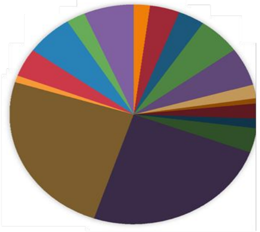
CAMPI3: Annotations

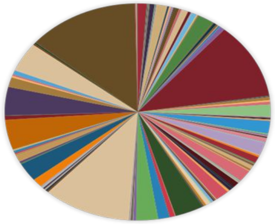
GOAL:

To Evaluate the impact of key decision points that are commonly encountered across bioinformatic methods on **taxonomic and functional annotation** using a ground-truth “synthetic” and a biological dataset.



SYNTHETIC COMMUNITY

- **Known** complexity and community
- **Up to 25** sequenced bacterial isolates
- **Four mixes** with variations on the composition of relative abundances



BIOLOGICAL DATASET

- **Unknown** complexity and community
- Parkinson Disease and healthy individuals gut microbiome samples.
- Shotgun Metagenomics and 16S rRNA data available

CAMPI3 TEAM

Benoît Kunath
Lucia Grenga
Zhibin Ning
Samantha Peters
Hamid Hachemi
Gelio Alves
Tim van den Bossche
Bart Mesuere
Thilo Muth
Robert Hettich
Pratik Jagtap

Synthetic Community

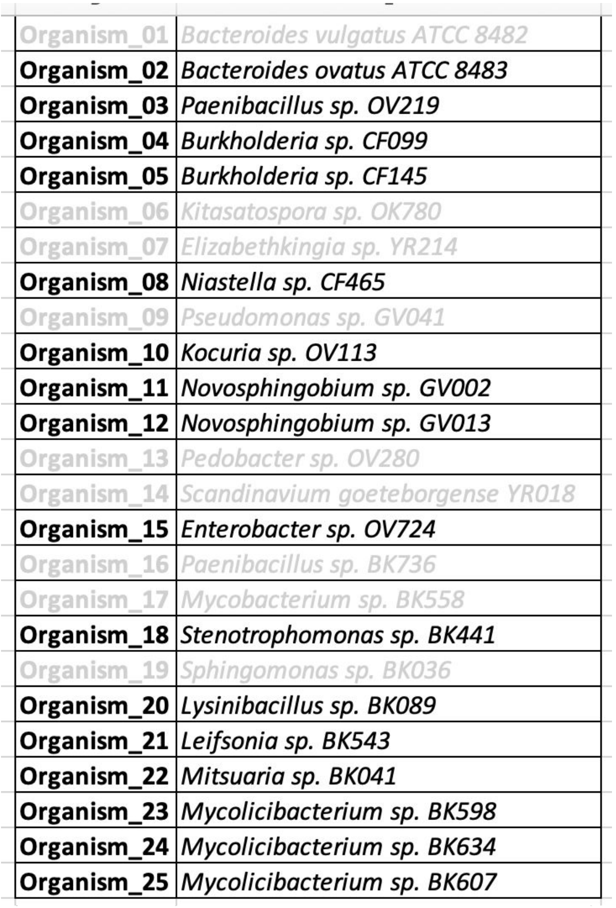
CAMPI3: Annotations

Organism_01	*Bacteroides vulgatus ATCC 8482*
Organism_02	*Bacteroides ovatus ATCC 8483*
Organism_03	*Paenibacillus sp. OV219*
Organism_04	*Burkholderia sp. CF099*
Organism_05	*Burkholderia sp. CF145*
Organism_06	*Kitasatospora sp. OK780*
Organism_07	*Elizabethkingia sp. YR214*
Organism_08	*Niastella sp. CF465*
Organism_09	*Pseudomonas sp. GV041*
Organism_10	*Kocuria sp. OV113*
Organism_11	*Novosphingobium sp. GV002*
Organism_12	*Novosphingobium sp. GV013*
Organism_13	*Pedobacter sp. OV280*
Organism_14	*Scandinavium goeteborgense YR018*
Organism_15	*Enterobacter sp. OV724*
Organism_16	*Paenibacillus sp. BK736*
Organism_17	*Mycobacterium sp. BK558*
Organism_18	*Stenotrophomonas sp. BK441*
Organism_19	*Sphingomonas sp. BK036*
Organism_20	*Lysinibacillus sp. BK089*
Organism_21	*Leifsonia sp. BK543*
Organism_22	*Mitsuaria sp. BK041*
Organism_23	*Mycolicibacterium sp. BK598*
Organism_24	*Mycolicibacterium sp. BK634*
Organism_25	*Mycolicibacterium sp. BK607*

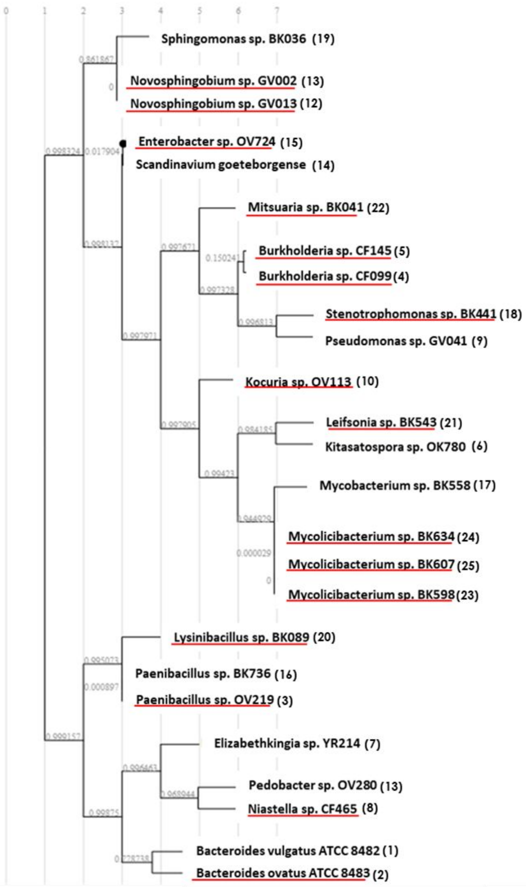
Samantha Peters

Synthetic Community

CAMPI3: Annotations



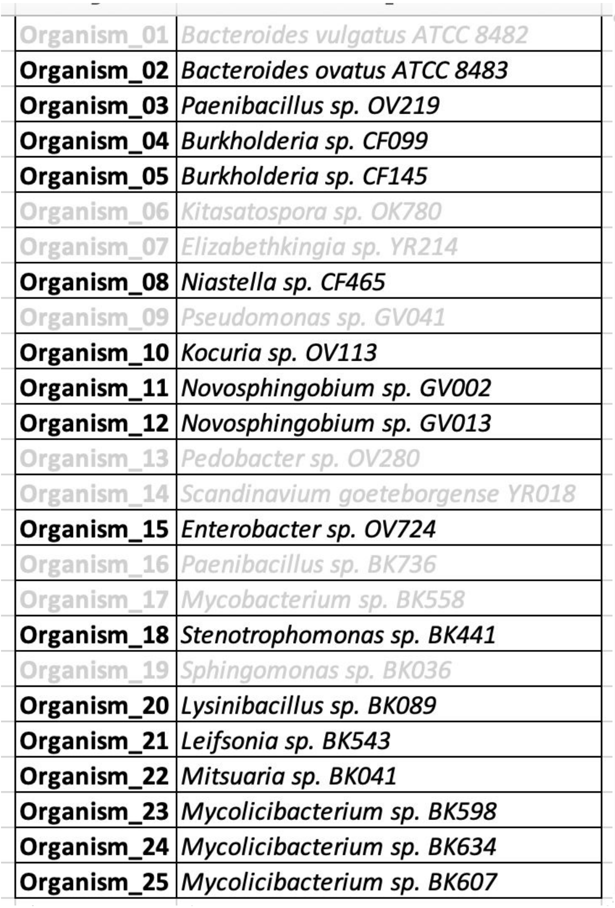
Organism_01	*Bacteroides vulgatus ATCC 8482*
Organism_02	*Bacteroides ovatus ATCC 8483*
Organism_03	*Paenibacillus sp. OV219*
Organism_04	*Burkholderia sp. CF099*
Organism_05	*Burkholderia sp. CF145*
Organism_06	*Kitasatospora sp. OK780*
Organism_07	*Elizabethkingia sp. YR214*
Organism_08	*Niastella sp. CF465*
Organism_09	*Pseudomonas sp. GV041*
Organism_10	*Kocuria sp. OV113*
Organism_11	*Novosphingobium sp. GV002*
Organism_12	*Novosphingobium sp. GV013*
Organism_13	*Pedobacter sp. OV280*
Organism_14	*Scandinavium goeteborgense YR018*
Organism_15	*Enterobacter sp. OV724*
Organism_16	*Paenibacillus sp. BK736*
Organism_17	*Mycobacterium sp. BK558*
Organism_18	*Stenotrophomonas sp. BK441*
Organism_19	*Sphingomonas sp. BK036*
Organism_20	*Lysinibacillus sp. BK089*
Organism_21	*Leifsonia sp. BK543*
Organism_22	*Mitsuaria sp. BK041*
Organism_23	*Mycolicibacterium sp. BK598*
Organism_24	*Mycolicibacterium sp. BK634*
Organism_25	*Mycolicibacterium sp. BK607*



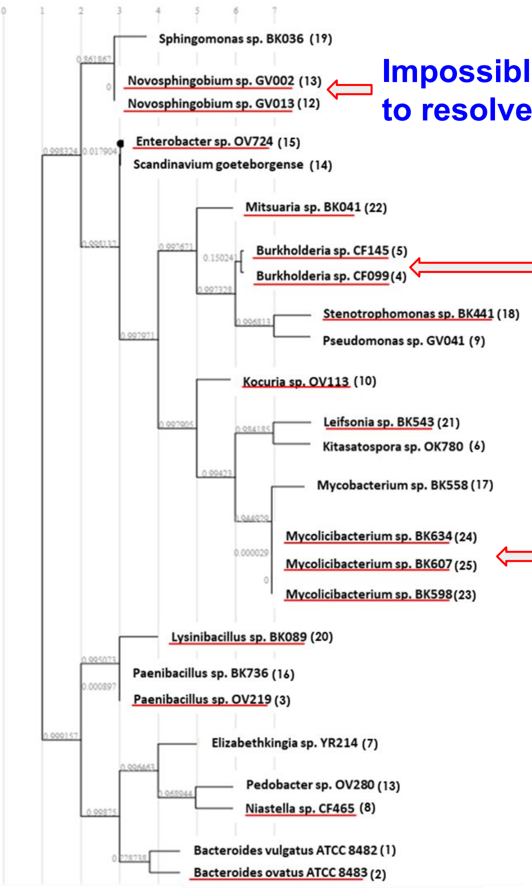
Samantha Peters

Synthetic Community

CAMPI3: Annotations



Organism_01	*Bacteroides vulgatus ATCC 8482*
Organism_02	*Bacteroides ovatus ATCC 8483*
Organism_03	*Paenibacillus sp. OV219*
Organism_04	*Burkholderia sp. CF099*
Organism_05	*Burkholderia sp. CF145*
Organism_06	*Kitasatospora sp. OK780*
Organism_07	*Elizabethkingia sp. YR214*
Organism_08	*Niastella sp. CF465*
Organism_09	*Pseudomonas sp. GV041*
Organism_10	*Kocuria sp. OV113*
Organism_11	*Novosphingobium sp. GV002*
Organism_12	*Novosphingobium sp. GV013*
Organism_13	*Pedobacter sp. OV280*
Organism_14	*Scandinavium goeteborgense YR018*
Organism_15	*Enterobacter sp. OV724*
Organism_16	*Paenibacillus sp. BK736*
Organism_17	*Mycobacterium sp. BK558*
Organism_18	*Stenotrophomonas sp. BK441*
Organism_19	*Sphingomonas sp. BK036*
Organism_20	*Lysinibacillus sp. BK089*
Organism_21	*Leifsonia sp. BK543*
Organism_22	*Mitsuaria sp. BK041*
Organism_23	*Mycolicibacterium sp. BK598*
Organism_24	*Mycolicibacterium sp. BK634*
Organism_25	*Mycolicibacterium sp. BK607*



Impossible to resolve

Easy to resolve

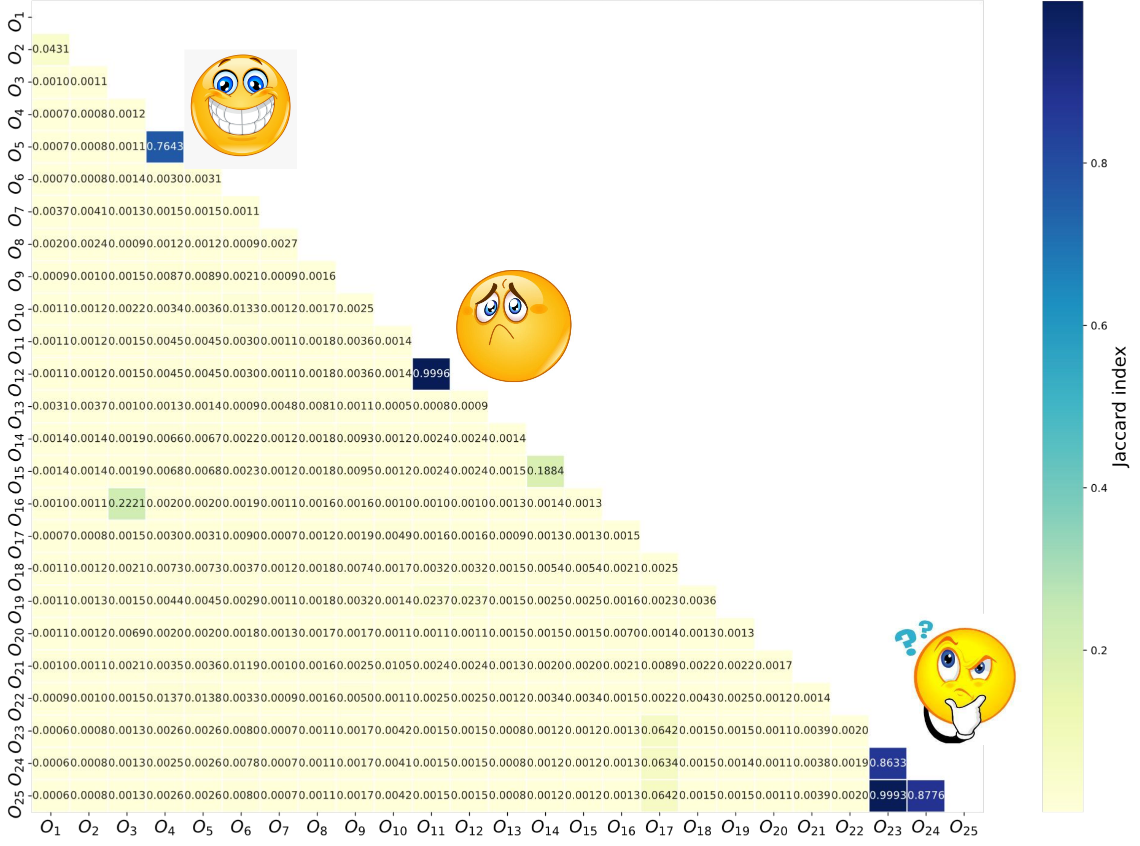
Difficult to resolve

Samantha Peters

Synthetic Community

CAMPI3: Annotations

Tryptic Peptidome Similarity



	O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8	O_9	O_{10}	O_{11}	O_{12}	O_{13}	O_{14}	O_{15}	O_{16}	O_{17}	O_{18}	O_{19}	O_{20}	O_{21}	O_{22}	O_{23}	O_{24}	O_{25}
O_1																									
O_2	0.0431																								
O_3	0.0010	0.0011																							
O_4	0.0007	0.0008	0.0012																						
O_5	0.0007	0.0008	0.0011	0.7643																					
O_6	0.0007	0.0008	0.0014	0.0030	0.0031																				
O_7	0.0037	0.0041	0.0013	0.0015	0.0015	0.0011																			
O_8	0.0020	0.0024	0.0009	0.0012	0.0012	0.0009	0.0027																		
O_9	0.0009	0.0010	0.0015	0.0087	0.0089	0.0021	0.0009	0.0016																	
O_{10}	0.0011	0.0012	0.0022	0.0034	0.0036	0.0133	0.0012	0.0017	0.0025																
O_{11}	0.0011	0.0012	0.0015	0.0045	0.0045	0.0030	0.0011	0.0018	0.0036	0.0014															
O_{12}	0.0011	0.0012	0.0015	0.0045	0.0045	0.0030	0.0011	0.0018	0.0036	0.0014	0.9996														
O_{13}	0.0031	0.0037	0.0010	0.0013	0.0014	0.0009	0.0048	0.0081	0.0011	0.0005	0.0008	0.0009													
O_{14}	0.0014	0.0014	0.0019	0.0066	0.0067	0.0022	0.0012	0.0018	0.0093	0.0012	0.0024	0.0024	0.0014												
O_{15}	0.0014	0.0014	0.0019	0.0068	0.0068	0.0023	0.0012	0.0018	0.0095	0.0012	0.0024	0.0024	0.0015	0.1884											
O_{16}	0.0010	0.0011	0.2221	0.0020	0.0020	0.0019	0.0011	0.0016	0.0016	0.0010	0.0010	0.0010	0.0013	0.0014	0.0013										
O_{17}	0.0007	0.0008	0.0015	0.0030	0.0031	0.0090	0.0007	0.0012	0.0019	0.0049	0.0016	0.0016	0.0009	0.0013	0.0013	0.0015									
O_{18}	0.0011	0.0012	0.0021	0.0073	0.0073	0.0037	0.0012	0.0018	0.0074	0.0017	0.0032	0.0032	0.0015	0.0054	0.0054	0.0021	0.0025								
O_{19}	0.0011	0.0013	0.0015	0.0044	0.0045	0.0029	0.0011	0.0018	0.0032	0.0014	0.0237	0.0237	0.0015	0.0025	0.0025	0.0016	0.0023	0.0036							
O_{20}	0.0011	0.0012	0.0069	0.0020	0.0020	0.0018	0.0013	0.0017	0.0017	0.0011	0.0011	0.0011	0.0015	0.0015	0.0015	0.0070	0.0014	0.0013	0.0013						
O_{21}	0.0010	0.0011	0.0021	0.0035	0.0036	0.0119	0.0010	0.0016	0.0025	0.0105	0.0024	0.0024	0.0013	0.0020	0.0020	0.0021	0.0089	0.0022	0.0022	0.0017					
O_{22}	0.0009	0.0010	0.0015	0.0137	0.0138	0.0033	0.0009	0.0016	0.0050	0.0011	0.0025	0.0025	0.0012	0.0034	0.0034	0.0015	0.0022	0.0043	0.0025	0.0012	0.0014				
O_{23}	0.0006	0.0008	0.0013	0.0026	0.0026	0.0080	0.0007	0.0011	0.0017	0.0042	0.0015	0.0015	0.0008	0.0012	0.0012	0.0013	0.0642	0.0015	0.0015	0.0011	0.0039	0.0020			
O_{24}	0.0006	0.0008	0.0013	0.0025	0.0026	0.0078	0.0007	0.0011	0.0017	0.0041	0.0015	0.0015	0.0008	0.0012	0.0012	0.0013	0.0634	0.0015	0.0014	0.0011	0.0038	0.0019	0.8633		
O_{25}	0.0006	0.0008	0.0013	0.0026	0.0026	0.0080	0.0007	0.0011	0.0017	0.0042	0.0015	0.0015	0.0008	0.0012	0.0012	0.0013	0.0642	0.0015	0.0015	0.0011	0.0039	0.0020	0.9993	0.8776	

Mix A: All sixteen organisms in equivalent amounts

~~Mix B: All sixteen organisms present, but in varying amounts~~

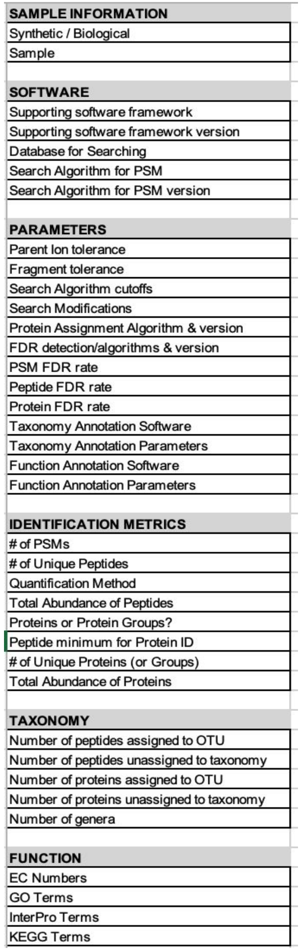
Mix C: Fifteen organisms present, with #24 left out

Mix D: Fifteen organisms present, with #15 left out

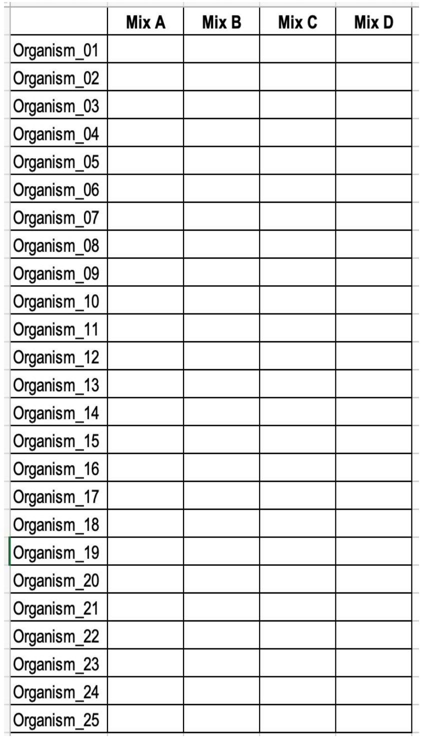
CAMPI3: Annotations

- **SYNTHETIC COMMUNITY**
- **BIOLOGICAL DATASET**

- **GENERAL METRICS**
- **TAXONOMY**
- **FUNCTIONS**
- **PROTEIN REPORT**
- **PEPTIDE REPORT**
- **PSM REPORT**



SAMPLE INFORMATION
Synthetic / Biological
Sample
SOFTWARE
Supporting software framework
Supporting software framework version
Database for Searching
Search Algorithm for PSM
Search Algorithm for PSM version
PARAMETERS
Parent Ion tolerance
Fragment tolerance
Search Algorithm cutoffs
Search Modifications
Protein Assignment Algorithm & version
FDR detection/algorithms & version
PSM FDR rate
Peptide FDR rate
Protein FDR rate
Taxonomy Annotation Software
Taxonomy Annotation Parameters
Function Annotation Software
Function Annotation Parameters
IDENTIFICATION METRICS
# of PSMs
# of Unique Peptides
Quantification Method
Total Abundance of Peptides
Proteins or Protein Groups?
Peptide minimum for Protein ID
# of Unique Proteins (or Groups)
Total Abundance of Proteins
TAXONOMY
Number of peptides assigned to OTU
Number of peptides unassigned to taxonomy
Number of proteins assigned to OTU
Number of proteins unassigned to taxonomy
Number of genera
FUNCTION
EC Numbers
GO Terms
InterPro Terms
KEGG Terms



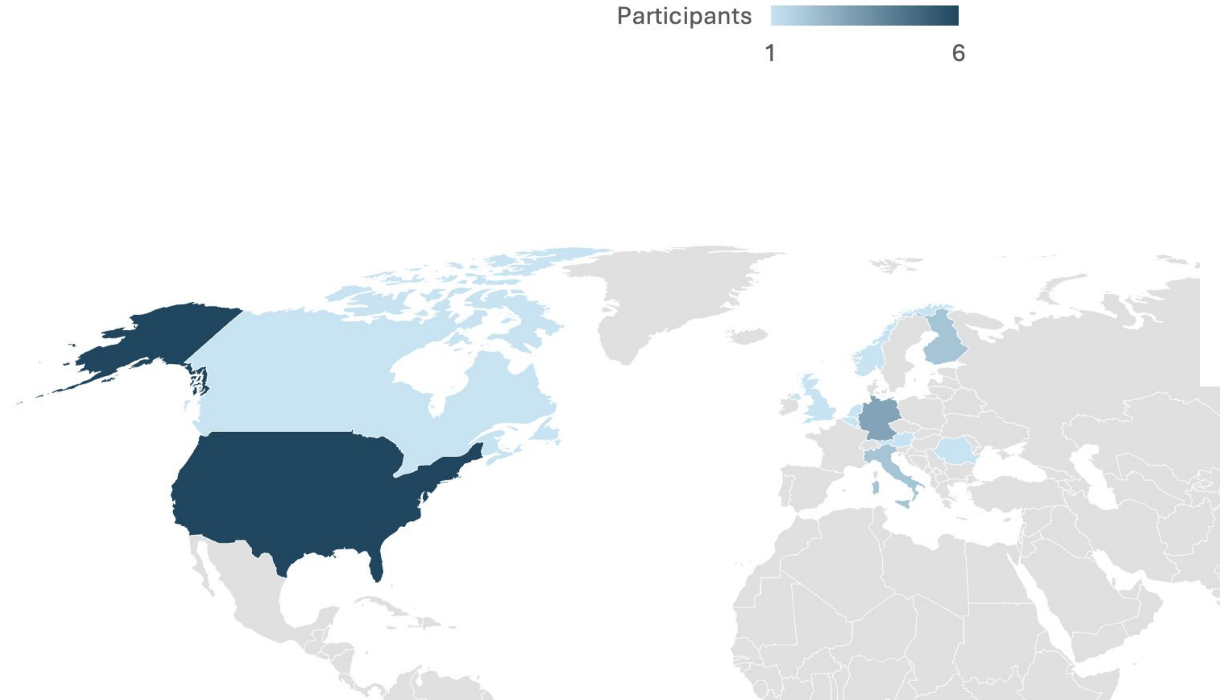
	Mix A	Mix B	Mix C	Mix D
Organism_01				
Organism_02				
Organism_03				
Organism_04				
Organism_05				
Organism_06				
Organism_07				
Organism_08				
Organism_09				
Organism_10				
Organism_11				
Organism_12				
Organism_13				
Organism_14				
Organism_15				
Organism_16				
Organism_17				
Organism_18				
Organism_19				
Organism_20				
Organism_21				
Organism_22				
Organism_23				
Organism_24				
Organism_25				

Submitted results

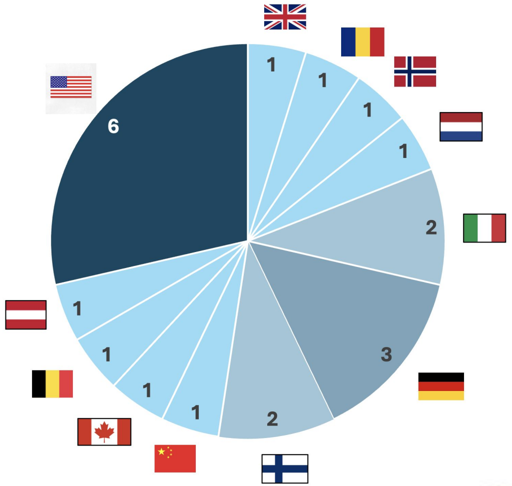
21 submissions (20 for synthetic and biological datasets + 1 synthetic dataset only)

20 laboratories from 12 countries

13 from Europe, 7 from America and 1 from Asia



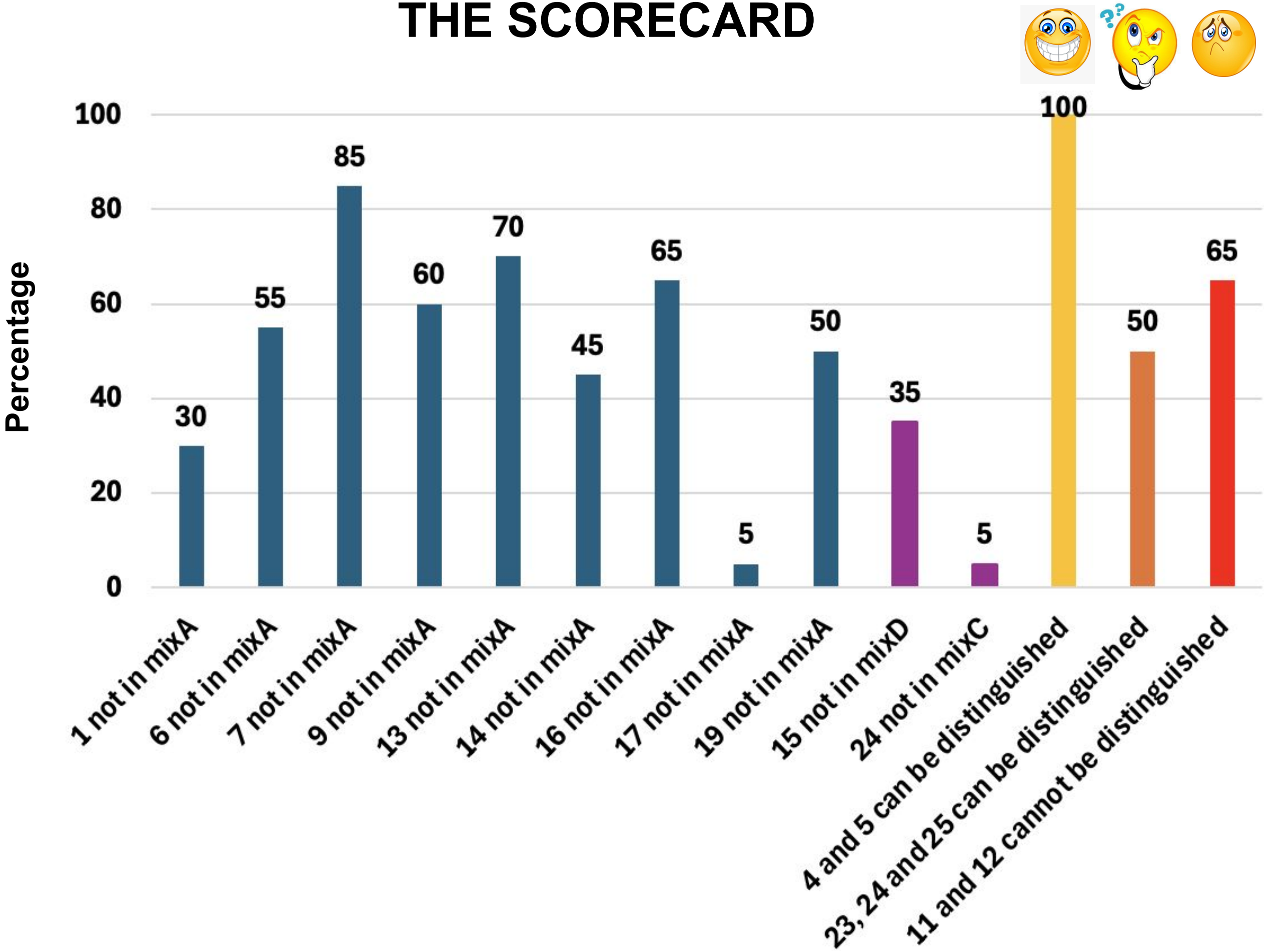
CAMPI3 Participants



Zhibin Ning

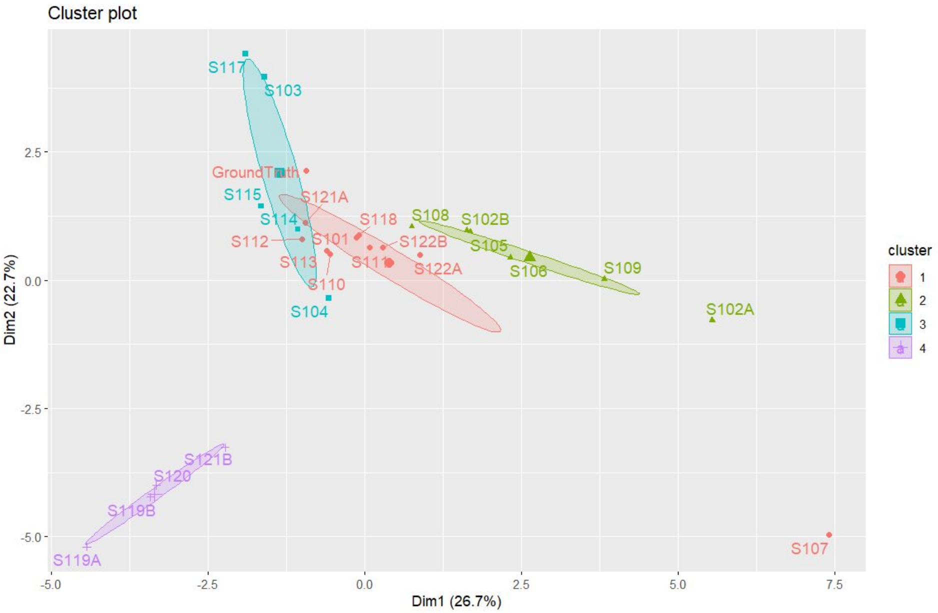
CAMPI3: Annotations

THE SCORECARD

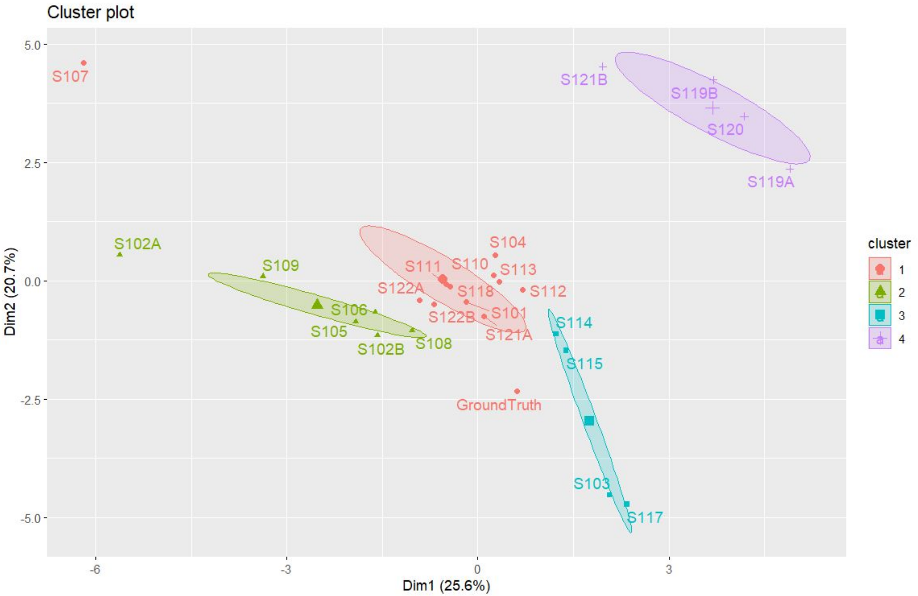


Results before interaction with participants

Mix A



Mix C



CAMPI3: Annotations

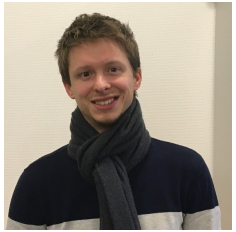
CURRENT STATUS OF TAXONOMIC PROFILING USING METAPROTEOMICS

IDENTIFY FACTORS THAT ARE RESPONSIBLE FOR THESE DIFFERENCES

SEEK FEEDBACK VIA INTERACTION WITH THE PARTICIPANTS

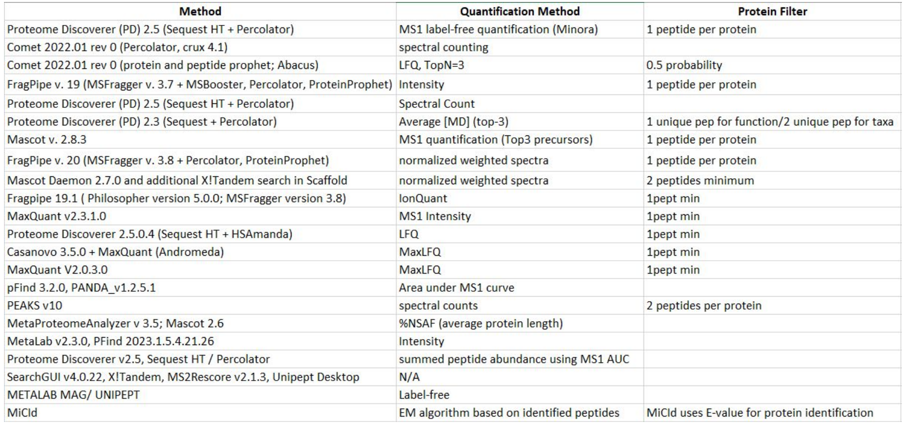
VALIDATION OF THE RECOMMENDATIONS THROUGH THE RE-ANALYSIS BY PARTICIPANTS

FINAL RECOMMENDATIONS BASED ON COMMUNITY LEVEL DISCUSSION

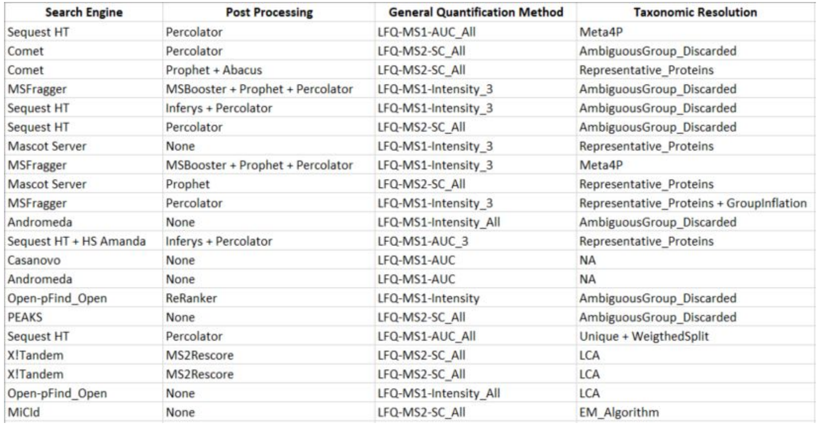


Ben Kunath

Interaction with participants



Method	Quantification Method	Protein Filter
Proteome Discoverer (PD) 2.5 (Sequest HT + Percolator)	MS1 label-free quantification (Minora)	1 peptide per protein
Comet 2022.01 rev 0 (Percolator, crux 4.1)	spectral counting	
Comet 2022.01 rev 0 (protein and peptide prophet; Abacus)	LFQ, TopN=3	0.5 probability
FragPipe v. 19 (MSFragger v. 3.7 + MSBooster, Percolator, ProteinProphet)	Intensity	1 peptide per protein
Proteome Discoverer (PD) 2.5 (Sequest HT + Percolator)	Spectral Count	
Proteome Discoverer (PD) 2.3 (Sequest + Percolator)	Average [MD] (top-3)	1 unique pep for function/2 unique pep for taxa
Mascot v. 2.8.3	MS1 quantification (Top3 precursors)	1 peptide per protein
FragPipe v. 20 (MSFragger v. 3.8 + Percolator, ProteinProphet)	normalized weighted spectra	1 peptide per protein
Mascot Daemon 2.7.0 and additional X!Tandem search in Scaffold	normalized weighted spectra	2 peptides minimum
Fragpipe 19.1 (Philosopher version 5.0.0; MSFragger version 3.8)	IonQuant	1pept min
MaxQuant v2.3.1.0	MS1 Intensity	1pept min
Proteome Discoverer 2.5.0.4 (Sequest HT + HSAmanda)	LFQ	1pept min
Casanovo 3.5.0 + MaxQuant (Andromeda)	MaxLFQ	1pept min
MaxQuant V2.0.3.0	MaxLFQ	1pept min
pFind 3.2.0, PANDA_v1.2.5.1	Area under MS1 curve	
PEAKS v10	spectral counts	2 peptides per protein
MetaProteomeAnalyzer v 3.5; Mascot 2.6	%NSAF (average protein length)	
MetaLab v2.3.0, PFind 2023.1.5.4.21.26	Intensity	
Proteome Discoverer v2.5, Sequest HT / Percolator	summed peptide abundance using MS1 AUC	
SearchGUI v4.0.22, X!Tandem, MS2Rescore v2.1.3, Unipept Desktop	N/A	
METALAB MAG/ UNIPEPT	Label-free	
MiCId	EM algorithm based on identified peptides	MiCId uses E-value for protein identification

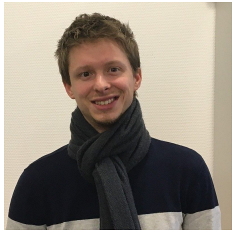


Search Engine	Post Processing	General Quantification Method	Taxonomic Resolution
Sequest HT	Percolator	LFQ-MS1-AUC_All	Meta4P
Comet	Percolator	LFQ-MS2-SC_All	AmbiguousGroup_Discarded
Comet	Prophet + Abacus	LFQ-MS2-SC_All	Representative_Proteins
MSFragger	MSBooster + Prophet + Percolator	LFQ-MS1-Intensity_3	AmbiguousGroup_Discarded
Sequest HT	Inferys + Percolator	LFQ-MS1-Intensity_3	AmbiguousGroup_Discarded
Sequest HT	Percolator	LFQ-MS2-SC_All	AmbiguousGroup_Discarded
Mascot Server	None	LFQ-MS1-Intensity_3	Representative_Proteins
MSFragger	MSBooster + Prophet + Percolator	LFQ-MS1-Intensity_3	Meta4P
Mascot Server	Prophet	LFQ-MS2-SC_All	Representative_Proteins
MSFragger	Percolator	LFQ-MS1-Intensity_3	Representative_Proteins + GroupInflation
Andromeda	None	LFQ-MS1-Intensity_All	AmbiguousGroup_Discarded
Sequest HT + HS Amanda	Inferys + Percolator	LFQ-MS1-AUC_3	Representative_Proteins
Casanovo	None	LFQ-MS1-AUC	NA
Andromeda	None	LFQ-MS1-AUC	NA
Open-pFind_Open	ReRanker	LFQ-MS1-Intensity	AmbiguousGroup_Discarded
PEAKS	None	LFQ-MS2-SC_All	AmbiguousGroup_Discarded
Sequest HT	Percolator	LFQ-MS1-AUC_All	Unique + WeigthedSplit
X!Tandem	MS2Rescore	LFQ-MS2-SC_All	LCA
X!Tandem	MS2Rescore	LFQ-MS2-SC_All	LCA
Open-pFind_Open	None	LFQ-MS1-Intensity_All	LCA
MiCId	None	LFQ-MS2-SC_All	EM_Algorithm

20 Groups

40 Meetings

Metadata correction and improvement

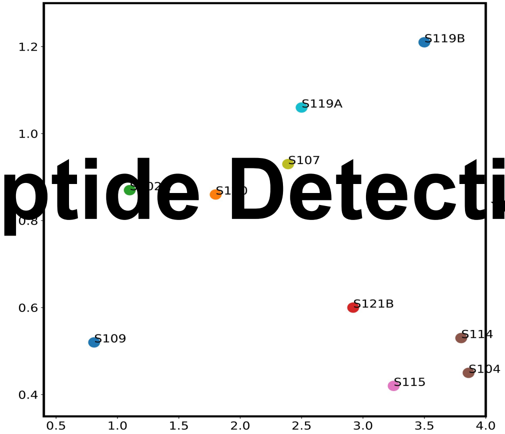


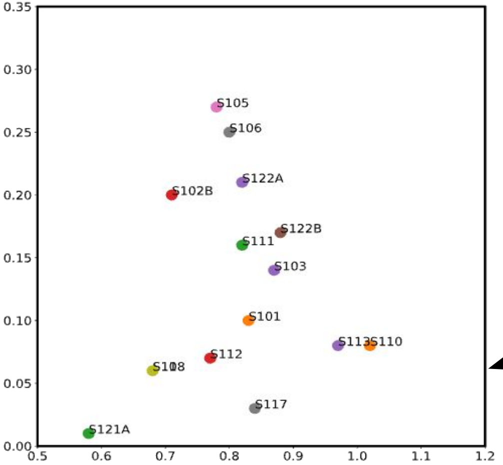
Ben Kunath

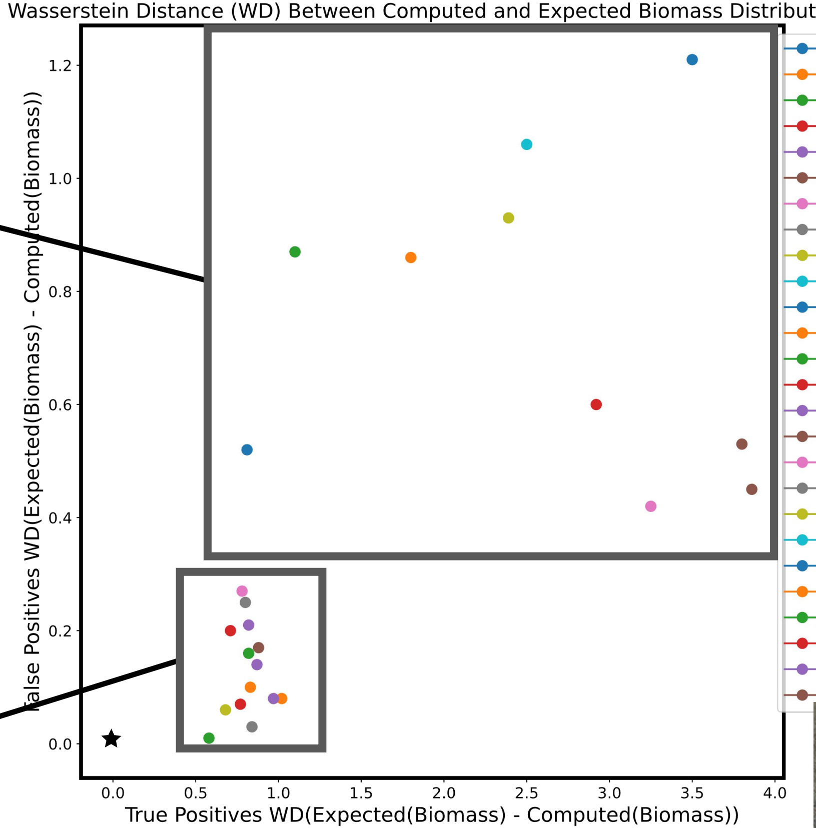
Break for “Mother-of-pearl-clouds” aka Polar stratospheric cloud aka Nacreous Clouds

Wasserstein distance between computed and expected biomass distributions for Mix A

Peptide Detection





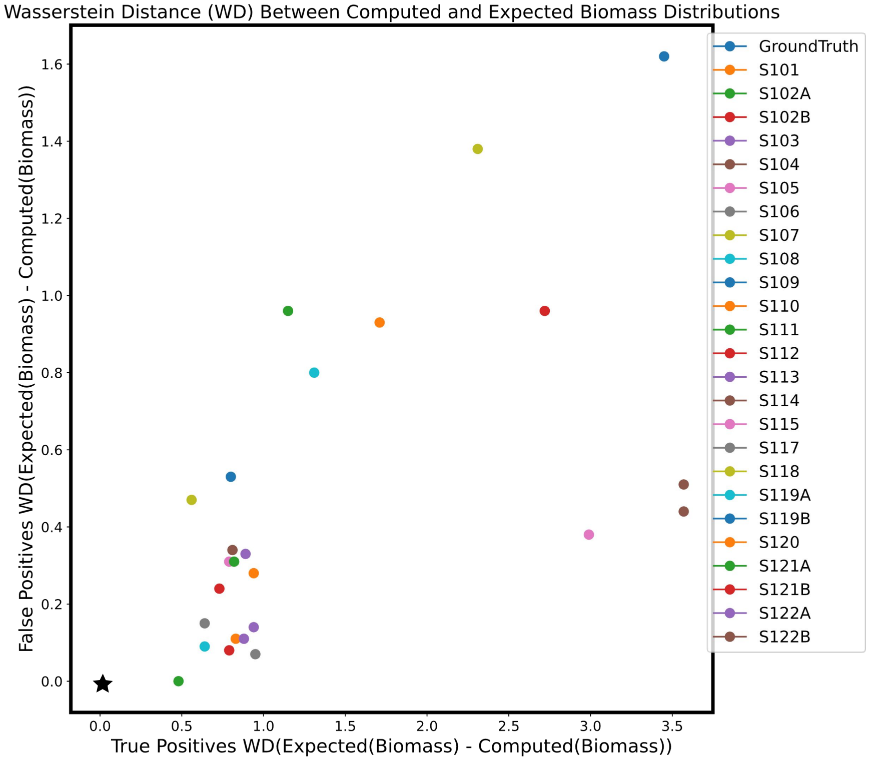


Gelio Alves

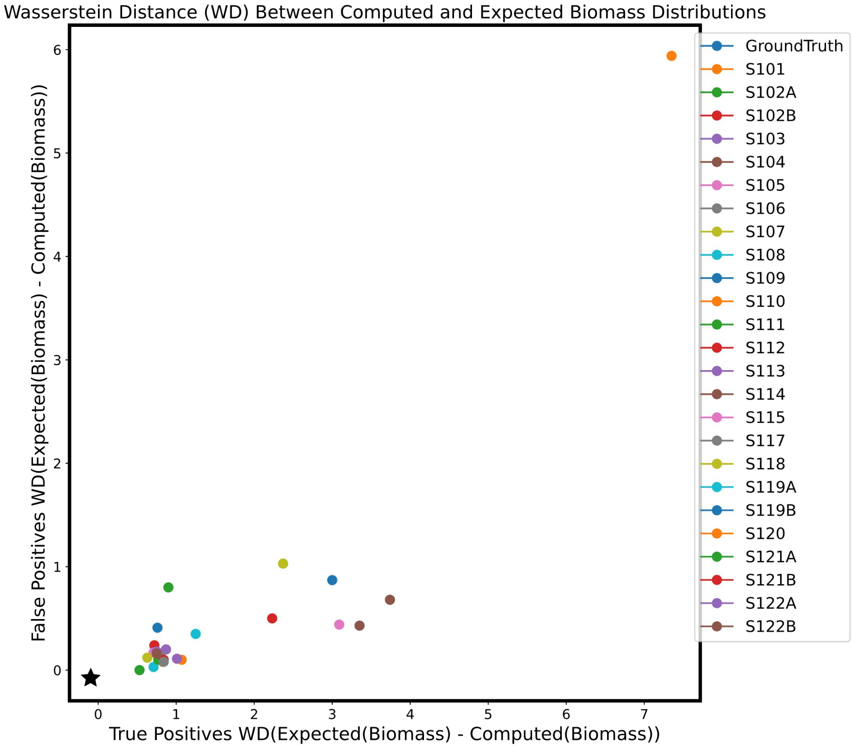
Peptide Detection

Wasserstein distance between computed and expected biomass distributions

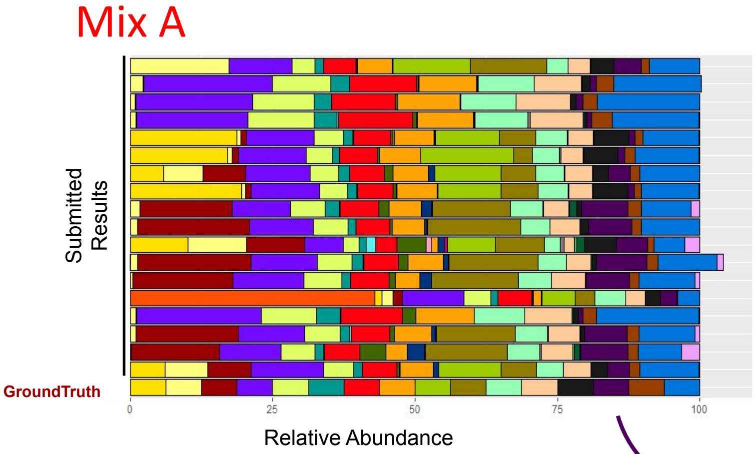
Mix C



Mix D

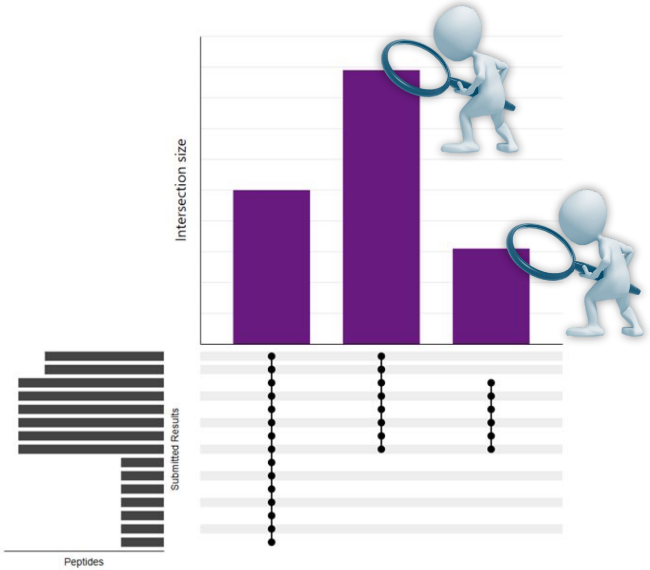


Peptide-level information



LIST OF PEPTIDES PER ORGANISM

Requested
(2nd interaction
with participants)

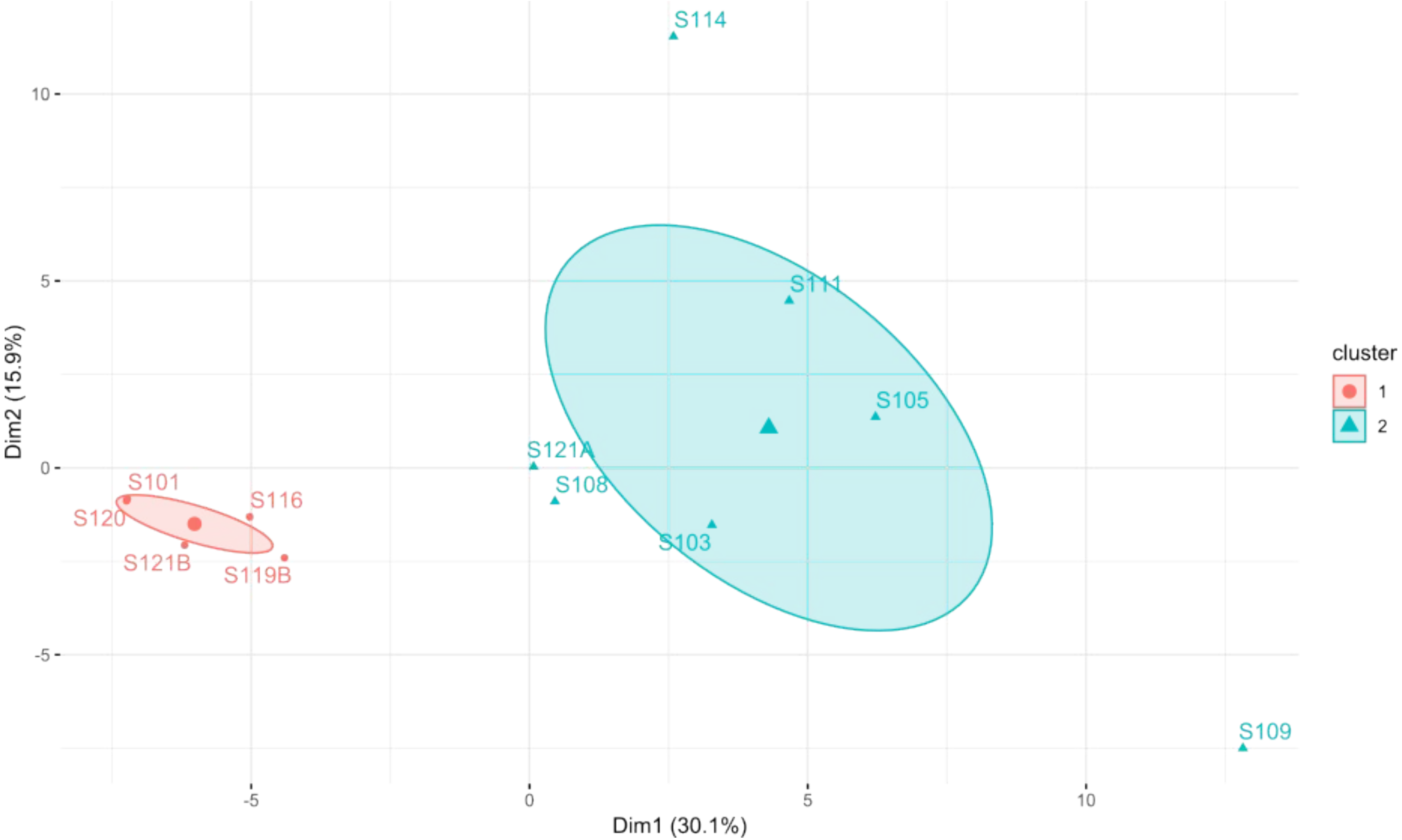


- Identify the peptides that could explain the differences in taxa abundance between submissions

Lucia Grenga

Insights from the Biological Dataset

Human Gut Microbiome (Parkinson’s Disease vs Healthy control)



Manuscript Plans

CAMPI3 Study
Decide on the title

Authors:
CAMPI3 Team with Ben, Lucia, Hamid as lead authors followed by Zhibin and Samantha. Decide on the order of the rest with Bob and Pratik as last authors.

Abstract:

Introduction:……………………………. ***March 2025***

Metaproteomics current status. Known and expected limitations notably compared to other omes

Introduce the Basic Aims:
- *Establish baseline/current situation within the community*
- *Identify strengths and limitations of the bioinformatic workflows for protein assignment and peptide spectral matching quality among users at varying levels of expertise and experience in the field of metaproteomics.*
- *Evaluate the impact of key decision points that are commonly encountered across bioinformatic methods on taxonomic and functional annotation.*
- *Provide recommendations and guidelines*

METHODS

Mass Spectrometry……………………………. ***December 2024***
Describe data acquisition parameters

Study promotion, enrollment, and timeline …………………………….***December 2024***
Describe the methodology used to seek results from 21 participants
https://z.umn.edu/campi3faqs

Information provided to the participants ……………………………. ***December 2024***
Describe the information provided to the 21 participants

Reporting Results ……………………………. ***January 2025***
Template (in supplement).
Results returned (supplement)

Interaction with the users ……………………………. ***January 2025***
One on one interaction with users.

Data processing and statistical analysis ……………………………. ***February 2025***
Methods used to generate outputs and figures.

CAMPI3 Summary

- **Synthetic microbiome dataset**: Three levels of difficulty for resolving bacterial species/strains were assessed.
- **Interaction with participants**: Feedback from the community will help us in identifying factors contributing to the differences.
- **Peptidome analysis**: We plan to identify peptides that could explain the differences in taxa abundance between submissions.
- **Biological data**: Assess the performance based on interaction with the participants.
- **Manuscript**: Recommendations for the community based on factors identified from both the synthetic and biological datasets.

QUESTIONS?