Supplement of Biogeosciences, 21, 4889–4908, 2024
https://doi.org/10.5194/bg-21-4889-2024-supplement
© Author(s) 2024. CC BY 4.0 License.

Biogeosciences
Open Access
EGU

Supplement of

Results from a multi-laboratory ocean metaproteomic intercomparison: effects of LC-MS acquisition and data analysis procedures

Mak A. Saito et al.

Correspondence to: Mak A. Saito (msaito@whoi.edu)

The copyright of individual parts of the supplement might differ from the article licence.

1
2 **Results from a Multi-Laboratory Ocean Metaproteomic Intercomparison: Effects of LC-**
3 **MS Acquisition and Data Analysis Procedures**
4
5 **Supplemental Materials**
6
7
8 **Table of Contents:**
9 Supplemental Tables S1-S12
10 Supplemental Ffigures S1-S4
11

12
13
14 **Table S1.** Participants in wet-lab (W) and informatic (I) components of the 2020-2021 OCB
15 ocean intercomparison study.

Institution(s)	Participants	Role
College of Charleston & NIST	Mike Janech, Ben Neely	W
Dalhousie University	Erin Bertrand, Scott McCain, Elden Rowland	W/I
Ghent University	Tim Van Den Bossche, Lennart Martens	I
Naval Research Laboratory	Judson Hervey, Dasha Leary, Jaimee Compton, Sophie Colston, Gary Vora	I
Rowan University and Rutgers University	Eli Moore, Haiyan Zheng	W
Oak Ridge National Laboratory	Bob Hettich, Samantha Peters, Richard Giannone	W/I
Ohio State University	Brian Searle	I
TU Delft	Martin Pabst and Hugo Kleikamp	I
University of Chicago	Jake Waldbauer	W
University of Minnesota	Pratik Jagtap, Tim Griffin, Subina Mehta	I
University of Vienna	Gerhard J. Herndl and Zihao Zhao	W/I
University of Washington Genome Sciences	Brook Nunn	W
University of Washington Oceanography	Rick Keil, Jacqui Neibauer, Megan Duffy	W
Woods Hole Oceanographic Institution	Mak Saito, Matthew McIlvin, Dawn Moran	W/I

16 **Table S2.** Metadata for laboratory intercomparison samples. Volumes filtered through 142 mm
17 pump heads and corresponding volume per slice.

18

Pump / Pump head / Sample name	Volume filtered (L)	Volume per 1/8th slice (L)
Pump 2L / BATS 1 / pump 1A	221.6*	27.7
Pump 2R / BATS 2 / pump 1B	167.3*	20.9
Pump 1L / BATS 3 / pump 2A	235.1+	29.4
Pump 1R / BATS 4 / pump 2B	211.1+	26.3

19 * Pump 1 total gauge = 447 L, sum of two pump gauges = 446.2 L

20 + Pump 2 total gauge = 478 L, sum of two gauges = 388.9 L, discrepancy of 89 L, gauges on pump head are assumed more accurate, as leaks in

21 system could create the additional flow for the total pump gauge.

22

23 **Table S3.** Sample metadata and accession numbers.

Expedition ID, Sample name	Location (Lat/Long)	Depth (m)	Date (mm-dd-yyyy)	Time (UTC; sampler recovery)	ProteomXchange ID
Laboratory Intercomparison	--	--	--	--	--
BATS 348, Lab 127	31.66 N 64.166 W	80	06-16-2018	08:00:00	PXD043218
BATS 348, Lab 135	31.66 N 64.166 W	80	06-16-2018	08:00:00	PXD043218
BATS 348, Lab 209	31.66 N 64.166 W	80	06-16-2018	08:00:00	PXD043218
BATS 348, Lab 438	31.66 N 64.166 W	80	06-16-2018	08:00:00	PXD043218
BATS 348, Lab 593	31.66 N 64.166 W	80	06-16-2018	08:00:00	PXD043218
BATS 348, Lab 652	31.66 N 64.166 W	80	06-16-2018	08:00:00	PXD043218
BATS 348, Lab 729	31.66 N 64.166 W	80	06-16-2018	08:00:00	PXD043218
BATS 348, Lab 774	31.66 N 64.166 W	80	06-16-2018	08:00:00	PXD043218
BATS 348, Lab 811	31.66 N 64.166 W	80	06-16-2018	08:00:00	PXD043218
BATS 348, paired metagenomic sample	31.66 N 64.166 W	80	06-16-2018	08:00:00	Bioproject Accession: PRJNA932835; SRA submission: SUB12819843
Informatics intercomparison	--	--	--	--	--
AE1913, Ocean 8 Clio020	33.128 N 65.967 W	120	06-19-2019	16:56:57	PXD044234
AE1913, Ocean 11 Clio020	33.128 N 65.967 W	20	06-19-2019	16:56:57	PXD044234

24

25
26

27

28 **Table S4. Experimental guidelines in ocean metaproteome intercomparison project**

Parameter	Guideline(s)
Extraction and digestion	Extraction of participant's choice, trypsin digestion
Chromatography	1-dimension of chromatography, at least 60 minutes of separation time, triplicate analyses
Total protein injected	1 μg suggested. Allowable range 0.25 - 2 μg
Isotope Tagging	No isotope tags
Mass spectrometry	Data Dependent Analyses (DDA), participant's choice of parameters
Informatics pipeline	Participants choice of software tools. Report in Spectral Counts. Protein and peptide results to be <1 % false discovery rate (FDR), 1 peptide per protein

29
30

31 **Table S5.** Laboratory intercomparison sample extraction method and LC method.
32
33

Lab ID	Extraction Method	LC method
127	2% SDS buffer 95oC; S-Trap purification and digestion	180 min run, 5% B (0.1% FA in acetontirile) to 30% B over 135 min, 30% B to 55% B over 12 min. A solvent 0.1% formic acid in water
135	5% SDS + 0.1M TEAB, tip sonication, S-trap digestion, c18 SPE	--
209	2% SDS, 95˚C + sonication; acetone precipitation; FASP cleanup & digestion	270 min run; 98% A (0.1 formic acid in water)/2% B (0.1% formic acid in acetonitrile) to 30% B over 130min, to 70% B over 45min
438	1% SDS buffer 95oC; SP3 bead purification and digestion	200 min run, 95% A (0.1 formic acid in water) to 95% B (0.1% formic acid in acetonitrile) nonlinear over 170 min, with a flow rate of 500nM min-1--
593	7M Urea 2M Thiourea, 1% DTT 2% CHAPS, vortex and sonicate, spin, ultrafiltration 30kD, filter aided sample prep (FASP) in solution digestion, desalt with C18 tips	180min gradient from 98% solution A (0.1% formic acid) and 2% solution B (90% acetonitrile and 0.1% formic acid) at 0 min to 40% solution B at 180 min with a flow rate of 300 nL min−1.
652	5% SDS + 0.1M TEAB applied to filters in ziplock, tip sonication, S-trap digestion, c18 SPE	120 min run, 5% B (0.1% FA in acetontirile) to 30% B over 90 min, 30% B to 55% B over 10 min. A solvent 0.1% formic acid in water
729	4% SDS sonication, protein aggregation capture	a linear organic gradient of 100% solvent A (95% water, 5% acetonitrile, 0.1% formic acid) to 25% solvent B (70% acetonitrile, 30% water, and 0.1% formic acid) for 180 minutes
774	2.1% SDS (2X Laemmeli buffer); SDS gel plug; 8M Urea; sonication	Sample was loaded on to a fused silica trap column (Acclaim PepMap 100, 75umx2cm, ThermoFisher). After washing for 5 min at 5 µl/min with 0.1% TFA, the trap column was brought in-line with an analytical column (Nanoease MZ peptide BEH C18, 130A, 1.7um, 75umx250mm, Waters) for LC-MS/MS. Peptides were fractionated at 300 nL/min using a segmented linear gradient 4-15% B in 30min (where A: 0.2% formic acid, and B: 0.16% formic acid, 80% acetonitrile), 15-25%B in 40min, 25-50%B in 44min, and 50-90%B in 11min. Solution B then returns at 4% for 5 minutes for the next run.
811	Bead beating and 3 freeze thaw cycles with Ammonium bicarbonate (50mM) and EDTA (5mM), centrifugation, then TCEP, iodoacetaminde, DTT, trypsin, desalted with C18 spin columns	Solvents of 100% LC/MS grade water with 0.1% formic acid (A) and 100% LC/MS grade acetonitrile with 0.1% formic acid (B) were used to elute peptides over a 90-minute gradient from 5-35% solvent B

34 **Table S6.** Chromatographic parameters and mass spectrometer and resolution employed. See

35 Table S5 for LC method and Table S7 for mass spectrometer parameters.

Lab ID	Column Length (cm)	Column Width (μm)	LC Resin	LC flow rate (nl/min)	LC gradient time (min)	Trap Column or Direct Injection	LC system	Mass Spectrometer	MS1 resolution
127	50	100	C18 Jupiter	250	147	direct	Dionex LC	QExactive	35,000 or 140,000
135	25	75	C18 Acclaim PepMap RSLC 2um	300	65	trap	Dionex LC 3000	Lumos Tribrid	60,000
209	200	100	C18 monolith (GL Sciences)	360	188	trap	Dionex 3000	Orbitrap Elite	120,000
438	25	100	3 μm C18 beads (Dr. Maisch)	500	200	trap	Dionex 3000	Fusion Tribrid	240,000
593	50	75	2μm C18 beads	300	270	--	Dionex UltiMate 3000	QExactive	120,000
652	30	75	3 μm C18 beads (Dr. Maisch)	250	90	trap	Thermo Easy-LC UPLC	QExactive	70,000
729	15	75	1.7μm Kinetex C-18 (Phenomenex)	150	180	trap	Vanquish Ultra-HPLC	QExactive Plus	70,000
774	25	75	Peptide BEH	300	120	trap	Dionex RSLC	Thermo Eclipse	120,000
811	37	75	C18 particles (Magic C18AQ, 100˚A, 5μm; Michrom)	300	90	precolumn	Easy-nLC 1200	Thermo Q Exactive Plus HRMS	70,000

36

37

38 **Table S7.** Chromatographic parameters and mass spectrometer and resolution employed. See

39 Table S5 for LC method and Table S6 for chromatographic parameters and mass spectrometer

40 and resolution employed.

Lab ID	MS1 AGC target	Max Injection Time (ms)	MS1 Scan Range	MS2 Detector	Resolution or Scan rate	Minimum AGC target	Max Injection Time (ms)	Loop count (N) or cycle time (s)	Isolation Window	Activation Type	Collision Energy	Charge States Included	Dynamic Exclusion (s)
127	3.00E+06	100	400-2000	orbitrap	17,500	5.00E+03	60	Top N 12, Top N 8	2	HCD	27	2,3,4	30
135	4.00E+05	50	375-1500	orbitrap	15,000	2.00E+05	30	TopN, 3sec	1.3	HCD	32	2,3,4,5,6	60
209	1.00E+06	100	300-1800	ion trap	rapid	1.00E+04	100	TopN 15	2	CID	35	>1	30
438	4.00E+05	50	380-1280	ion trap	normal rate	2.00E+04	150	2 s cycle	1.6	HCD	30	2,3,4,5,6,7,8	15
593	--	--	350-1800	orbitrap	--	--	--	20	--	CID	--	>1	30
652	1.00E+06	100	400-1400	orbitrap	35,000	5.00E+04	50	20	1.2	HCD	30	2,3,4,5	10
729	1.00E+06	25	300-1500	orbitrap	17,500	1.00E+05	50	20	1.8	HCD	27	2,3,4,5	30
774	8.00E+05	auto	375-1500	orbitrap	15,000	1.00E+05	50	3	1.2	HCD	30	2-7	60
811	5.00E+04	50	375-1575	orbitrap	17,500	5.00E+04	50	20	1.2	HCD	25	2,3,4,5	30

41
42

43

44 **Table S8.** Participant laboratory results: User provided results from diverse informatic pipelines.

45 NA – not available. Multiple values reported if protein groupings were used, based on the output

46 formats and protein inference methods of the various informatic pipelines used.

Lab	Total Unique Peptides	Protein IDs
127	22382	3520
135	9797	NA
209	2363	4359 / 1049
438	15903	5771
593	131	89
652	11979	2089
729	11204	4907
774	18859	5946
811	3515	NA

47

48 **Table S9.** Participant laboratory results using the single pipeline re-analysis. Raw data files
49 were processed SEQUEST HT and Scaffold resulted in these sums of total unique peptides,
50 total proteins, and protein groups.

Lab	Total Unique Peptides	Total Protein IDs	Protein Groups
135	9600	3919	3533
209	3354	1586	1461
438	15646	6221	5621
593	0	0	0
652	9106	3518	3189
729	6626	3522	3202
774	16500	5676	5111
811	14	12	12
127	12615	5080	4595

51

52 **Table S10.** Participant laboratory results passed through the single pipeline re-analysis, using
53 alternate chromatographic techniques. Raw data files were processed SEQUEST HT and
54 Scaffold resulted in these sums of total unique peptides, total proteins, and protein groups.

Lab	Total Unique Peptides	Total Protein IDs
Alt-1 (12h run)	7060	2832
Alt-2 (2D)	18477	7765
Alt-3 (2D)	5852	2746

55

56 **Table S11.** Informatic intercomparison study: anonymous laboratory identification numbers,

57 software used, and results. NA – not available.

58

ID	Software	Unique Peptides Oceans 8	Unique Peptides Oceans 11
109	Peaks Studio X	4522	4898
321	SearchGUI / Peptide Shaker	2768	7389
321	MaxQuant	3342	4751
362	X!Tandem / SearchGUI	4890	8079
458	SEQUEST-HT / Percolator	6369	8288
501	MSGF+ OpenMS	4025	7463
828	SEQUEST-HT PD	NA	NA
902	SEQUEST-HT / Percolator	4653	7649
932	MASCOT	1724	3019
957	MsFragger / PeptideProphet / ProteinProphet	3687	6144

59

60

61

62

Table S12. Summary Table of Laboratory Intercomparison Results

	Lab_127	Lab_135	Lab_209	Lab_438	Lab_652	Lab_729	Lab_774	Average	Std Dev	Sum 7 SC
Sum of Spectral Counts (SC)	73828	38784	63198	126642	69677	53166	70606	70843	27455	495901
Number of Peptide IDs	12615	9600	3354	15646	9106	6626	16500	10492	4757	--
Number of Protein IDs	5080	3919	1586	6221	3518	3522	5676	4217	1574	--
Number of Protein Groups	4595	3533	1461	5621	3189	3202	5111	3816	1411	--
								Average*	Std Dev*	
Average Shared Peptides (pairwise 7 labs)	2821.0	2422.8	1304.2	2945.0	2325.7	2241.5	2769.2	2404	554	--
Average R2 (pairwise 7 labs)*	0.708	0.586	0.589	0.713	0.652	0.604	0.583	0.63	0.06	--
Average Slope (pairwise 7 labs)	1.099	3.014	1.617	1.386	1.297	1.028	0.710	1.45	0.75	--
Dynamic Exclusion Time (s)	30	60	30	15	10	30	60	33.57	19.73	--

*average and standard deviation of all pairwise comparisons

63

Figure S1. Results from user submitted data reports for laboratory intercomparison. a) Total number of unique peptide identifications by laboratory. A total of 35715 unique peptides were detected across all six laboratories. Note any peptides with PTMs were removed and not counted. b) Total number of protein identifications, note that some laboratory groups did not provide protein results (135 and 811). c) Pairwise comparisons of shared peptides between six laboratories ranged from 3844 to 10877 and averaged 7142 +/- 2074 identified peptides, demonstrating reproducibility of peptides identifications between laboratories. Note that PTMs were not taken into account for the uniqueness of peptides.

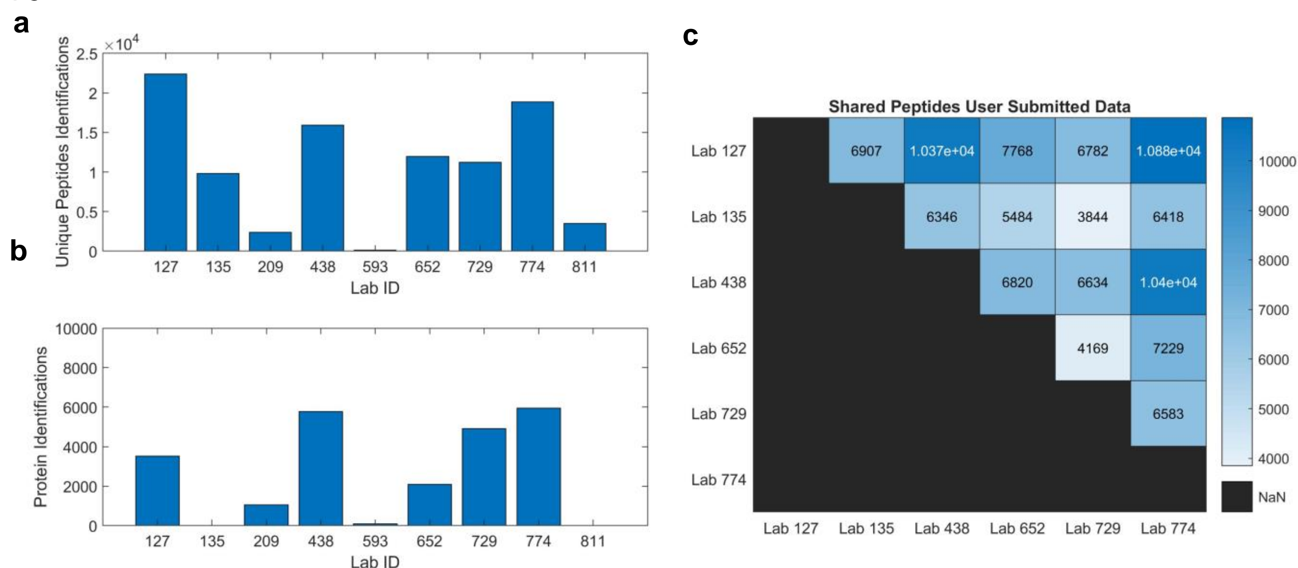
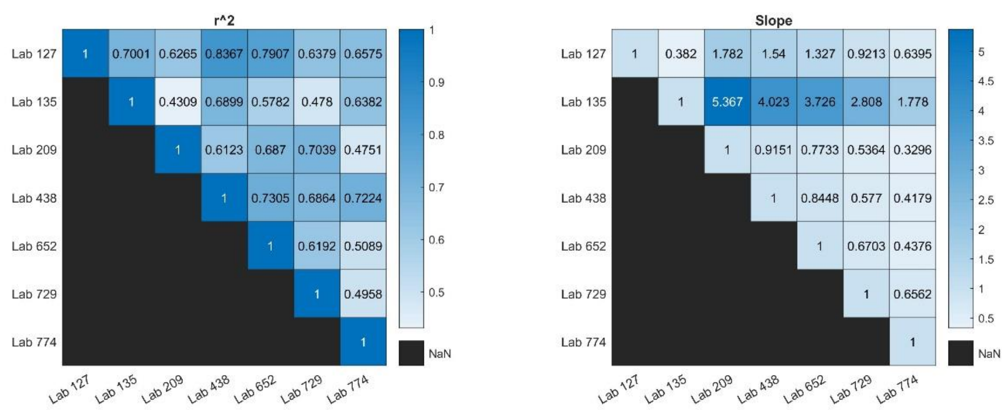


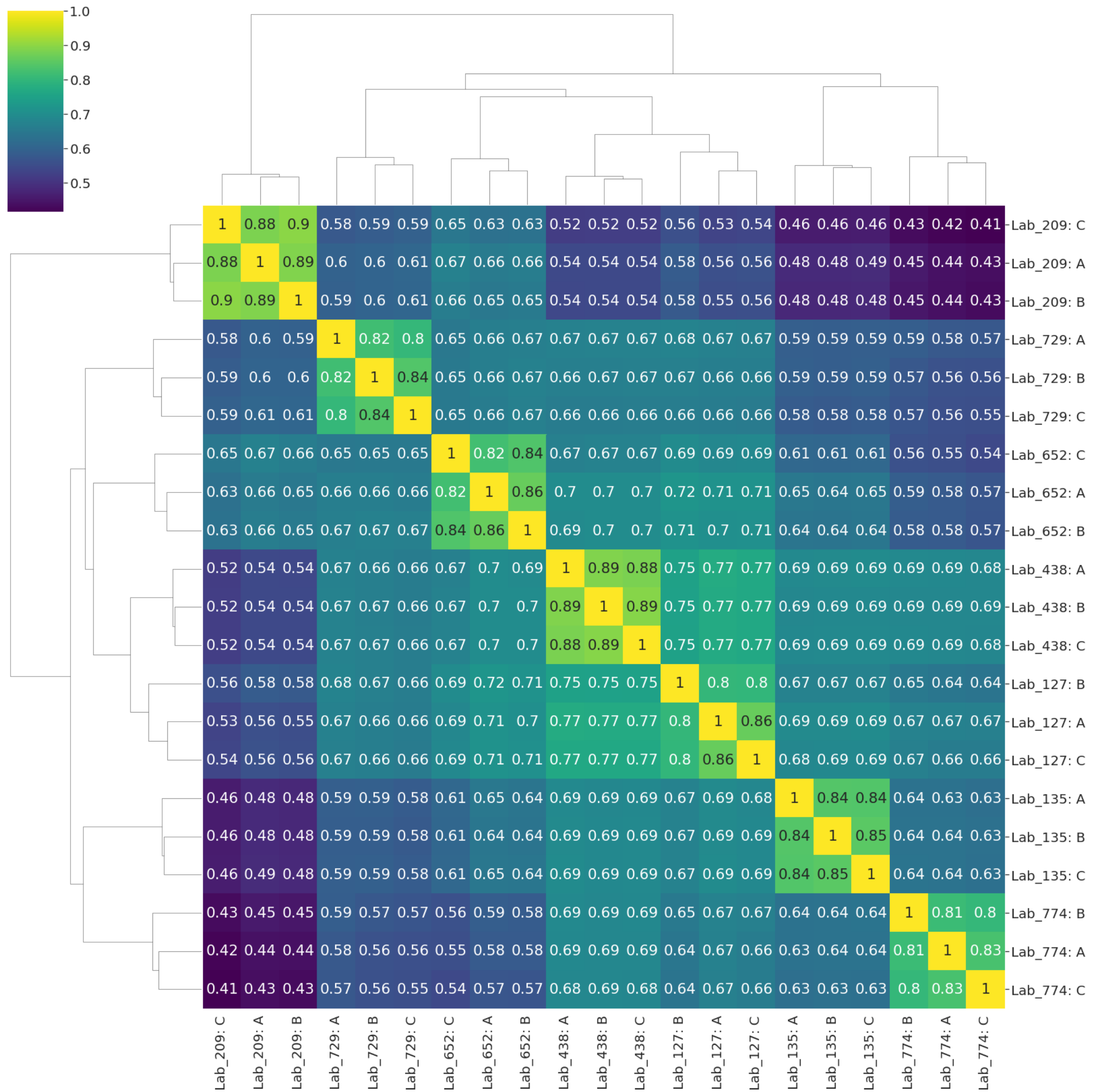
Figure S2. Results of pair-wise two-way linear regression analyses for re-analysis of submitted raw data from laboratory intercomparison, corresponding to Figure 4.



88 **Figure S3. Quantitative Sørensen similarity analysis.** Sørenson similarity analysis on full

89 protein dataset. See Fig. 6 for analysis of top 1000 proteins.

90



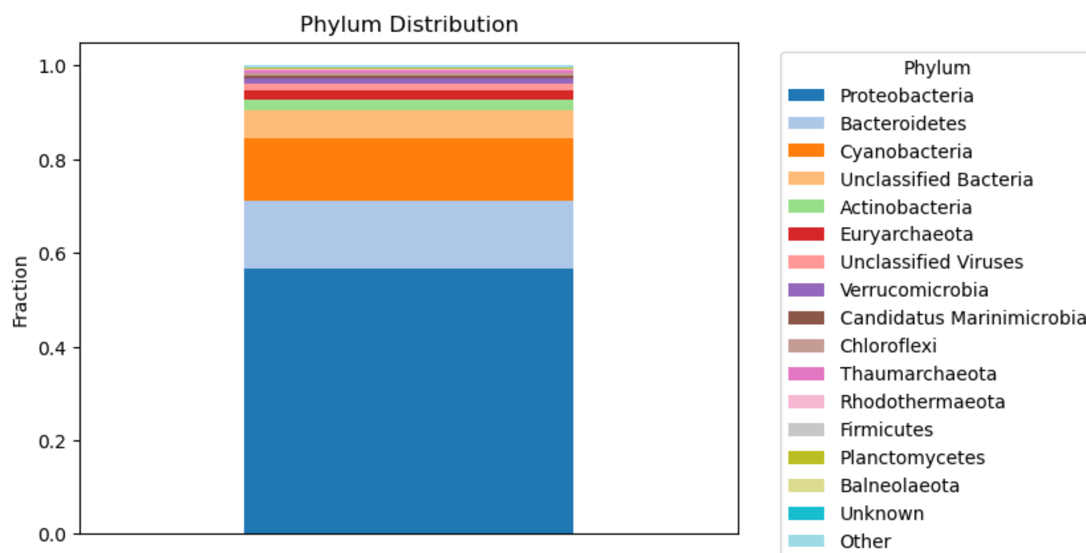
91

92

93

94

95 **Figure S4.** Phylum distribution within metagenomic annotations with sum of each taxa as a
96 fraction of all annotated genes.
97



98