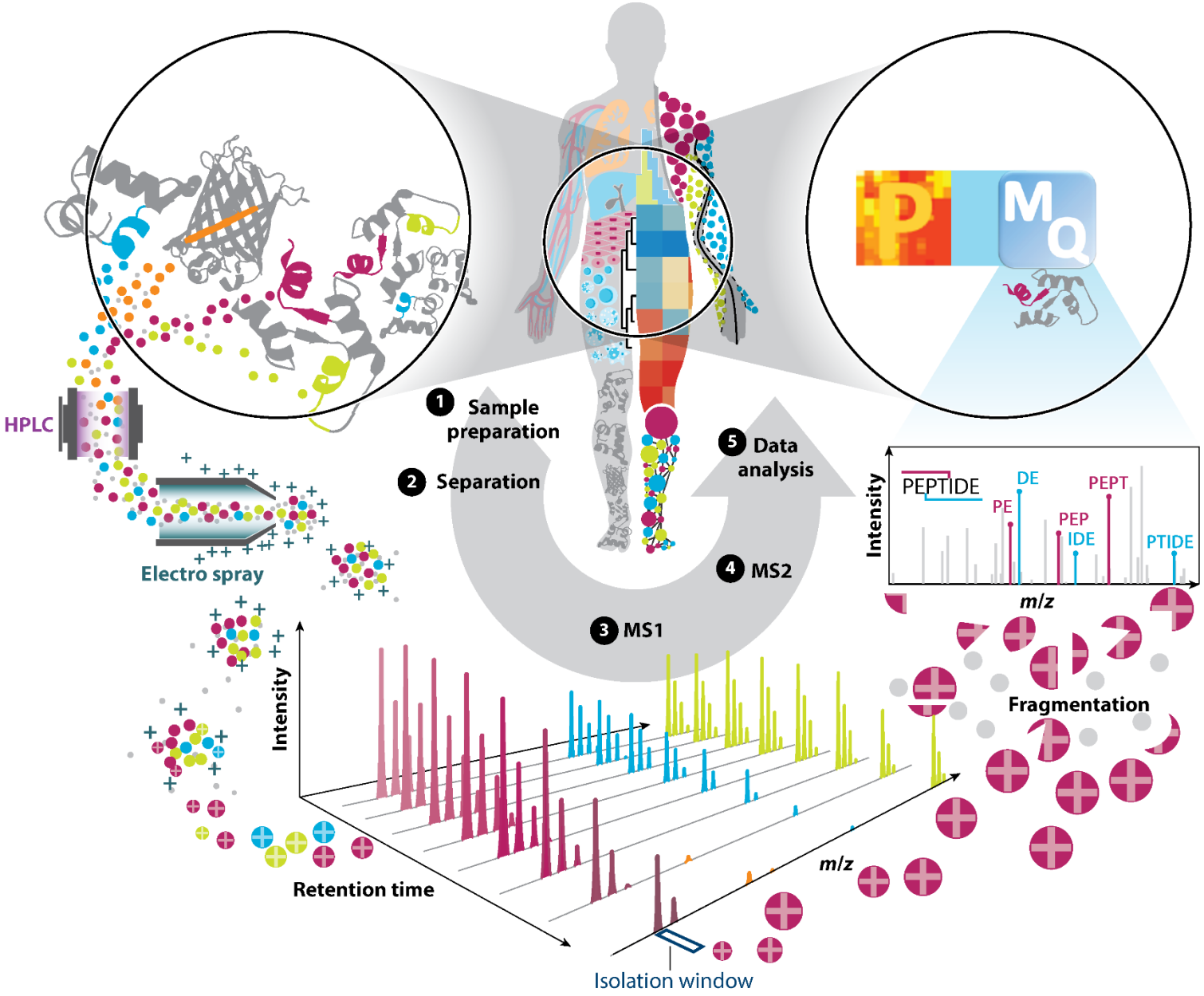




Enabling Proteo-Informatic Analysis via Galaxy-P Platform

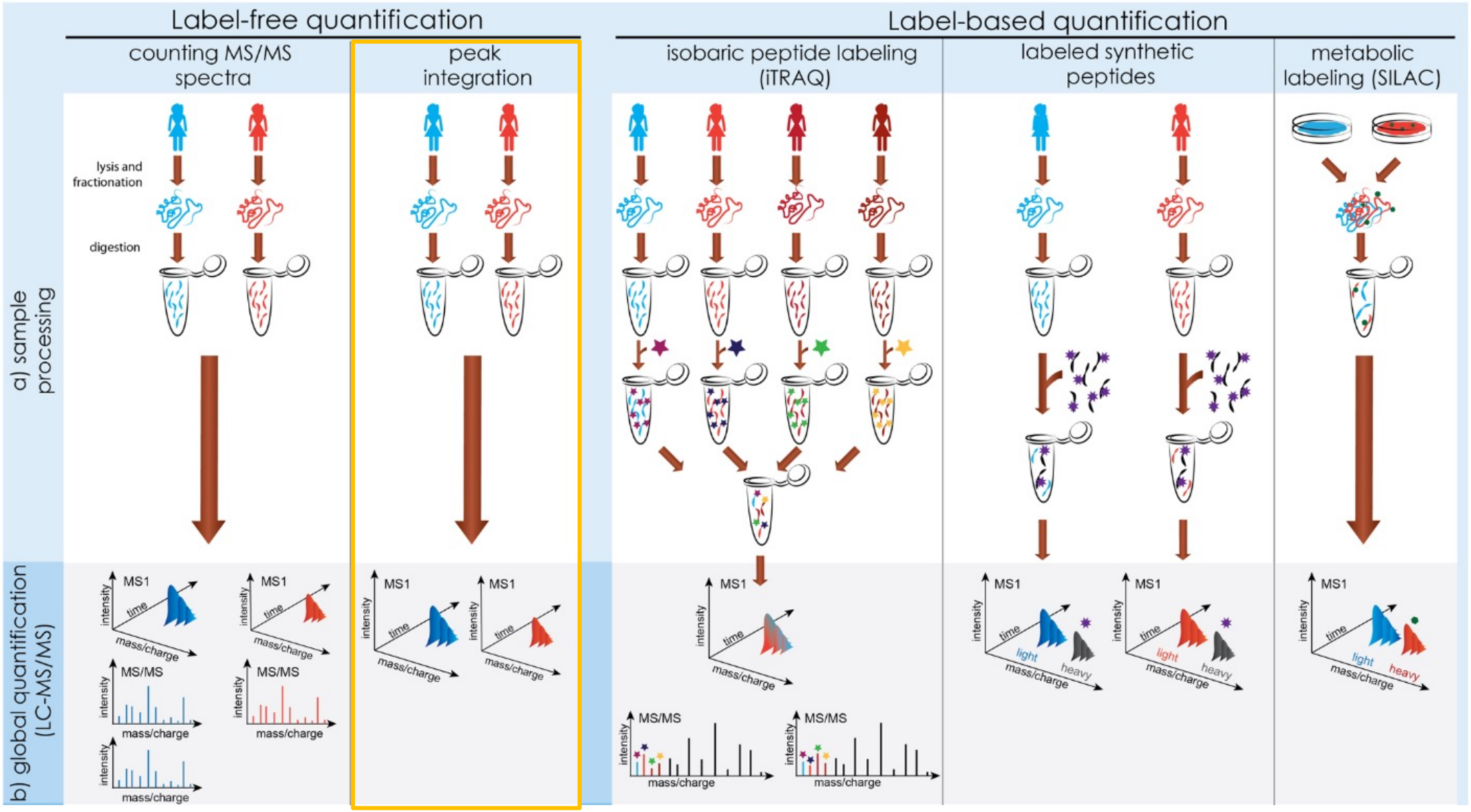
- *MULTIOMICS*
- ***MASS SPECTROMETRY***
- *GALAXY BIOINFORMATICS PLATFORM*
- *METAPROTEOMICS*

Mass Spectrometry and Proteomics



Sinitcyn P, et al. 2018.
Annu. Rev. Biomed. Data Sci. 1:207–34

Quantification methods in proteomics



Enabling Proteo-Informatic Analysis via Galaxy-P Platform

- *MULTIOMICS*
- *MASS SPECTROMETRY*
- ***GALAXY BIOINFORMATICS PLATFORM***
- *METAPROTEOMICS*

https://galaxyproject.org

Data Intensive *analysis* for everyone

Introduction to Galaxy PROJECT

- **Web-based** platform for computational biomedical research
- Developed at Penn State, Johns Hopkins, OHSU and Cleveland Clinic
- Community driven

- **Open source** under Academic Free License
- More than 10,000 citations

- More than 125 **public Galaxy servers**
- Usegalaxy.* instances:
  - Usegalaxy.org, usegalaxy.org.au, usegalaxy.eu

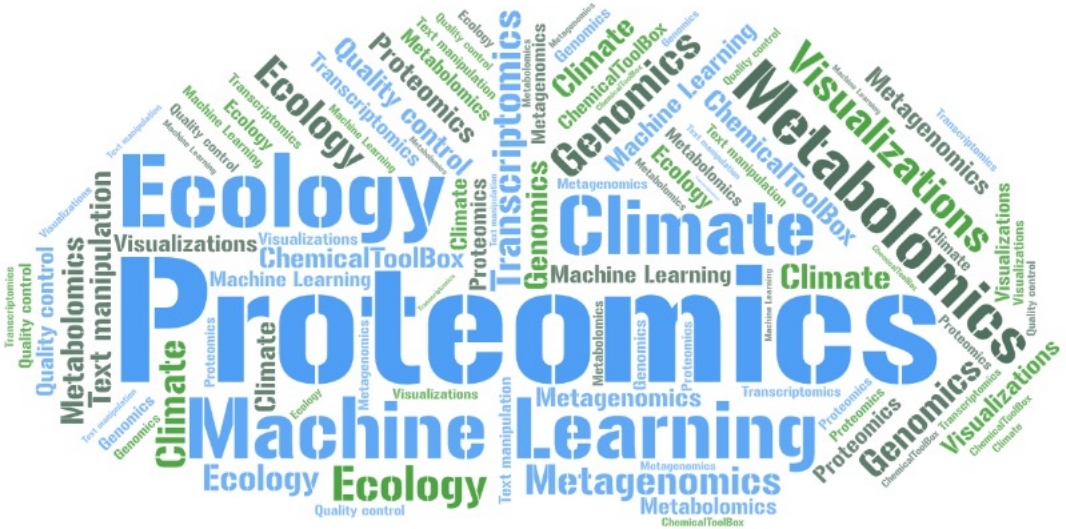
Core values

Galaxy PROJECT

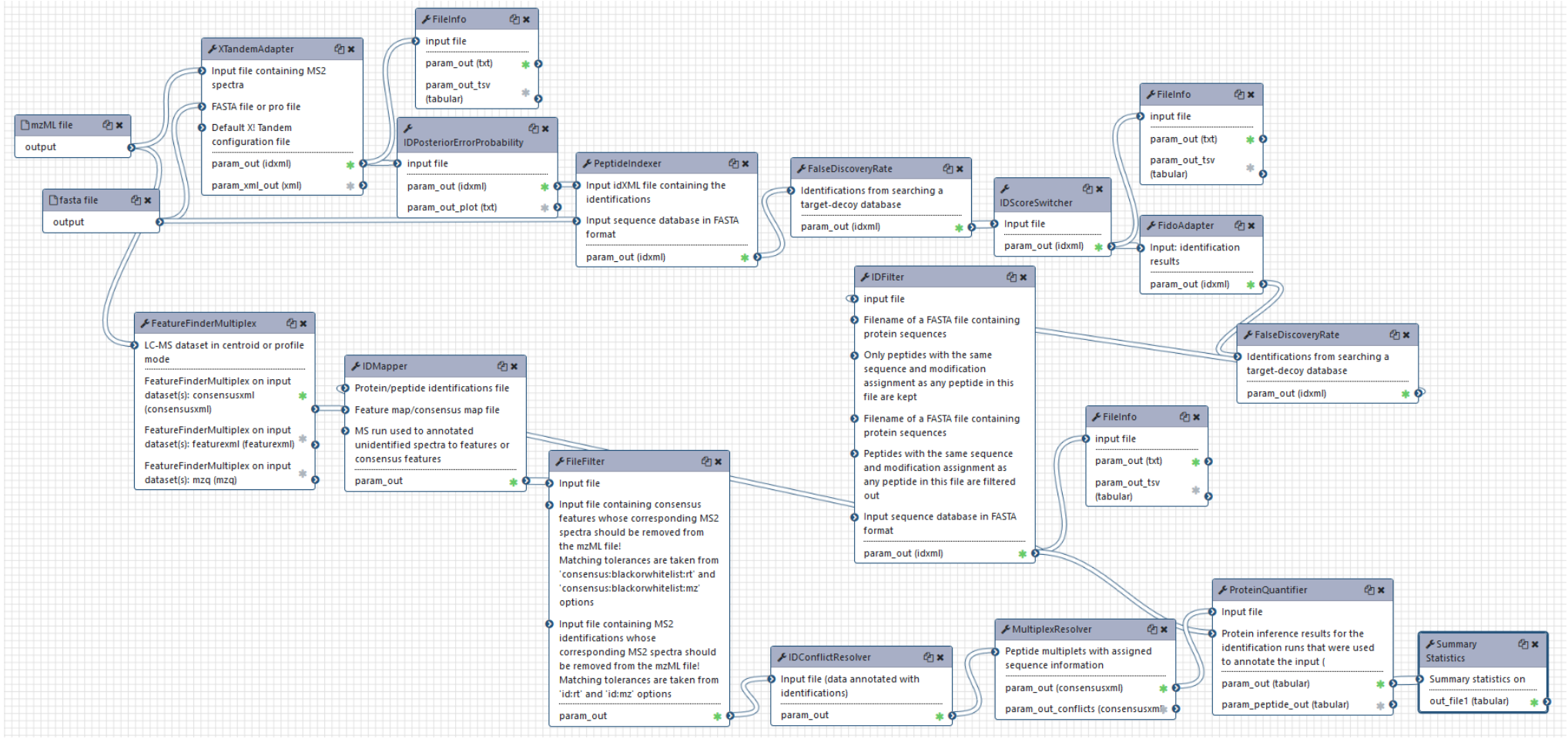
- **Accessibility**
  - Graphical user interface
  - Pre-installed tools
- **Reproducibility**
  - All information is captured in histories
  - Controlled tool versions
- **Transparency**
  - Sharing of histories and workflows

Available tools

- Galaxy tool shed: more than 7,000 tools
- Any open source tool can be integrated
- Areas:
  - Genomics
  - Transcriptomics
  - Proteomics
  - Metabolomics
  - ChemicalToolBox
  - Statistics
  - Ecology
  - Text manipulation
  - …



Tools can be connected into workflows



Hands-on training material

Welcome to Galaxy Training!

Collection of tutorials developed and maintained by the worldwide Galaxy community

Galaxy for Scientists

Topic	Tutorials
Introduction to Galaxy Analyses	10
Assembly	5
Climate	2
Computational chemistry	6
Ecology	5
Epigenetics	6
Genome Annotation	3
Imaging	3
Metabolomics	4
Metagenomics	6
Proteomics	15
Sequence analysis	2
Statistics and machine learning	8
Transcriptomics	23
Variant Analysis	8
Visualisation	2

Galaxy Tips & Tricks

Topic	Tutorials
User Interface and Data Manipulation	16

Galaxy for Developers and Admins

Topic	Tutorials
Galaxy Server administration	35
Development in Galaxy	13

How to contribute?

First off, thanks for taking the time to contribute!

You can report mistakes or errors, create more contents, etc. Whatever is your background, there is probably a way to do it: via the GitHub website, via command-line. If you feel it is too much, you can even write it with any text editor and contact us: we will work together to integrate it.

To get you started, check our dedicated tutorials or our Frequently Asked Questions

Galaxy for Contributors and Instructors

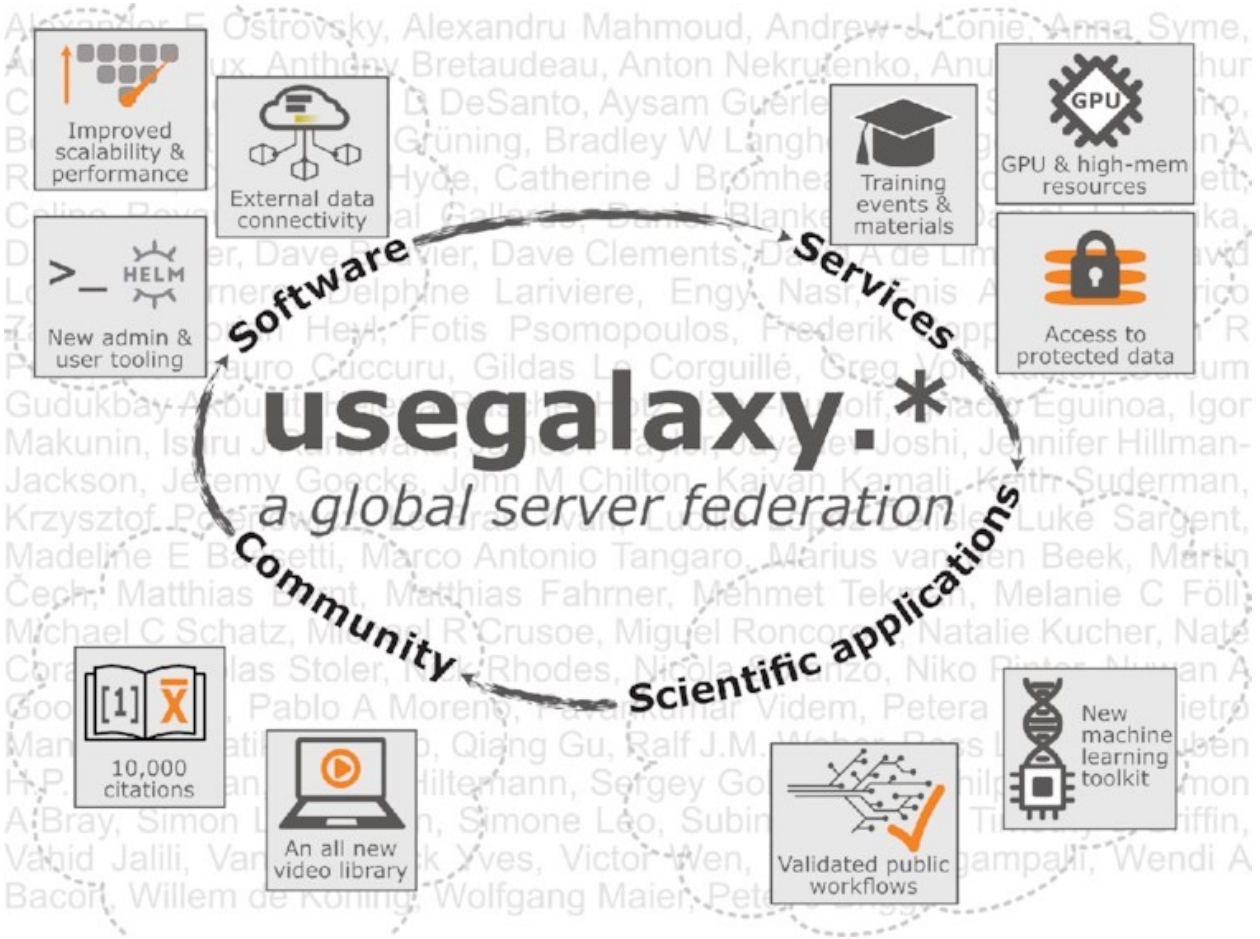
Topic	Tutorials
Contributing to the Galaxy Training Material	10
Teaching and Hosting Galaxy training	5

Galaxy Training Network

> 130 training materials

https://training.galaxyproject.org/training-material

Graphical Abstract The Galaxy Project is represented by a combination of software, managed services, application of the software ...



Nucleic Acids Res, gkac247, https://doi.org/10.1093/nar/gkac247

The content of this slide may be subject to copyright: please see the slide notes for details.

OXFORD UNIVERSITY PRESS

One solution: Galaxy

Goecks, J, Nekrutenko, A, Taylor, J and The Galaxy Team.. ***Genome Biol.*** 2010, **11**: R86.

A web-based, community developed bioinformatics workbench for integrating disparate software -- flexible

Geared towards use by bench scientists; many training resources available

Already home to genomic/transcriptomic tools

Provenance tracking, sharing and reproducibility

Amenable to other 'omic tools (e.g. Galaxy for proteomics project, Galaxy-P)

Working philosophy:

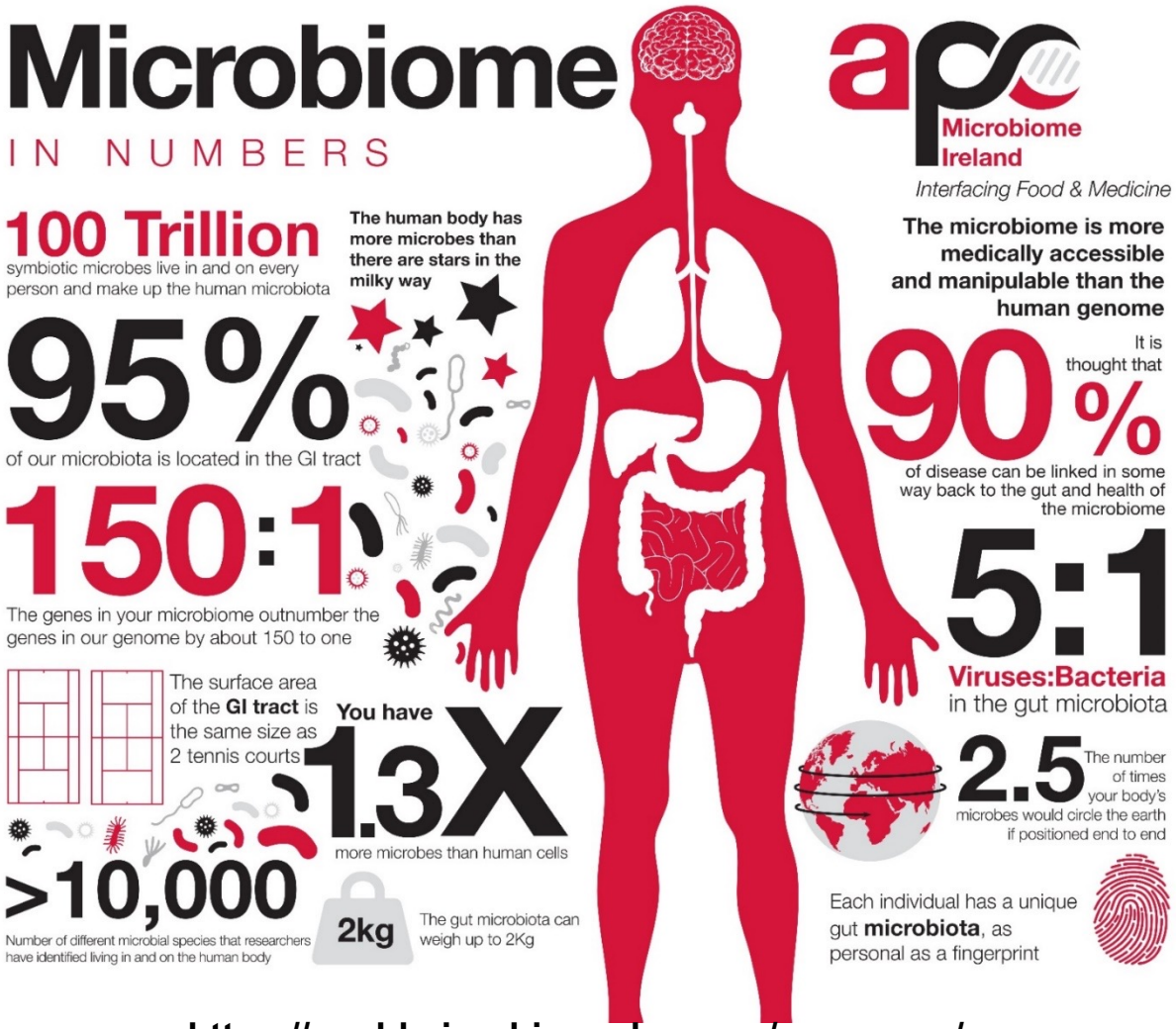
MINNEAPOLIS

COME TOGETHER #UseGalaxy2022

Enabling Proteo-Informatic Analysis via Galaxy-P Platform

- *MULTIOMICS*
- *MASS SPECTROMETRY*
- *GALAXY BIOINFORMATICS PLATFORM*
- ***METAPROTEOMICS***

MICROBIOME RESEARCH



https://worldmicrobiomeday.com/resources/

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

MICROBIOME

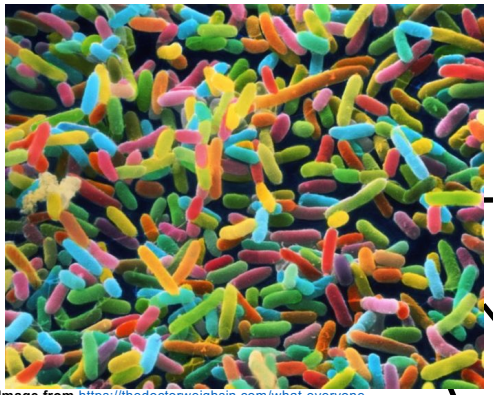
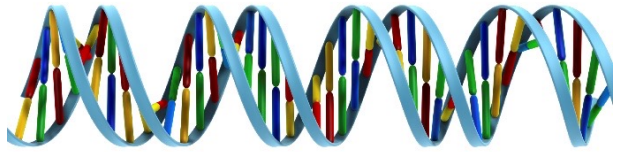


Image from https://thedoctorweighsin.com/what-everyone-should-know-about-the-infant-microbiome/

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



METAGENOMICS

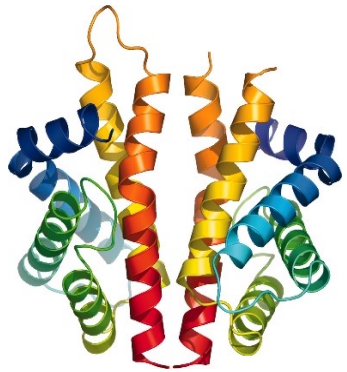
TAXONOMY

function

METATRANSCRIPTOMICS

TAXONOMY

function



METAPROTEOMICS

TAXONOMY

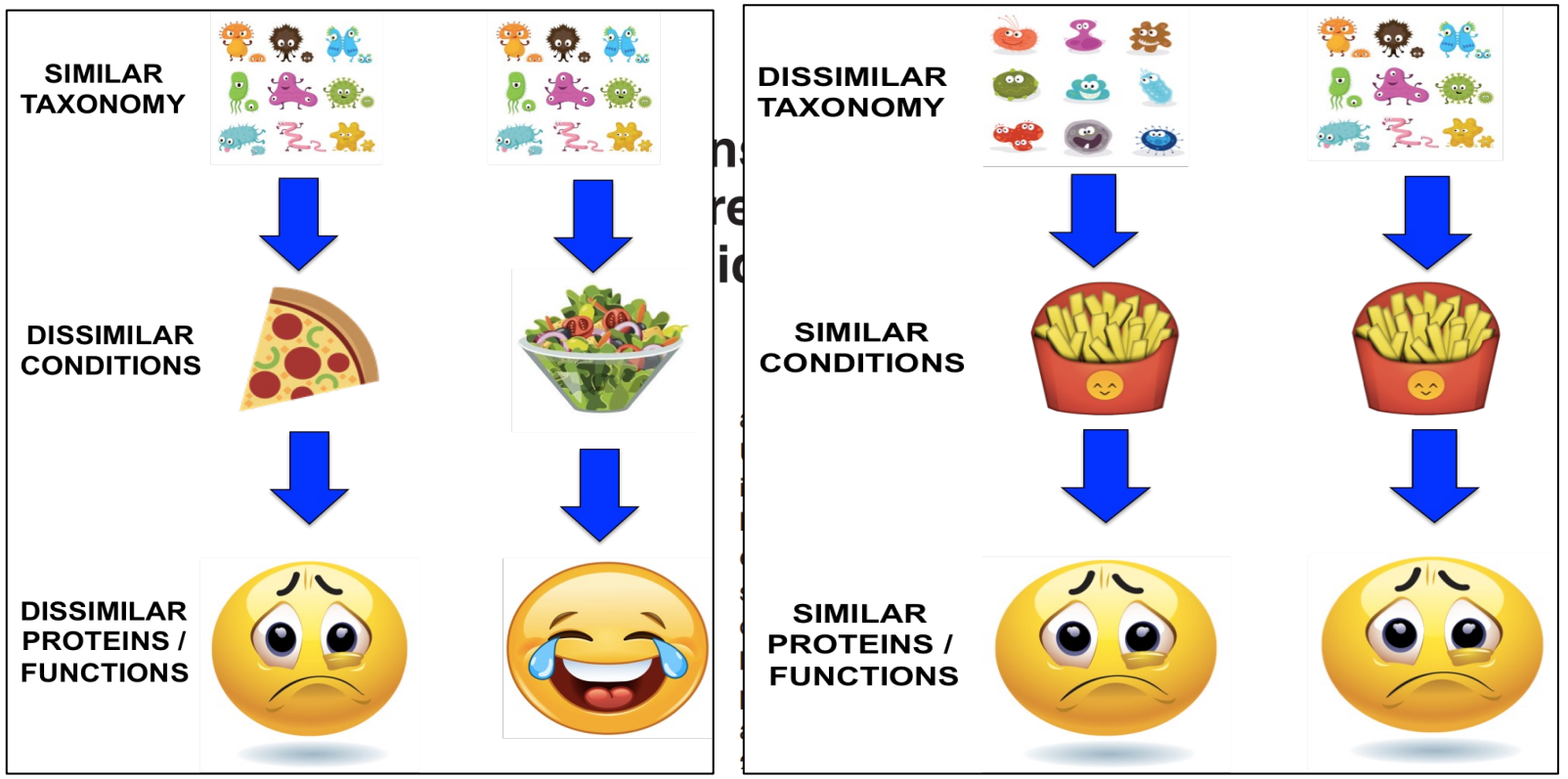
FUNCTION

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.

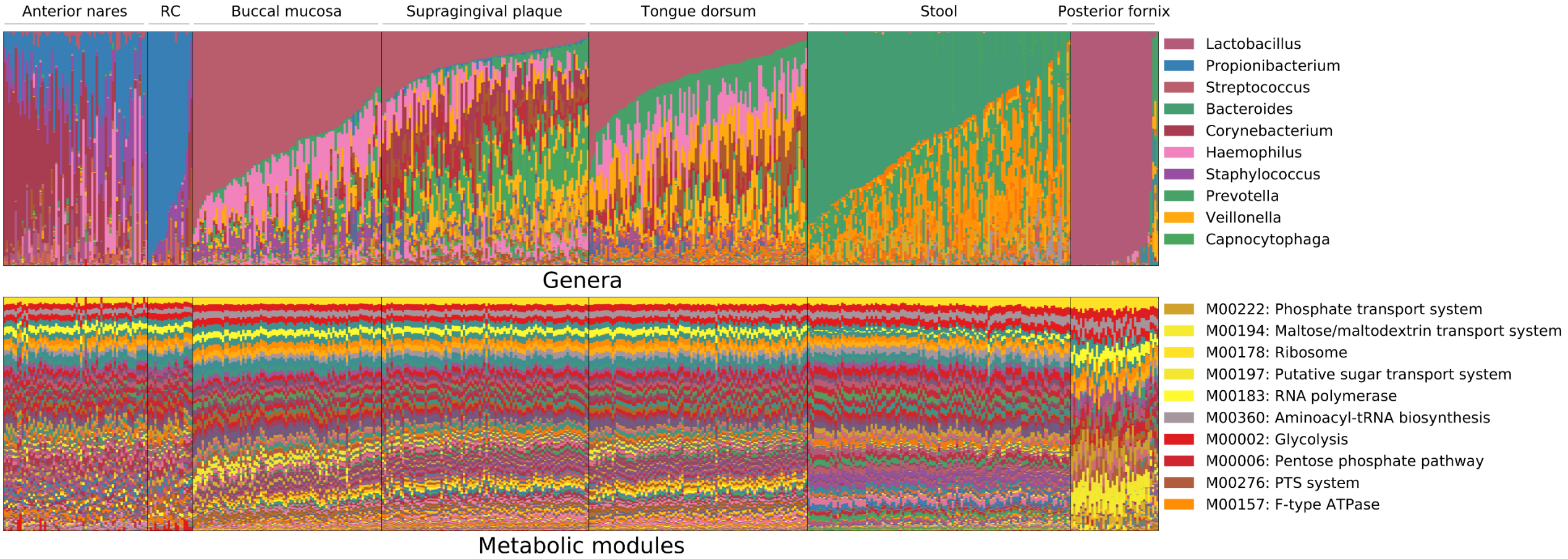


Bond and Wilmes 2015

"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

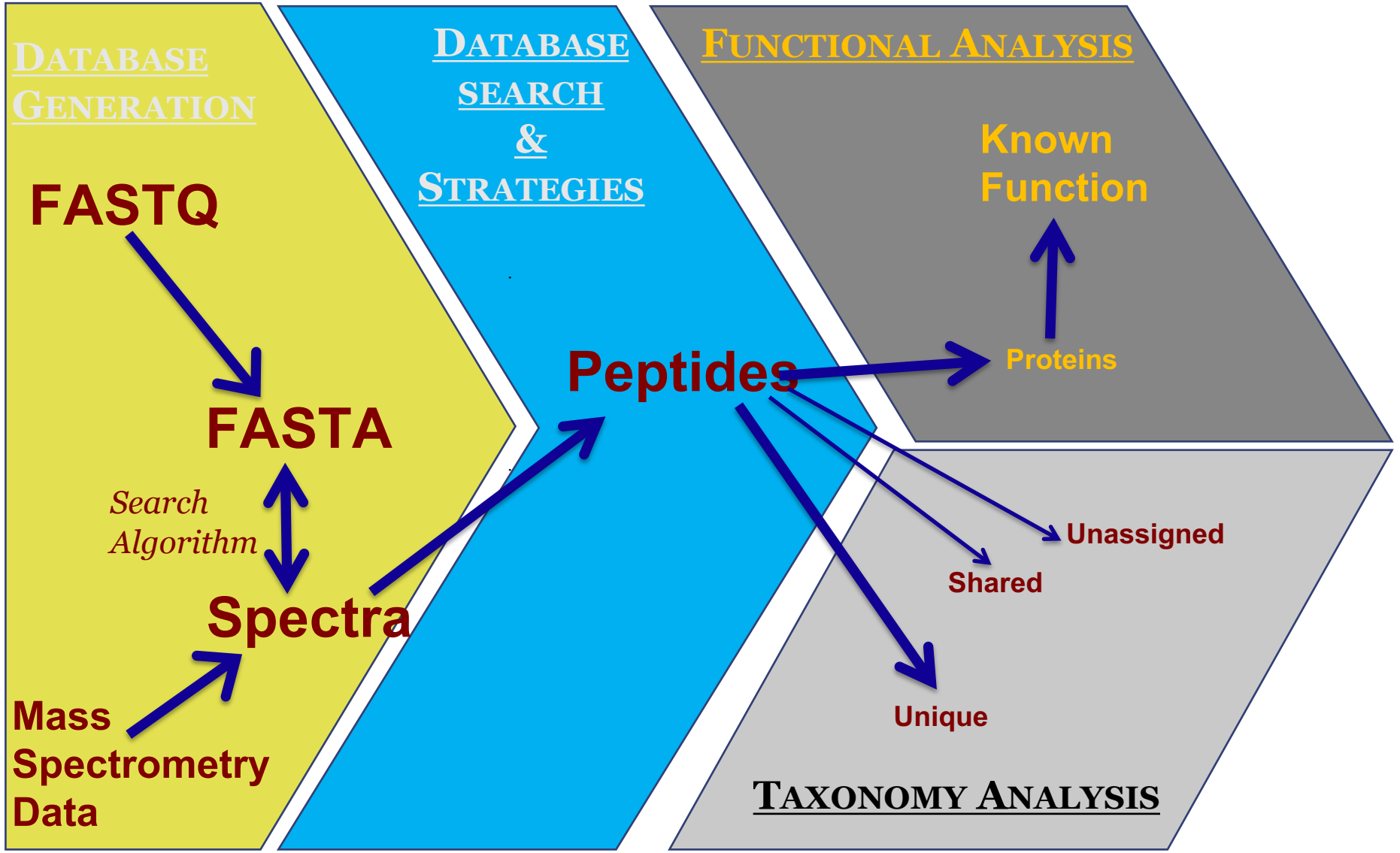


Human Microbiome Project Consortium (2012). Nature. 486(7402):207-214. doi: 10.1038/nature11234.

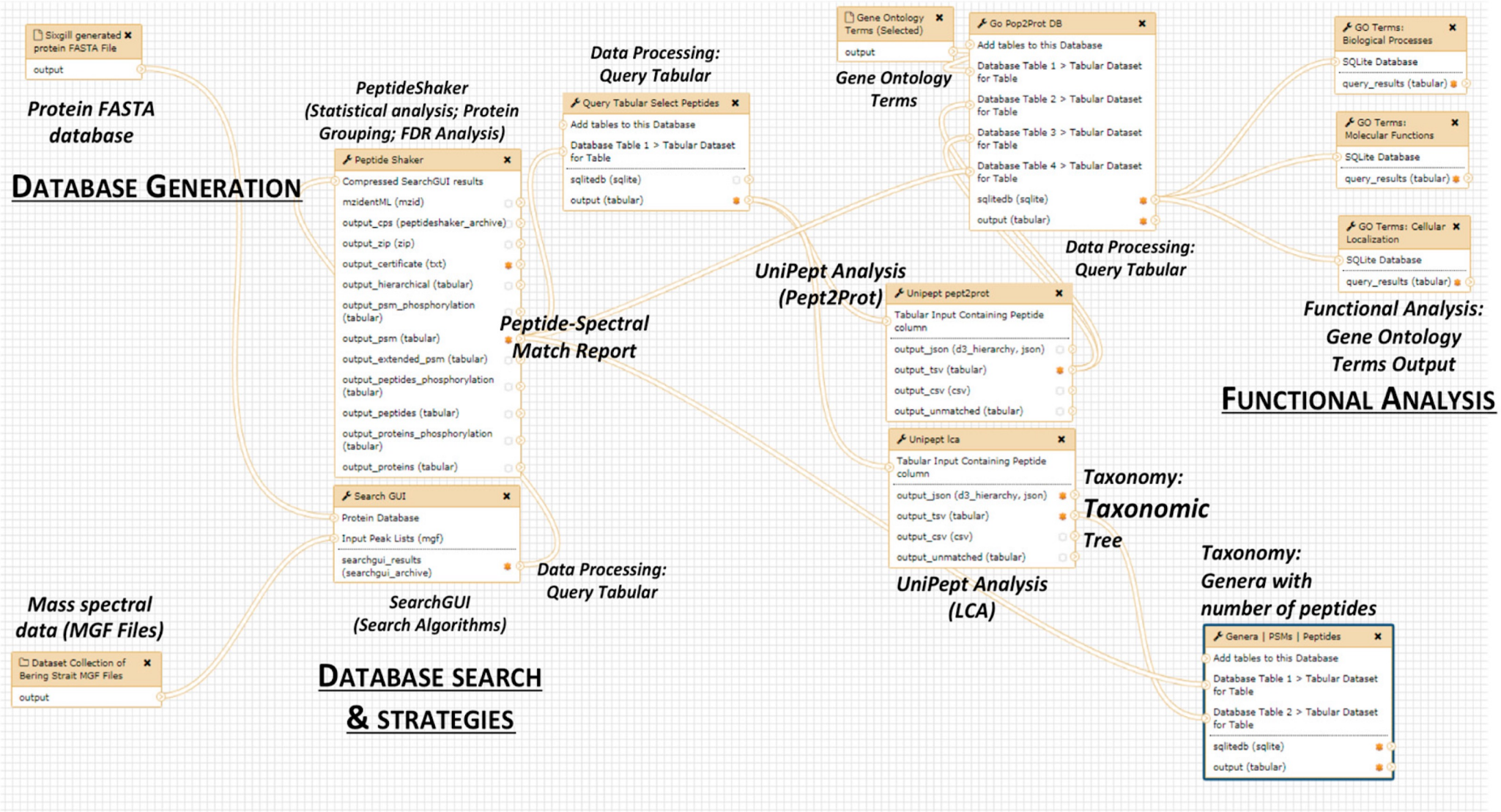
Metaproteomics Analytical Challenges

	Single-organism proteomics	Metaproteomics
Search Database Size	Small to medium size (10 K to 100K sequences)	LARGE (1 million and above)
	SINGLE + Contaminants	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



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

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

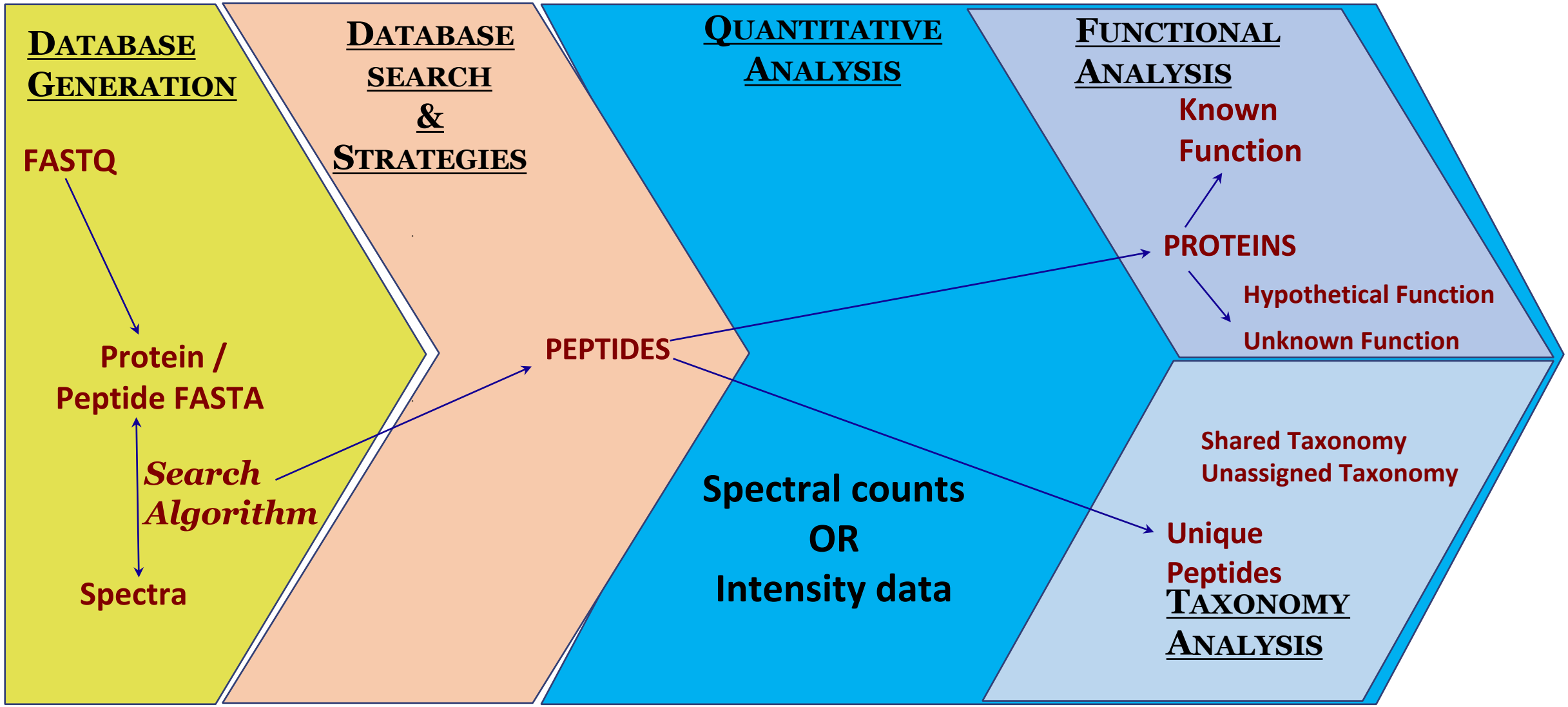
Jagtap PD et al (2018) **BAL Fluid Metaproteome** in Acute Respiratory Failure. Am J Respir Cell Mol Biol. 59(5):648-652. doi: 10.1165/rcmb.2018-0068LE.

ACS Publications Most Trusted. Most Cited. Most Read.

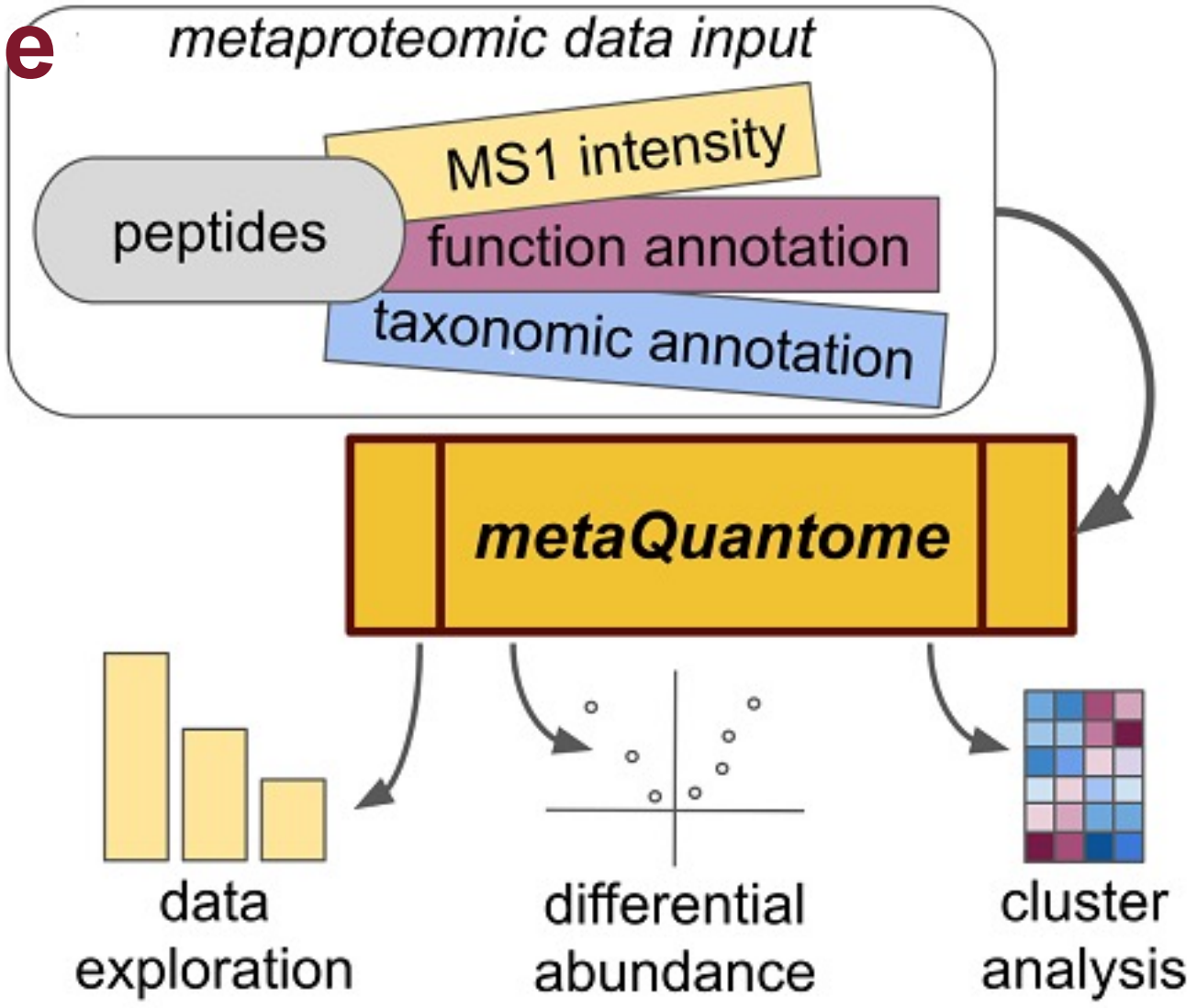
Journal of proteome research

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

Metaproteomics Workflow



metaQuantome

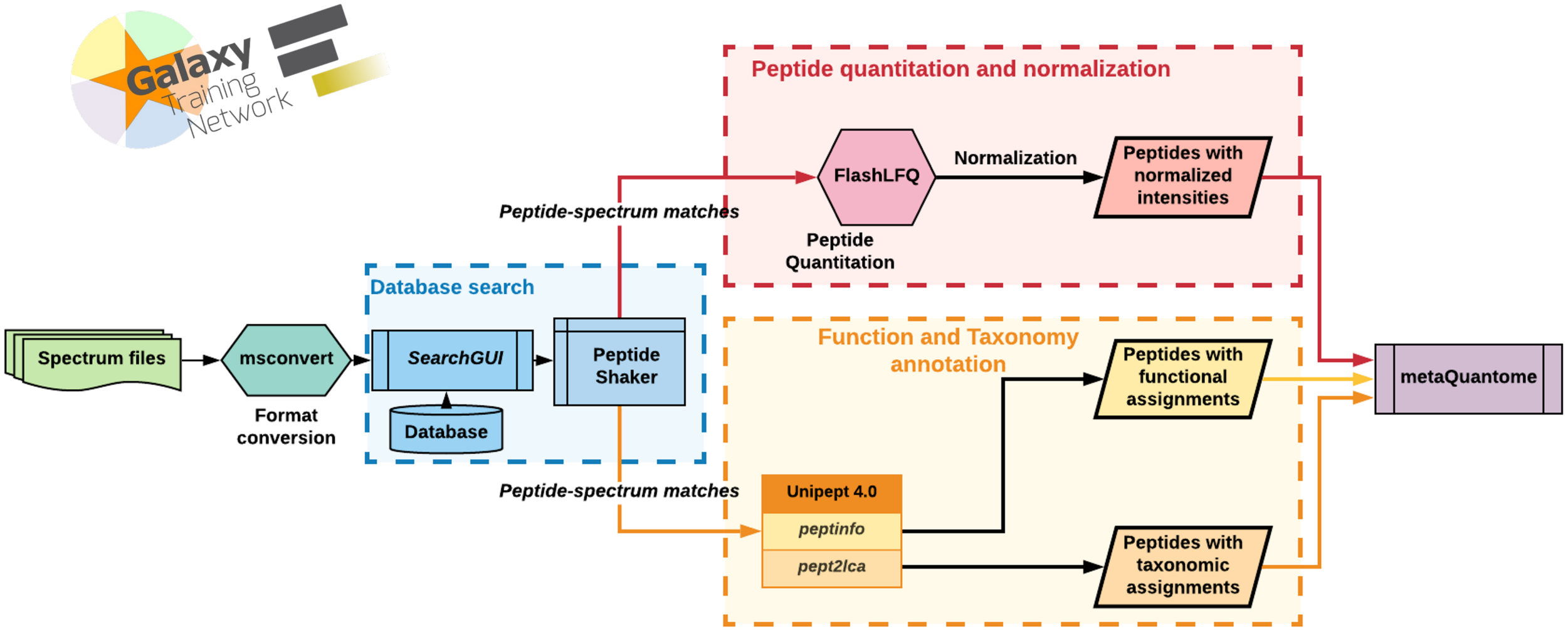


Caleb Easterly

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

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

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

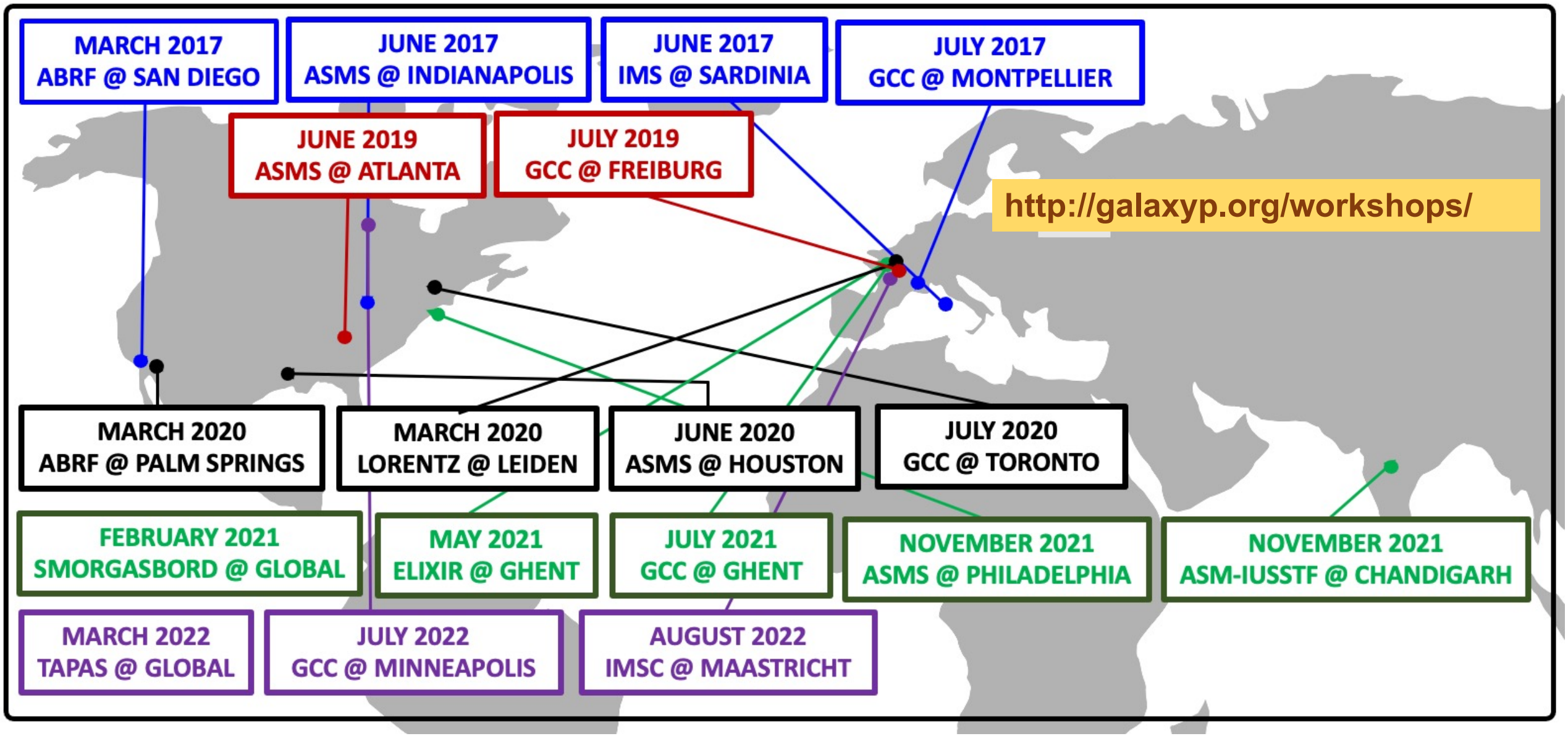
Galaxy Training Network

https://galaxyproject.org/events/gcc2021/training/

https://training.galaxyproject.org/training-material/topics/proteomics/

The Galaxy-P team seeks to continue integrating promising new software tools and workflows from leading laboratories into the Galaxy platform and encourages researchers to share their interests so that newer bioinformatics workflows can be made accessible via this resource.

Workshops


MARCH 2017
ABRF @ SAN DIEGO
JUNE 2017
ASMS @ INDIANAPOLIS
JUNE 2017
IMS @ SARDINIA
JULY 2017
GCC @ MONTPELLIER
JUNE 2019
ASMS @ ATLANTA
JULY 2019
GCC @ FREIBURG
http://galaxyp.org/workshops/
MARCH 2020
ABRF @ PALM SPRINGS
MARCH 2020
LORENTZ @ LEIDEN
JUNE 2020
ASMS @ HOUSTON
JULY 2020
GCC @ TORONTO
FEBRUARY 2021
SMORGASBORD @ GLOBAL
MAY 2021
ELIXIR @ GHENT
JULY 2021
GCC @ GHENT
NOVEMBER 2021
ASMS @ PHILADELPHIA
NOVEMBER 2021
ASM-IUSSTF @ CHANDIGARH
MARCH 2022
TAPAS @ GLOBAL
JULY 2022
GCC @ MINNEAPOLIS
AUGUST 2022
IMSC @ MAASTRICHT

ANALYSIS OF FUNCTIONS EXPRESSED BY MICROBIOMES

Online Workshop (15-24 November 2021) *Hosted by CSIR-IMTech, Chandigarh and Galaxy-P Team, Minneapolis*

ANALYSIS OF FUNCTIONS EXPRESSED BY MICROBIOMES

Online Workshop

15-24 November 2021

Funded by American Society of Microbiology (ASM) and Indo-United States Science and Technology Forum (IUSSTF).

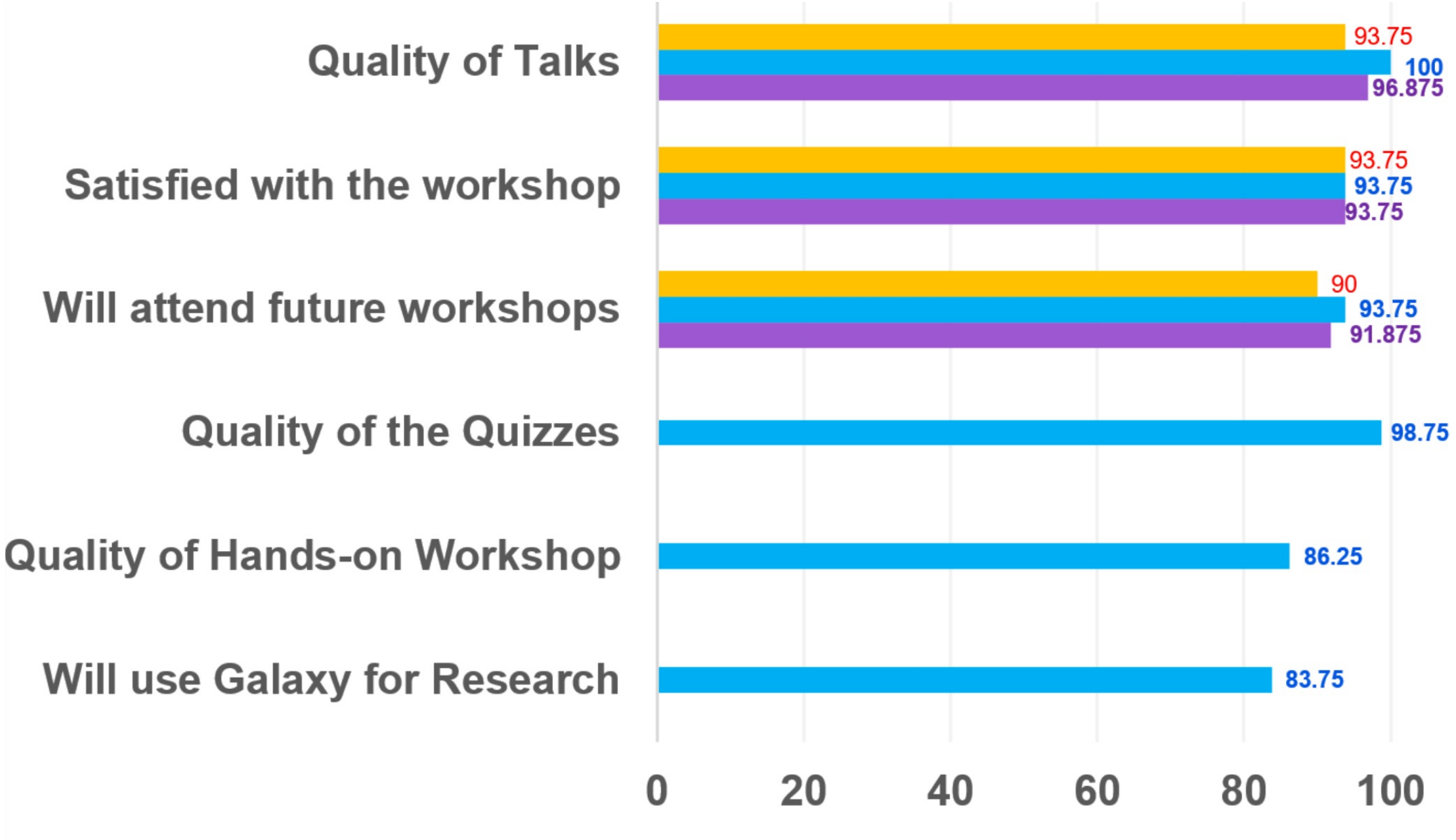
Hosted by CSIR-IMTech, Chandigarh and Galaxy-P Team, Minneapolis.

https://galaxyproject.org/events/2021-11-microbiomes/home/

https://gallantries.github.io/galaxy-workshop/events/functional-microbiome-2021/

ANALYSIS OF FUNCTIONS EXPRESSED BY MICROBIOMES

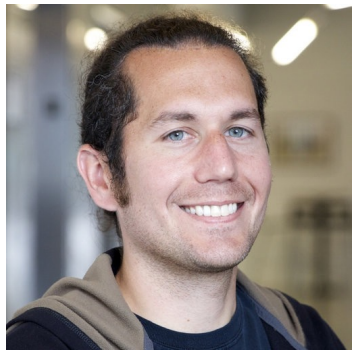
Online Workshop (15-24 November 2021) *Hosted by CSIR-IMTech, Chandigarh and Galaxy-P Team, Minneapolis*



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



Trans Proteomic Pipeline

Michael Hoopmann - *ISB, Seattle, WA*

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

October 2022

Galaxy

Melanie Foell - *Freiburg University, Germany*

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

November 2022

nextflow

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

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

Accessing Galaxy Tools and Workflows

Tools and Workflows also available on :
https://proteomics.usegalaxy.eu/

Galaxy Training Network:
https://training.galaxyproject.org/training-material/topics/proteomics

Galaxy Europe: https://proteomics.usegalaxy.eu/

Contact: http://galaxyp.org/contact/

TEN YEARS OF GALAXY-P !

PubMed

z.umn.edu/galaxypreferences

https://github.com/galaxyproteomics

GalaxyP

galaxyp.org

https://training.galaxyproject.org

Galaxy Training Network

http://galaxyp.org/workshops/

ACKNOWLEDGMENTS

Funding

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

Biologists / collaborators
Chris Wendt
Teresa Laguna
Maneesh Bhargava
Ryan Hunter
David Largaespada

Saskia Hiltemann

Thilo Muth
Robert Koch Institut, Germany

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

http://galaxyp.org/contact/

galaxyp.org

twitter.com/usegalaxy