

Supporting Information

Progress and Challenges in Ocean Metaproteomics and Proposed Best Practices for Data Sharing

Mak A. Saito^{,a}, Erin M. Bertrand^b, Megan E. Duffy^c, David A. Gaylord^a, Noelle A. Held^a, William Judson Hervey^d, Robert L. Hettich^e, Pratik D. Jagtap^f, Michael G. Janech^g, Danie B. Kinkade^a, Dagmar H. Leary^d, Matthew R. McIlvin^a, Eli K. Moore^h, Robert M. Morris^c, Benjamin A. Neelyⁱ, Brook L. Nunn^j, Jaclyn K. Saunders^{a,c}, Adam I. Shepherd^a, Nick I. Symmonds^a, David A. Walsh^k*

a Woods Hole Oceanographic Institution

b Dalhousie University, Department of Biology

c University of Washington, School of Oceanography

d Naval Research Laboratory

e Oak Ridge National Laboratory and Microbiology Department, University of Tennessee

f University of Minnesota, Department of Biochemistry, Molecular Biology and Biophysics

g College of Charleston

h Rowan University, Department of Environmental Science

i National Institute of Standards and Technology

j University of Washington, Department of Genome Sciences

k Concordia University, Department of Biology

Corresponding Author

*Mak Saito. Marine Chemistry and Geochemistry Department, Woods Hole Oceanographic Institution, 360 Woods Hole Rd, Woods Hole MA 02543. msaito@whoi.edu, 1-508-289-2393

Table of Contents

Table S1. Examples of Environmental and Biomedical Metaproteomics Studies.....**S-2**

Table S2. Additional metadata associated with sample processing and analysis for, but not exclusive to, metaproteomics data.....**S-5**

Table S1. Examples of Environmental and Biomedical Metaproteomics Studies

Dataset	Database	Search algorithm and strategy	Taxonomy Analysis	Functional Analysis	Reference (PubMed)
Activated sludge 2D Electrophoresis (2004)	NCBI nr	Peptide Mass Fingerprinting using Mascot and De novo analysis and BLAST -P	-	-	15305916
Human Infant Gastrointestinal Tract metaproteome (2007)	NCBI nr	Peptide Mass Fingerprinting and Protein Prospector, De novo and BLAST -P	-	-	17158612
Ocean scale metaproteomics (2010)	GOS Combined Assembly Protein (P) database	Sequest, PeptideProphet and ProteinProphet and BLAST -P	-	-	20164862
Salivary metaproteomics (2010)	Uniprot/Swiss-Prot protein sequence database release 6.0	Sequest and PeptideProphet	BLAST -P and MEGAN	Clusters of Orthologous Groups and MetaCyc databases.	20331792
Salivary metaproteomics on highly fractionated database (2012)	HOMD database	ProteinPilot and two-step method	MEGAN analysis	MEGAN analysis	22522805
Temporal analysis of Human Intestinal microbiota (2012)	Publically available gut microbiome databases (including MetaHIT, human and food databases)	OMSSA and Progenesis (Quantitative analysis)	Mapped against UniProtKB and metagenomic analysis	Clusters of Orthologous Genes (COG) and BLAST analysis	22279554
Gutless marine worm metaproteomics (2012)	Olavius symbiont metagenomes and the genomes of related organisms	Sequest	Species assignment using metagenomic and metaproteomic data	Literature survey and METACYC database	22517752

Urine microbiome from asymptomatic bacteriuria (2012)	Human UniRef90 and 19 bacterial proteomes	Mascot and Percolator	Used MOTHUR for metagenomics data	Functions based on proteins identified	22929533
Rat Intestinal Microbiota (2012)	NCBI nr	Mascot and Proteome Discoverer	NCBI Taxonomy identifications	Clusters of Orthologous Genes (COG) and BLAST analysis	23016992
Metagenomics and Metaproteomics of Crohn's Disease in twins (2012)	Database derived from 51 reference proteomes and matched metagenome	Sequest and DT Aselect	Using metagenomics data	eggNOG, CAZY and KEGG orthologous groups	23209564
Metaproteomics of semiarid soil (2014)	UniRef100	Mascot and Scaffold	Scaffold identifications	Scaffold identifications	24530626
Rice plant root metaproteomics (2014)	Rice Root Microbiome database generated using Metagene Annotator on metagenome data	Mascot	Mascot Identification	Mascot Identification	24928870
Pre-term Infant Gut Metaproteome (2015)	Translated metagenomic databases and 16SrRNA analysis	Sequest and DT A Select and Usearch	-	Clusters of Orthologous Genes (COG) analysis	26077811
Biogas plant sludge metaproteomics (2015)	Biogas plant proteomes with SwissProt database	Mascot - standard search	MetaProteome Analyzer	KEGG Analysis	26152594
Targeted metaproteomics of Pacific Ocean (2015)	Pacific Ocean Metaproteome using MET ATRYP on microbial genome		-	-	26152594
Metaproteomics of aquatic estuary microbial communities (2015)	205 marine and archaeal proteomes and databases generated from metagenomics and metatranscriptomic databases	Sequest and Proteome Discoverer	BLAST -P and MEGAN analysis	Clusters of Orthologous Genes (COG) analysis	26223443

Active metaproteome and obesity (2015)	In-house human intestinal metaproteome database	OMSSA and X!tandem	UniPept	EggNOG analysis	26255997
Dental caries metaproteomics in presence of sucrose (2015)	HOMD database	ProteinPilot and two-step method	MEGAN analysis	MEGAN analysis	26684897
Metaproteomics of dysbiotic gut microbiota (2016)	NCBI nr	Mascot and Proteome Discoverer	PROteomics results Pruning & Homology group ANotation Engine" (PROPHANE)	PROteomics results Pruning & Homology group ANotation Engine" (PROPHANE)	25887379
Bering Strait and Chukchi Sea Metaproteome (2016)	Metapeptide database generated from shotgun metagenomic analysis	Comet	UniPept	UniPept based GO term analysis	27396978
Healthy and diseased stony coral analysis (2016)	NCBI nr data-base	Mascot	NCBI Taxonomy identifications	SEED subsystems profiles and the UniProt classification.	27492757
Mouse Fecal Sample Metaproteome (2016)	Combined database from host, food and bacterial (eleven) protein database	Sequest and Proteome Discoverer	-	Gene Ontology (GO) annotation	27559751
Periodontitis and Dental caries samples along with healthy controls (2016)	HOMD and UniProt SwissProt database	MaxQuant - two-step search method	Lowest Common Ancestor (LCA) analysis	Gene Ontology (GO)-terms and KEGG pathways from eggNOG analysis	27672500
Rhizosphere of plants living in serpentine soil (2017)	16SrRNA derived database	Mascot - standard search	UniProtKB taxonomy database	BLAST -P, InterPro, BLAST2GO	28083663

Table S2. Additional metadata associated with sample processing and analysis for, but not exclusive to, metaproteomics data.

Sample preparation	
Extraction method	link to published method when available
Reducing agent used	
Alkylating agent used	
Digestion enzyme used	
Data acquisition	
Experiment type (DDA, DIA, SRM-MRM)	(parameters in published methods)
MS Instrument	
Mass accuracy MS1 and MS2	
Activation method	
LC instrument	(parameters in published methods)
Column type	
2D chromatography column type	
Standardization	
Internal protein standard	provide sequences
Internal peptide standard	provide sequences
External protein standard	provide sequences
External peptide standard	provide sequences
Retention time standard	provide sequences
Targeted standard(s)	provide sequences
Data analysis	
Database used	Link to sequence repository
Contaminant database	Link to sequence repository
PSM Search engine used	
Primary peptide scoring metric	
Peptide false discovery rate (FDR)	
Protein false discovery rate (FDR)	

Quantitative data type and unit (spectral counts, MS1 intensities, etc.)	
FDR calculation method	
# missed cleavages permitted	
Min # peptides permitted for protein inference	
Post-translational modifications permitted (masses and residues)	