Houston, We Have a Microbiome Problem
(...and how the Metaproteomics Initiative aims to solve it)

June 7, 2023
General Assembly B
Workshop structure

5:45 – Welcome and brief introduction to the workshop

5:55 PM: Tim Van Den Bossche (Ghent University) – Introduction to the Metaproteomics Initiative.

6:05 PM: Bob Hettich (Oak Ridge National Lab) - updates from recent Metaproteomics Initiative meeting and the rationale for the Critical Assessment of MetaProteome Investigation (CAMPI) initiative

6:20 PM: Pratik Jagtap (University of Minnesota) and Samantha Peters (Oak Ridge National Lab) – description of CAMPI-3 goals, logistics and how you can be involved!

6:40 PM: Questions and discussion

Invitation to attend the metaproteomics biinformatics hub tomorrow at 10:30-12:30 in the INFORMATICS HUB in Poster Hall.
Metaproteomics: Promise and challenge

- Characterization of functional state (beyond genomic functional potential)
- Taxa-function relationships
- Host-microbiome interactions
- Guide for small molecule identification
- Annotating genes/proteins of unknown function

- Complexity!
- Sample preparation and limited abundance
- Taxonomic assignment
- Proteins of unknown function

https://www.pnnl.gov/projects/soil-microbiome/research

Microbial community

Microbes contributing to the health of a host

Microbes contributing to environmental systems
Objectives of workshop

• Emphasize the value and success of metaproteomics

• Promote resources available to those engaging in metaproteomics

• Outline challenges and solutions across critical aspects of metaproteomics:
  *Introduction to CAMPI Studies and launch of CAMPI3 Study*

• Invite you to join the community (help wanted!)

• Answer your questions!
Why an Initiative?

- Need for better **communication** and **collaboration**
  - In the next years, major new analytical developments are likely to facilitate unprecedented insights into microbiome function and dynamics
  - Many metaproteomics researchers are working solitary in single-species proteomics labs, or in metagenomics/metatranscriptomics labs
We organized the first, community-driven multi-lab benchmark study in the field

Critical Assessment of MetaProteome Investigation (CAMPI): a multi-laboratory comparison of established workflows


Nature Communications 12, Article number: 7305 (2021) | Cite this article

CAMPI 2 has been launched, CAMPI 3 will be launched very, very soon...
We also organize every 1 - 1.5 years the International Metaproteomics Symposium

- Magdeburg, Germany (2016)
- Alghero, Italy (2017)
- Leipzig, Germany (2018)
- Online symposium (2021)
- Luxembourg, Luxembourg (2021)
- Avignon, France (2023)
January 2025 (tbd)
Winter meeting at a Ski resort
Why an Initiative?

• Need for better **communication** and **collaboration**
  • In the next years, major new analytical developments are likely to facilitate unprecedented insights into microbiome function and dynamics
  • Many metaproteomics researchers are working solitary in single-species proteomics labs, or in metagenomics/metatranscriptomics labs
  • First CAMPI is out, CAMPI2 and CAMPI3 ongoing

• Need for better **education**
  • the microbiome community needs better education on the technical details and capabilities offered by metaproteomics, and easy access to experts
On our website (metaproteomics.org), we provide an overview of useful tutorials and lectures.

**Tutorials**
- Metaproteomics tutorial using Galaxy-P
- Integrative meta-omics using Galaxy-P
- Unipept
- MetaProteomeAnalyzer
- MetaQuantome, data creation
- MetaQuantome, functional annotation
- MetaQuantome, taxonomic annotation
- Ocean Protein Portal

**Lectures**
- “Metaproteomics to Investigate Functional Interactions in Microbiota” by Dr. Manuel Kleiner
- “Systems Ecology of the Human Gut Microbiome” by Dr. Paul Wilmes
- “Complexity in Ocean MetaProteomics: 20,000 Spectral Counts under the Sea.” by Dr. Mak Saito
- “Metaproteomics @ GTN Smörgåsbord” by Dr. Pratik Jagtap
- “New Pathways for Old Metals” by Dr. Nadia Szeinbaum
- “What can the tiniest organisms teach us about the vast ocean?” by Dr. Brook L. Nunn
- “pepFunk: an R shiny app and workflow for peptide-centric functional analysis of metaproteomic microbiome data” by Dr. Caitlin Simopoulos
- “Designer Fibre for Beneficial Microbes: A Path to Microbiome Modulation” by Dr. Phil Pope
- “Metaproteomics session at TSMC 2021” by Dr. Pratik Jagtap, Dr. Bob Hettich, and Dr. Mak Saito
We organize metaproteomics workshops at several conferences and institutes

2023: EuPA conference (Newcastle, UK), FEMS conference (Hamburg, Germany), ASMS (Houston, TX, USA), ...
Why an Initiative?

• Need for better **communication** and **collaboration**
  • In the next years, major new analytical developments are likely to facilitate unprecedented insights into microbiome function and dynamics
  • Many metaproteomics researchers are working solitary in single-species proteomics labs, or in metagenomics/metatranscriptomics labs
  • First CAMPI is out, CAMPI2 and CAMPI3 ongoing

• Need for better **education**
  • the microbiome community needs better education on the technical details and capabilities offered by metaproteomics, and easy access to experts

• Need for (better) **standardization**
  • efforts to standardize methodologies, will accelerate experimental and bioinformatic innovations
Standardization in metaproteomics was first discussed at the IMS in Avignon (April 2023)

Please fill in the following 5-minute survey:
https://tinyurl.com/metaproteomics-standardization

www.metaproteomics.org | @MetaP_Init
More information can be found in our commentary in *Microbiome*

**The Metaproteomics Initiative: a coordinated approach for propelling the functional characterization of microbiomes**


*Microbiome* 9, Article number: 243 (2021) | [Cite this article]
Who are we?

>200 members from 77 labs in 19 countries (12 in Europe, Australia, Canada, China, India and USA)*

Who’s already a member? 😊

* In this count, only the labs and countries are included of the researchers who want to have their lab listed on our website.
Becoming a member is really easy!

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
Robert Hettich
Oak Ridge National Lab
Famous Avignon Bridge
VENUE

Palais des Papes - Avignon

The Palace of the Popes is the biggest building ever built during the Gothic period. It stands as a powerful testimony to the presence of 9 popes who lived in and reigned from Avignon in the 1300s. Construction of the Palace took less than 20 years, and took place between 1334 and 1342. Two popes were the primary builders of the Palace: Pope Benedict XII, who built the first pontifical palace (now referred to as the "Old Palace"), and Pope Clement VI, who built new extensions, referred to as the "New Palace".
Session 1. Metaproteomics in Health and Disease
Session 2. Environmental Metaproteomics
Session 3. New Methods and Trends in Metaproteomics
Session 4. Modelling and Managing Microbial Communities
Session 5. Integrative Multiomics and Bioinformatics
Session 6. Metaproteomic of Complex/Technical Habitats and Eukaryotes
Critical Assessment of Metaproteome Investigation, CAMPI
Proposed next round of CAMPI projects
Avignon 2023

Scientific Committee
Robert Hettich, USA
Jean Armengaud, France
Pratik Jagtap, USA
Sergio Uzzau, Italy
Dan Figeys, Canada
Paul Wilmes, Luxembourg
Dirk Benndorf, Germany
Nico Jehmlich, Germany
Critical Assessment of MetaProteome Investigation (CAMPI): a multi-laboratory comparison of established workflows

1. The primary objective of this meeting was to better define and launch the next series of CAMPI projects.

2. Based on the society survey, a number of small and large projects were discussed.

3. There was consensus to run two small challenge CAMPI projects simultaneously.

4. We should be careful to define what CAMPI “is” and “is not.” What are our desired endpoints? We should define our challenges carefully.

5. We should set concrete goals for each of the projects and define the work scopes.
Scientific Committee CAMPI consideration recommendations

**CAMPI Goals:**

1. Take into account the ranking below from the survey table.

<table>
<thead>
<tr>
<th>Ideas</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. [Validation/Benchmark] Statistical analysis of metaproteomic data. Different levels of aggregation: peptides, (meta)proteins, taxa, functions, taxon-specific functions, pathways, etc. Comparison of different approaches for multivariate analysis, differential analysis, discriminant analysis, enrichment analysis, etc. Association of (clinical) metadata.</td>
<td>36</td>
</tr>
<tr>
<td>2. [Validation] How to do differential metaproteomics? How can statistical approaches be adapted (i.e. normalization, imputation and testing)? SC or XIC approach on specific peptides for quantification?</td>
<td>31</td>
</tr>
<tr>
<td>3. [Validation/Benchmark] What quantification approaches are appropriate for proteins in metaproteomics? Can area under the curve or peak intensity approaches be used reliably with the high mass peak density in metaproteomic spectra? and if yes can features be matched between runs?</td>
<td>31</td>
</tr>
<tr>
<td>4. [Challenges] Work on low biomass samples</td>
<td>28</td>
</tr>
<tr>
<td>5. [Benchmark] 1-Compare the best bioinformatic/statistical solutions for extracting functional information from a metaproteomics dataset acquired on a single sample (without any preliminary information). 2-Compare the best solutions for extracting functional information (biological insight) from the very same microbiota in two or three different conditions (with biological triplicates).</td>
<td>27</td>
</tr>
<tr>
<td>6. [Validation] What about DIA in metaproteomics? Comparison between DDA and direct DIA and DIA-PASEF: identification and quantification</td>
<td>26</td>
</tr>
<tr>
<td>7. [Challenges] Protein inference and attribution of shared proteins groups to their correct microbe</td>
<td>25</td>
</tr>
<tr>
<td>8. Can we test the “power” of metaproteomics in revealing biological function, by correlating to metatranscriptomics? :might be controversial, but could be good to show how metaproteomics is SO MUCH MORE revealing than metatranscriptomics due to suspect correlations :)</td>
<td>25</td>
</tr>
<tr>
<td>9. [Validation] Can label based (e.g. TMT) approaches be used to increase throughput and reliability of quantification in metaproteomics?</td>
<td>25</td>
</tr>
<tr>
<td>10. [Challenges] Soil/Rhizosphere metaproteomics</td>
<td>24</td>
</tr>
<tr>
<td>11. Challenges in clinical metaproteomics</td>
<td>24</td>
</tr>
<tr>
<td>12. The human metaproteomic project. Cataloging all the expressed proteins and their functions. This would stimulate the development of technology and software[Daniel Figeys]</td>
<td>24</td>
</tr>
<tr>
<td>13. [Challenges] generating a repository of human microbiome biological functions (colon, oral, vagina, stomach, ...)</td>
<td>24</td>
</tr>
<tr>
<td>14. Can we use a microbiome of known taxonomy and/or functional content and generate a deep metagenomic dataset? We can share the metagenomic database with expert laboratories to generate a metaproteomics database using assembly and annotation methods (both taxonomy and functional). We can then evaluate the databases against a MS dataset generated from this sample to evaluate which methods give the best results with respect to taxonomy and function analysis. This will offer a list of features that should be considered to generate an optimal metaproteomic database. It might also get some metagenomics researchers interested in this challenge.</td>
<td>21</td>
</tr>
<tr>
<td>15. [Limitations] Increasing MAG database sizes and their effect on identification rates</td>
<td>19</td>
</tr>
<tr>
<td>16. Targeted Metaproteomics.</td>
<td>17</td>
</tr>
<tr>
<td>17. [Limitations] Computational capacity/software to work with big datasets</td>
<td>12</td>
</tr>
</tbody>
</table>
Scientific Committee CAMPI consideration recommendations

CAMPI Goals:

1. Objective is another “group publication,” in which we deliver a study that could not have been performed by only one or a few research groups. The goal would be to generate and publish a mega-study in a high-impact paper. This could address multiple issues as proposed by the survey table.

2. Involve the largest possible IMS community, preferably at the international level

3. The following draft projects outline on how the IMS community can approach the high-impact goal

4. These would serve as the springboard for the next series of CAMPI research projects.
<table>
<thead>
<tr>
<th>IG name</th>
<th>idea(s) rank #</th>
<th>Short description</th>
<th>Start</th>
<th>end</th>
<th>Expected result(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample preservation</td>
<td>11,12,13</td>
<td>Assess robust methods to fossilize and preserve snapshot(s) of microbial protein expression obtained by metaproteomics. Methods should: 1. be alternative to simply freezing samples 2. lead to user friendly standards 3. fit for purpose also “on field” (any environmental /rural area, patients point of care, etc.). This might be achieved by critically and comparatively assessing commercially available tools or lab protocols, to collect and preserve samples with “metaproteomics grade” quality. Fresh and preserved paired samples will be extracted and measured by participating labs and/or sent to other participating labs. All labs will test all selected sampling/preservation protocols.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The establishment of a standardized protocol to collect and preserve microbiome samples, enabling shipment of inactivated biological samples to near or far away proteomic labs.
### CAMPI-3 Avignon

<table>
<thead>
<tr>
<th>IG name</th>
<th>idea(s) rank #</th>
<th>Short description</th>
<th>Start</th>
<th>end</th>
<th>Expected result(s)</th>
</tr>
</thead>
</table>
| Bioinformatics | 1, 2, 3, and 5 | Collect metagenome and mass spectrometry data from a gut microbiome and conduct a thorough bioinformatics study to compare and contrast the best possible methods to extract metaproteome information, specifically focusing on database builds, protein redundancy, enrichment analyses, annotation of protein groups and functional information pathways, statistical analysis approaches, and quantification. |       |     | 1) The community will be provided with mass spectrometry data search results from a gut microbiome dataset with associated metadata.  
2) Community members will perform informatic analysis on the samples to extract functional information on the data at various levels: peptides, proteins, pathways, etc. |
CAMPI3: Annotations

GOALS:

• Evaluate bioinformatic methods used for taxonomic and functional annotation using a ground-truth “synthetic” dataset and a biological dataset.

CAMPI3 TEAM
Samantha Peters
Benoît Kunath
Zhibin Ning
Caitlin Simopoulos
Tim Van Den Bossche
Jaclyn Saunders
Bart Mesuere
Lucia Grenga
Thilo Muth
Paul Wilmes
Daniel Figeys
Robert Hettich
Pratik Jagtap
CAMPI-1 results
the impetus for CAMPI3

SIHUMI

Human feces

- TAXANOMY
- FUNCTIONS
- PROTEIN INFESSION
- PEPTIDE SPECTRAL MATCHING QUALITY

https://metaproteomics.org/campi
CAMPI3: Annotations

SYNTETIC COMMUNITY

• Known complexity and community
• Up to 25 sequenced bacterial isolates
• 4 mixes with variations on the composition or relative abundances

BIOLOGICAL DATASET

• Unknown complexity and community
• Family case study of Type 1 diabetes
• Multiple time points analyzed
• Strong meta-omics coverage highlighting strain-level diversity
CAMPI3: Annotations

**SYNTHETIC COMMUNITY (TAXONOMY)**

<table>
<thead>
<tr>
<th>Organism</th>
<th>Mix A</th>
<th>Mix B</th>
<th>Mix C</th>
<th>Mix D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism_01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism_25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Pre-analysis results:**

- **Synthetic Community**
  - Simplest mix
  - Altered abundances / composition
CAMPI3: Annotations

- SYNTHETIC COMMUNITY
- BIOLOGICAL DATASET
- GENERAL METRICS
  - TAXONOMY
  - FUNCTIONS
- PROTEIN INFERENCE (PROTEIN REPORT, PEPTIDE REPORT)
- PEPTIDE SPECTRAL MATCHING QUALITY (PSM REPORT)

Pre-Analysis
- JUNE 2023
  - ASMS (Houston)
  - Announcement

Promotion & Analysis
- DECEMBER 2023
  - Results reported by participants.
- Promotion
- Analysis by Participants

Analysis by CAMPI3 Team
- JUNE 2024
  - Data Analysis & Summarization
  - Results Report
  - Manuscript
Interested in participating in CAMPI3?

The Metaproteomics Initiative has announced the CAMPI3 study to assess the informatic methods used in the field of metaproteomics research. This is the latest study in the series of CAMPI studies (https://metaproteomics.org/campi/) initiated by the Metaproteomics Initiative. Through this study, the Metaproteomics Initiative intends to understand how decisions in different workflows impact assignments. We are looking for diverse perspectives and areas of expertise (tool developers, experimentalists, analysts, etc.)

Interested participants will be sent mass spectrometry files, protein sequence database, and parameters for searches for a ‘synthetic community’ dataset and a ‘biological dataset’. Reporting templates and metrics will be provided so that you can share the results. The results will be kept anonymous.
INTERESTED?

Fill in the survey:  
z.umn.edu/campi3ing

INFORMATION


DOWNLOAD

https://metaproteomics.org/campi3/download

RESULTS SUBMISSION

Contact Anonymizer ningzhibin@gmail.com

QUESTIONS?

Contact pjagtap@umn.edu and peterssl@ornl.gov
Bioinformatics Hub @ ASMS : Metaproteomics

• THURSDAY, JUNE 8, 2023
• 10:30 AM to 12:30 PM
• METAPROTEOMICS RESEARCHERS AND BIOINFORMATICIANS DISCUSS THEIR CHALLENGES AND OPPORTUNITIES
CAMPI3: Annotations

**INTERESTED?**
Fill in the survey: [z.umn.edu/campi3ing](https://z.umn.edu/campi3ing)

**INFORMATION**

**DOWNLOAD**
[https://metaproteomics.org/campi3/download](https://metaproteomics.org/campi3/download)

**RESULTS SUBMISSION**
Contact Anonymizer ningzhibin@gmail.com

**QUESTIONS?**
Contact pjagtap@umn.edu and peterssl@ornl.gov