Metaproteomics: Bioinformatic Workflow for Metaproteomic Analysis of Clinical Samples

Pratik Jagtap

University of Minnesota

Google Scholar: z.umn.edu/pjgs
PubMed: z.umn.edu/pjagtapreferences
Email: pjagtap@umn.edu
Twitter: pratikomics

Learn more at galaxyP.org
z.umn.edu/itcrgalaxyvideo
Metaproteomics: Bioinformatic Workflow for Metaproteomic Analysis of Clinical Samples

- **MICROBIOME RESEARCH**
- MASS SPECTROMETRY DATA ANALYSIS
- METAPROTEOMIC RESEARCH
- GALAXY BIOINFORMATICS PLATFORM
- QUANTITATIVE METAPROTEOMICS
- CLINICAL METAPROTEOMICS: COVID-19 PANDEMIC
- CLINICAL METAPROTEOMICS: CYSTIC FIBROSIS
- METAPROTEOMICS EDUCATION
Microbiome Research

Microbiome in Numbers

100 Trillion
symbolic microbes live in and on every person and make up the human microbiota

95%
of our microbiota is located in the GI tract

150:1
The human body has more microbes than there are stars in the milky way

1.3X
The surface area of the GI tract is the same size as 2 tennis courts

You have

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

>10,000
Number of different microbial species that scientists have identified so far in and on the human body

2kg
The gut microbiota can weigh up to 2kg

5:1
You have viruses:bacteria in your gut microbiota

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

90%
of disease can be linked in some way back to the gut, and health of the microbiome

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
Potential to unravel the mechanistic details of microbial interactions with host / environment by analyzing the functional dynamics of the microbiome.
MICROBIAL TAXA VARY WHILE METABOLIC PATHWAYS REMAIN STABLE WITHIN A HEALTHY POPULATION

Metaproteomics: Bioinformatic Workflow for Metaproteomic Analysis of Clinical Samples

- MICROBIOME RESEARCH
- **MASS SPECTROMETRY DATA ANALYSIS**
  - METAPROTEOMICS RESEARCH
  - GALAXY BIOINFORMATICS PLATFORM
  - QUANTITATIVE METAPROTEOMICS
  - CLINICAL METAPROTEOMICS: COVID-19 PANDEMIC
  - CLINICAL METAPROTEOMICS: CYSTIC FIBROSIS
  - METAPROTEOMICS EDUCATION
Mass Spectrometry and Proteomics

1. Sample preparation
2. Separation
3. MS1
4. MS2
5. Data analysis
6. Intensity
7. Retention time
8. Isolation window
9. m/z
10. Fragmentation

HPLC
Electro spray
PEP
DE
PEPT
IDE
PTIDE

Metaproteomics: Bioinformatic Workflow for Metaproteomic Analysis of Clinical Samples

- Microbiome Research
- Mass Spectrometry Data Analysis
- **Metaproteomics Research**
- Galaxy Bioinformatics Platform
- Clinical Metaproteomics: COVID-19 Pandemic
- Clinical Metaproteomics: Cystic Fibrosis
- Metaproteomics Education
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.”

**Bond and Wilmes 2015**

**Metaproteomics**

Bond and Wilmes 2004

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


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.”

**Metaproteomics Analytical Challenges**

**Single-Organism Proteomics**
- **Search Database Size**: Small to medium size (10 K to 100K sequences)
- **Complexity**: Single + Contaminants

**Metaproteomics**
- **Search Database Size**: Large (1 million and above)
- **Complexity**: 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 Workflow**

**Database Generation**
- FASTQ
- FASTA
- Mass Spectrometry Data

**Database Search & Strategies**
- Spectra
- Search Algorithm

**Functional Analysis**
- Peptides
- Known Function
- Proteins

**Taxonomy Analysis**
- Unique
- Shared
- Unassigned
Metaproteomics: Bioinformatic Workflow for Metaproteomic Analysis of Clinical Samples

- MICROBIOME RESEARCH
- MASS SPECTROMETRY DATA ANALYSIS
- METAPROTEOMICS RESEARCH

**GALAXY BIOINFORMATICS PLATFORM**

- CLINICAL METAPROTEOMICS: COVID-19 PANDEMIC
- CLINICAL METAPROTEOMICS: CYSTIC FIBROSIS
- METAPROTEOMICS EDUCATION
Software tools can be used in a sequential manner to generate analytical workflows that can be reused, shared and creatively modified.
The Galaxy Interface

Main viewing window (workflow development, results visualization etc)
**Workflows:** Multiple software tools used in a sequential manner for an analysis

**History:** recorded, complete analysis (workflow + all data and results)


Metaproteomics: Bioinformatic Workflow for Metaproteomic Analysis of Clinical Samples

- MICROBIOME RESEARCH
- MASS SPECTROMETRY DATA ANALYSIS
- METAPROTEOMICS RESEARCH
- GALAXY BIOINFORMATICS PLATFORM

- QUANTITATIVE METAPROTEOMICS
- CLINICAL METAPROTEOMICS: COVID-19 PANDEMIC
- CLINICAL METAPROTEOMICS: CYSTIC FIBROSIS
- METAPROTEOMICS EDUCATION
**DATABASE GENERATION**

FASTQ

Protein / Peptide FASTA

Search Algorithm

Spectra

**DATABASE SEARCH & STRATEGIES**

**QUANTITATIVE ANALYSIS**

Spectral counts OR Intensity data

**FUNCTIONAL ANALYSIS**

Known Function

PROTEINS

Hypothetical Function

Unknown Function

Shared Taxonomy

Unassigned Taxonomy

Unique Peptides

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

Case Study: Cellulose Degradation in a BioGas Reactor

Biogas-plant (60°C) Fredrikstad, Norway
Lab-scale reactor (55°C)
Anaerobic bottles (65°C)

Food waste
Manure

Serial dilution

Cellulose

Food waste
Manure

0h 8h 13h 18h 23h 28h 33h 38h 43h

Magnus Arntzen
NMBU, Norway
MetaQuantome Workflow
metaQuantome ANALYSIS: Principal Component Analysis

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

Taxonomic Contribution to a Function

GLYCOSIDE HYDROLASE

Gene Ontology Terms
- Cellulose binding
- Cellulase activity
- Cellulose 1,4-beta-cellobiosidase activity

Functions Associated With Cellulose Degradation in Hungateiclostridium

Taxa associated with Glycosyl hydrolases and transferases
Metaproteomics: Bioinformatic Workflow for Metaproteomic Analysis of Clinical Samples

- MICROBIOME RESEARCH
- MASS SPECTROMETRY DATA ANALYSIS
- METAPROTEOMICS RESEARCH
- GALAXY BIOINFORMATICS PLATFORM

**CLINICAL METAPROTEOMICS**

- CLINICAL METAPROTEOMICS: CYSTIC FIBROSIS
- METAPROTEOMICS EDUCATION
CLINICAL METAPROTEOMIC STUDIES DETECT LIMITED NUMBER OF MICROBIAL PEPTIDES

![Bar chart showing distinct peptides]

- **Microbial**: 2292
- **Human**: 85713

**Cystic Fibrosis dataset**


Metaproteomics: Bioinformatic Workflow for Metaproteomic Analysis of Clinical Samples

- MICROBIOME RESEARCH
- MASS SPECTROMETRY DATA ANALYSIS
- METAPROTEOMICS RESEARCH
- GALAXY BIOINFORMATICS PLATFORM
- CLINICAL METAPROTEOMICS: COVID-19 PANDEMIC
- CLINICAL METAPROTEOMICS: CYSTIC FIBROSIS
- METAPROTEOMICS EDUCATION
Sample processing

Data Acquisition

Nasal Swabs from COVID-19 patients

COVID-19 PANDEMIC DATASETS

bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

bioRxiv posts many COVID-19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

New Results

Metaproteomic analysis of nasopharyngeal swab samples to identify microbial peptides and potential co-infection status in COVID-19 patients

Surbhi Bihani, Aryan Gupta, Subrina Mehta, Andrew Rajczewski, James Johnson, Dhanush Borishetty, Timothy J. Griffin, Sanjeeva Srivastava, Pratik Jagtap
doi: https://doi.org/10.1101/2023.01.31.525328

Surbhi Bihani
COVID-19 PANDEMIC DATASETS

Metaproteomic analysis of nasopharyngeal swab samples to identify microbial peptides and potential co-infection status in COVID-19 patients

Surbhi Bhani, Aryan Gupta, Subha Mehta, Andrew Ruzekowski, James Johnson, Dhanush Borishetty, Timothy J. Griffin, Sanjeet Sivastava, Pratik Jagtap

doi: https://doi.org/10.1101/2023.01.31.525328
MICROBIAL PROTEINS FROM COVID-19 PATIENTS

Severe Negative

Log$_2$(FC)

PC 2 (11.1%)

PC 1 (49%)

Log$_{10}$(p)

Log$_2$(FC)

Aryan Gupta

**Clostridium**
tRNA pseudouridine synthase B

**Pseudomonas fluorescens**
T2SSF domain-containing protein

**Penicillium steckii**
DNA repair and recombination protein

**Fusarium vanettii**
MARVEL domain-containing protein

**Enterobacter**
Thiamine biosynthesis protein ThiH

Log$_2$(Abundance)
**Targeted Analysis of Potential Pathogens in COVID Patients**

**Streptococcus parasanguinis**
Dominant isolate of dental plaque
Opportunistic pathogen associated with subacute endocarditis

**Burkholderia cenocepacia**
Opportunistic pathogen
Cause systemic infections in immunocompromised individuals including cystic fibrosis patients.

**Cryptococcus floricola**
Infect immunocompromised hosts
Infection initiates in the lungs.
Metaproteomic analysis of nasopharyngeal swab samples to identify microbial peptides and potential co-infection status in COVID-19 patients

Surbhi Bihani, Aryan Gupta, Subina Mehta, Andrew Rajczewski, James Johnson, Dhanush Borishetty, Timothy J. Griffin, Sanjeeva Srivastava, Pratik Jagtap

doii: https://doi.org/10.1101/2023.01.31.525328
Metaproteomics: Bioinformatic Workflow for Metaproteomic Analysis of Clinical Samples

- Microbiome Research
- Mass Spectrometry Data Analysis
- Metaproteomics Research
- Galaxy Bioinformatics Platform
- Clinical Metaproteomics: COVID-19 Pandemic
- Clinical Metaproteomics: Cystic Fibrosis
- Metaproteomics Education
Cystic Fibrosis and the Microbiome

- Life-shortening Mendelian disease.
- Mutations in the CF transmembrane conductance regulator (CFTR) gene, which encodes an epithelial anion channel.
- Abnormal anion transport across epithelia of secretory glands including lung.
- Thick mucus which predisposes patients to chronic bacterial infections and airway inflammation.
- Progressive and irreversible airway damage.
- Reduced quality and length of life in CF.

- Understanding CF respiratory microbiome (‘song along with the singers’) may help in improved therapeutics.
- Understanding of host–microbiome interactions may help in defining relationships between microbiome, disease status, and treatment response.
CYSTIC FIBROSIS DATASETS

BALF from Cystic Fibrosis or Disease Control Samples

Sample Collection → Sample processing

Data Acquisition

RAW

Monica E. Kruk
Bioinformatic Workflow

16S rRNA Data → Comprehensive Protein Database → MetaNovo → Reduced Database → MS Datasets

Identification of Microbial Peptides

Targeted Analysis of Microbial Peptides

Data Analysis

Protein Database of PepQuery verified organisms

Unipept

PepQuery

Uniprot

FragPipe

MaxQuant

MQ
MICROBIAL PEPTIDES FROM CYSTIC FIBROSIS PATIENTS

**PF00069**

**TAXONOMY:** Pseudomonas

**FUNCTION:** Type II secretion system protein

**DWLDSLQR**

**TAXONOMY:** Streptococcus agalactiae

**FUNCTION:** DNA recombination enzyme

**MKIGNLGGAYR**

**TAXONOMY:** Bacteria (UniPept)

**FUNCTION:** Membrane kinase activity

**VLGARGHR**

**FUNCTION:** Nucleoside triphosphate hydrolase

**AALGAYDLR**

**TAXONOMY:** Unassigned (UniPept)

**FUNCTION:** Uncharacterized protein

**ATVEELHLEGI**

**TAXONOMY:** Unassigned (UniPept)

**FUNCTION:** Uncharacterized protein
MICROBIAL PEPTIDES FROM CYSTIC FIBROSIS PATIENTS

A)

B)
**Targeted Assay for Host-Microbial Protein Dynamics in Clinical CF Samples**

- Lipoprotein particle remodeling
- Cholesterol transport
- Phospholipid efflux.

- Inflammatory response
- Leukocyte migration
- Neutrophil degranulation
- Cytokine production
- Defense response to bacteria

Gene expressions for protein remodelling in CF and DC samples:

- **APOC2**: CF, DC
- **APOE**: CF, DC
- **S100A9**: CF, DC
- **MPO**: CF, DC
- **FGR**: CF, DC
- **MMP8**: CF, DC
- **S100A12**: CF, DC
- **S100A8**: CF, DC
- **ELANE**: CF, DC
- **PRSS57**: CF, DC
- **PPP1CB**: CF, DC
Clinical Metaproteomics

• We have developed a MS-based Galaxy-driven bioinformatics workflow for processing of microbial and host proteins, generating verified microbial peptide candidates suitable for targeted analysis within individual patient samples.

• We have utilized this workflow to detect and validate microbial peptides during co-infection during the COVID-19 pandemic.

• The workflow was also used to generate a promising microbial and host peptide panel for application to CF disease progression studies by comparing it with disease control (DC).

• We intend to use this workflow in other clinical metaproteomic studies to detect differentially expressed host and microbial proteins in disease state.
Metaproteomics: Bioinformatic Workflow for Metaproteomic Analysis of Clinical Samples

- MICROBIOME RESEARCH
- MASS SPECTROMETRY DATA ANALYSIS
- METAPROTEOMICS RESEARCH
- GALAXY BIOINFORMATICS PLATFORM
- CLINICAL METAPROTEOMICS: COVID-19 PANDEMIC
- CLINICAL METAPROTEOMICS: CYSTIC FIBROSIS
- METAPROTEOMICS EDUCATION
METAPROTEOMICS: STRENGTHS & CHALLENGES

FUNCTIONAL COMPOSITION

HOST-MICROBIOME INTERACTION

QUANTIFYING BIOMASS CONTRIBUTIONS

PROTEIN EXTRACTION

VARIABILITY & ABUNDANCE

LARGE DATABASES FALSE POSITIVES
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.
EDUCATION PORTAL @ METAPROTEOMICS INITIATIVE

Want to stay up to date about the metaproteomics field and the Initiative? Become a member via www.metaproteomics.org!

@MetaP_Init
info@metaproteomics.org

https://metaproteomics.org/education/
ANALYSIS OF FUNCTIONS EXPRESSED BY MICROBIOMES

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

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/