1   **Results from a Multi-Laboratory Ocean Metaproteomic Intercomparison:**

2   **Effects of LC-MS Acquisition and Data Analysis Procedures**

3   *Participants of the Ocean Metaproteome Intercomparison Consortium:*

4   Mak A. Saito[*1], Jaclyn K. Saunders[1+], Matthew R. McIlvin[1], Erin M. Bertrand[2], John A. Breier[3],
5   Margaret Mars Brisbin[1], Sophie M. Colston[4], Jaimee R. Compton[4], Tim J. Griffin[5], W. Judson
6   Hervey[4], Robert L. Hettich[6], Pratik D. Jagtap[5], Michael Janech[7], Rod Johnson[8], Rick Keil[9], Hugo
7   Kleikamp[10], Dagmar Leary[4], Lennart Martens[17,18], J. Scott P. McCain[2,11], Eli Moore[12], Subina
8   Mehta[5], Dawn M. Moran[1], Jacqui Neibauer[7], Benjamin A. Neely[13], Michael V. Jakuba[1], Jim
9   Johnson[5], Megan Duffy[7], Gerhard J. Herndl[14], Richard Giannone[6], Ryan Mueller[15], Brook L.
10   Nunn[9], Martin Pabst[9], Samantha Peters[6], Andrew Rajczewski[5], Elden Rowland[2], Brian
11   Searle[16],Tim Van Den Bossche[17,18], Gary J. Vora[4], Jacob R. Waldbauer[19], Haiyan Zheng[20],
12   Zihao Zhao[14]
13
14   [1]Woods Hole Oceanographic Institution, Woods Hole, MA, USA
15   [2]Department of Biology, Dalhousie University, Halifax, NS, Canada
16   [3]The University of Texas Rio Grande Valley, Edinburg, TX
17   [4]Center for Bio/Molecular Science & Engineering, Naval Research Laboratory, Washington, DC, USA
18   [5]University of Minnesota at Minneapolis, Minneapolis, Minnesota, USA
19   [6]Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
20   [7]College of Charleston, Charleston, South Carolina, USA
21   [8]Bermuda Institute of Ocean Sciences, Bermuda
22   [9]University of Washington, Seattle, Washington, USA
23   [10]Department of Biotechnology, Delft University of Technology, Netherlands
24   [11]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA
25   [12] United States Geological Survey, USA
26   [13]National Institute of Standards and Technology, Charleston, South Carolina, USA
27   [14]University of Vienna, Dept. of Functional and Evolutionary Ecology, Austria
28   [15]Oregon State University, Corvallis, Oregon, USA
29   [16]Ohio State University, Columbus, Ohio, USA
30   [17]Department of Biomolecular Medicine, Faculty of Medicine and Health Sciences, Ghent University, 9000
31   Ghent, Belgium
32   [18]VIB – UGent Center for Medical Biotechnology, VIB, 9000 Ghent, Belgium
33   [19]Department of Geophysical Sciences, University of Chicago, Chicago, Illinois, USA
34   [20]Rutgers University, Piscataway, New Jersey, USA
35   [+]Present address: University of Georgia, Department of Marine Sciences
36   * corresponding author, msaito@whoi.edu
37
38

39

40

41 **Abstract**

42 Metaproteomics is an increasingly popular methodology that provides information regarding the

43 metabolic functions of specific microbial taxa and has potential for contributing to ocean ecology

44 and biogeochemical studies. A blinded multi-laboratory intercomparison was conducted to

45 assess comparability and reproducibility of taxonomic and functional results and their sensitivity

46 to methodological variables. Euphotic zone samples from the Bermuda Atlantic Time-Series

47 Study in the North Atlantic Ocean collected by *in situ* pumps and the AUV *Clio* were distributed

48 with a paired metagenome, and one-dimensional liquid chromatographic data dependent

49 acquisition mass spectrometry analyses was stipulated. Analysis of mass spectra from seven

50 laboratories through a common informatic pipeline identified a shared set of 1056 proteins from

51 1395 shared peptides constituents. Quantitative analyses showed good reproducibility: pairwise

52 regressions of spectral counts between laboratories yielded R^2 values ranging from 0.43 to 0.83,

53 and a Sørensen similarity analysis of the top 1,000 proteins revealed 70-80% similarity between

54 laboratory groups. Taxonomic and functional assignments showed good coherence between

55 technical replicates and different laboratories. An informatic intercomparison study, involving 10

56 laboratories using 8 software packages successfully identified thousands of peptides within the

57 complex metaproteomic datasets, demonstrating the utility of these software tools for ocean

58 metaproteomic research. Future efforts could examine reproducibility in deeper metaproteomes,

59 examine accuracy in targeted absolute quantitation analyses, and develop standards for data

60 output formats to improve data interoperability. Together, these results demonstrate the

61 reproducibility of metaproteomic analyses and their suitability for microbial oceanography

62 research including integration into global scale ocean surveys and ocean biogeochemical

63 models.

64

1. Introduction

65

66 Microorganisms within the oceans are major contributors to global biogeochemical cycles,

67 influencing the cycling of carbon, nitrogen, phosphorus, sulfur, iron, cobalt and other elements.

68 'Omic methodologies can provide an expansive window into these communities, with genomic

69 approaches characterizing the diversity and potential metabolisms, and transcriptomic and

70 proteomic methods providing insights into expression and function of that potential. Of these,

71 proteomics is increasingly being applied to natural ocean environments—when applied to

72 complex communities with diverse taxa present, the technique is commonly referred to as

73 metaproteomics (Wilmes and Bond, 2006). Metaproteomic samples contain an extraordinary

74 level of complexity relative to single organism proteomes (at least 1-2 orders of magnitude) due

75 to the simultaneous presence of many different organisms in widely varying abundances

76 (McCain and Bertrand, 2019). In particular, ocean metaproteome samples are significantly more

77 complex than the human proteome, the latter of which is itself considered to be a highly

78 complex sample (Saito et al., 2019). Proteomics provides a perspective distinct from other

79 'omics methods: as a direct measurement of cellular functions it can be used to examine the

80 diversity of ecosystem biogeochemical capabilities, to determine the extent of specific nutrient

81 stressors by measurement of transporters or regulatory systems, to determine cellular resource

82 allocation strategies in-situ, estimate biomass contributions from specific microbial groups, and

83 even to estimate potential enzyme activity (Bender et al., 2018; Bergauer et al., 2018; Cohen et

84 al., 2021; Fuchsman et al., 2019; Georges et al., 2014; Hawley et al., 2014; Held et al., 2021;

85 Leary et al., 2014; McCain et al., 2022; Mikan et al., 2020; Moore et al., 2012; Morris et al.,

86 2010; Saito et al., 2020; Sowell et al., 2009; Williams et al., 2012). The functional perspective

87 that proteomics allows is often complementary to metagenomic and metatranscriptomic

88 analyses. Moreover, the measurement of microbial proteins has improved greatly in recent

89   years, due to the advancements in nanospray-liquid chromatography and high-resolution mass

90   spectrometry approaches (Mueller and Pan, 2013; Ram et al., 2005; McIlvin and Saito, 2021).

91   With increasing interest in the measurement of proteins and their biogeochemical

92   functions within the oceans, the metaproteomic datatype is beginning to establish itself as a

93   valuable research and monitoring tool. However, given rapid changes in technology and

94   methods, as well as the overall youth of the metaproteomic field, demonstrating the

95   reproducibility and robustness of metaproteomic measurements to microbial ecology and

96   oceanographic communities is an important goal. This is particularly true as applications for

97   metaproteomics expand in research and monitoring of the changing ocean environment, for

98   example in global scale efforts such as the developing BioGeoSCAPES program

99   (www.biogeoscapes.org; (Tagliabue, 2023)), which aims to characterize the ocean metabolism

100   and nutrient cycles on a changing planet. As a result, there is a pressing need to assess inter-

101   laboratory consistency, and to understand the impacts of sampling, extraction, mass

102   spectrometry, and informatic analyses on the biological inferences that can be drawn from the

103   data.

104   There have been efforts to conduct intercomparisons of metaproteomic analyses in both

105   biomedical and environmental sample types in recent years that provide precedent for this

106   study. A recent community best practice effort in ocean metaproteomics data-sharing also

107   identified major challenges in ocean metaproteomics research, including sampling, extraction,

108   sample analysis, informatics pipelines, and data sharing, and conducted a quantitative

109   assessment of sample complexity in ocean metaproteome samples (Saito et al., 2019). A

110   previous benchmark study, driven by the Metaproteomics Initiative (Van Den Bossche et al.,

111   2021), was the "Critical Assessment of Metaproteome Investigation study" (CAMPI) that

112   employed a laboratory-assembled microbiome and human fecal microbiome sample to

113   successfully demonstrate reproducibility of results between laboratories. CAMPI found

114   robustness in results across datasets, while also observing variability in peptide identifications

115   largely attributed to sample preparation. This observation was consistent with prior findings on

116   single organism samples that determined >70% of the variability was due to sample processing,

117   rather than chromatography and mass spectrometry (Piehowski et al., 2013). Finally, the

118   Proteomics Informatics Group (iPRG) from the Association of Biomolecular Resources Facilities

119   (ABRF) conducted a study examining the influence of informatics pipelines on metaproteomics

120   analyses that found consistency among research groups in taxonomic attributions (Jagtap et al.,

121   2023), and previous research has demonstrated the impact of database choices on final

122   functional annotations and biological implications (Timmins-Schiffman et al., 2017).

123   Here we describe the results from the first ocean metaproteomic intercomparison. In this

124   study, environmental ocean samples were collected from the euphotic zone of the North Atlantic

125   Ocean and partitioned into subsamples and distributed to an international group of laboratories

126   (Fig. 1). The study was designed to examine inter-laboratory consistency rather than maximal

127   capabilities, stipulating one-dimensional chromatographic analyses from each laboratory (with

128   optional deeper analysis). Users were invited to use their preferred extraction, analytical, and

129   informatic procedures. The effort focused on the data dependent analysis (DDA) methods that

130   are currently common in ocean and other environmental and biomedical metaproteomics and

131   associated spectral abundance units of relative quantitation, which have been shown to be

132   reproducible in metaproteomics (Kleiner et al., 2017; Pietilä et al., 2022). Blinded results were

133   submitted, compared and discussed at a virtual community workshop in September of 2021. An

134   additional informatic pipeline comparison study was also conducted where participants were

135   provided metaproteomic raw data and associated metagenomic sequence database files and

136   were encouraged to use the informatic pipeline of their choice.

137   **2. Methods**

2.1 Sample Collection and Metadata

138

139   Ocean metaproteome filter samples for the wet lab comparison (Figure 1) were collected

140   at the Bermuda Atlantic Time-series Study (31° 40'N 64 ° 10'W) on expedition BATS 348 on

141   June 16th, 2018, between 01:00 and 05:00 am local time. *In situ* (underwater) large volume

142   filtration was conducted using submersible pumps to produce replicate biomass samples at a

143   single depth in the water column for intercomparisons. All filter subsamples are matched for

144   location, time, and depth. To collect the samples, two horizontal McLane pumps were clamped

145   together (Figure 1c) and attached at the same depth (80 m) with two filter heads (Mini-MULVS

146   design) on each pump and a flow meter downstream of each filter head. Each filter head

147   contained a 142 mm diameter 0.2 μm pore-size Supor (Pall Inc.) filter with an upstream 142 mm

148   diameter 3.0 μm pore-size Supor (Figure 1b, d). Only the 0.2 – 3.0 μm size fraction was used in

149   this study. The pumps were set to run for 240 min at 3 L per min. Volume filtered was measured

150   by three gauges on each pump, one downstream of each pump head, and one on the total

151   outflow (Table S2). Individual pump head gauges summed to the total gauge for pump 1 (within

152   1 L; 447 L and 446.2 L), but deviated by 89 L on pump 2 (478 L and 388.9 L). Given that the

153   total gauge is further downstream, we report the pump head gauges as being more accurate.

154   The pump heads were removed from the McLane pumps immediately upon retrieval,

155   decanted of excess seawater by vacuum, placed in coolers with ice packs, and brought into a

156   fabricated clean room environment aboard the ship. The 0.2 μm pore-size filters were cut in

157   eight equivalent pieces and frozen at -80°C in 2 mL cryovials, creating 16 samples per pump

158   that were co-collected temporally and in very close proximity (<1 m) to each other for a total of

159   32 samples used in this study (Figure 1d). The 3.0 μm pore-size filters are not included in this

160   study but are archived for future efforts. The sample naming scheme associated with the

161   different pumps and pump heads is described in Table S2. Note that pump 1A and 1B samples

162   accidentally had two 3.0 μm filters superimposed above the 0.2 μm filter, and 1B had a small

163   puncture in it, although neither of these seemed to affect the biomass collected, presumably the

164   puncture occurred after sampling was completed.

165   Samples for the informatic component were collected by the autonomous underwater

166   vehicle *Clio*. The vehicle and its sampling characteristics were used as previously described

167   (Breier et al., 2020; Cohen et al., 2023). Specifically, samples Ocean-8 and Ocean-11 were

168   also collected from the BATS station on R/V *Atlantic Explorer* expedition identifier AE1913 (also

169   described as BATS validation track BV55 32.75834º N 65.7374º W). The samples were

170   collected by autonomous underwater vehicle (AUV) *Clio* on June 19th 2019, dive Clio020, with

171   samples collected at 20 m (Ocean-11) and 120 m (Ocean-8) with 66.6 L and 92.6 L filtered,

172   respectively, used for this study. These samples were analyzed by 1D DDA analysis using

173   extraction and mass spectrometry for laboratory 438 (Tables S5-S7). Sample metadata for both

174   arms of this intercomparison study and corresponding repository information is provided in

175   Table S3 and repository links are in the Data Availability Statement.

176   *2.2 Metagenomic Extraction, Sequencing, and Assembly*

177   A metagenomic (reference sequence) database was created for peptide to spectrum

178   matching (PSMs) for the metaproteomic studies using a 1/8th sample split from the exact

179   sample used in the intercomparison as described above. Samples were shipped on dry ice to

180   the Naval Research Laboratory in Washington D.C. (USA), where DNA was extracted and

181   sequenced. Preserved filters were cut into smaller pieces using a sterile blade and placed into a

182   PowerBead tube with a mixture of zirconium beads and lysis buffer (CD1) from the Dneasy

183   PowerSoil Pro kit (Qiagen, Hilden Germany). The bead tube with filter sample was heated at

184   65°C for 10 min then placed on a vortex adapter and vortexed at maximum speed for 10 min.

185   After sample homogenization/lysis, the bead tube was centrifuged at 16 k x g for 2 min. The

186   supernatant was transferred to a DNA LoBind tube and processed using the manufacturer's

7

187   recommendations. The purified DNA was further concentrated by adding 10 µL 3 M NaCl and

188   100 µL cold 100% ethanol. The sample was incubated at -30°C for 1 hour, followed by

189   centrifugation at 16 k x g for 10 min. The supernatant was removed and precipitated DNA was

190   air-dried and resuspended in 10 mM Tris. DNA concentration was quantified with the Qubit

191   dsDNA High Sensitivity assay (Thermo Fisher Scientific, Waltham, MA, USA) and DNA quality

192   was assessed using the NanoDrop (ThermoFisher) and gel electrophoresis. Processing controls

193   included reagent only and blank filter samples.

194   Sequencing libraries were created from purified sample DNA using the IonExpress Plus

195   gDNA Fragment Library Preparation kit (Thermo Fisher) for a 200 bp library insert size. No

196   amplification of the library was required as determined by qPCR using the Ion Library TaqMan

197   Quantitation Kit. A starting library concentration of 100 pM was used in template generation and

198   chip loading with the Ion 540 Kit on the Ion Chef instrument prior to single-end sequencing on

199   the S5 benchtop sequencer.

200   Sequencing used a mix of Ion Torrent and Oxford Nanopore sequencing and resulting

201   sequencing reads were assembled using SPAdes v. 3.13.1 with Python v. 3.6.8. Following

202   metagenome assembly, contigs smaller than 500 bases were discarded. Open reading frame

203   (ORF) calling was performed on contigs 500 bps or longer using Prodigal v. 2.6.3 (Hyatt et al.,

204   2010) run with metagenomic settings as well as MetaGeneMark by submitting to the

205   MetaGeneMark server (http://exon.gatech.edu/meta_gmhmmp.cgi) using GeneMark.hmm

206   prokaryotic program v. 3.25 on August 11, 2019. ORFs called from both programs were

207   combined and made non-redundant using in-house Python scripts that utilize BioPython v. 1.73.

208   Non-redundant ORFs were annotated using the sequence alignment program DIAMOND (v 0.9.29)

209   with the NCBI nr database (downloaded 12/17/2019). ORFs were also annotated with InterProScan

210   (v 5.29) and with GhostKOALA (Kanehisa et al., 2016) (submitted to server 1/2/2020). Taxonomy

211 lineages were generated by using the best DIAMOND (Buchfink et al., 2015) hit and pulling lineage

212 information from NCBI Taxonomy database using BioPython v. 1.73.

213 *2.3 Proteomic methodologies: Extraction, instrumentation, and informatics*

214   Some basic protocol stipulations were provided to study participants regarding analytical

215 conditions to set a uniformity of experimental design. While users were encouraged to use the

216 extraction method of their preference, constraints on chromatography and mass spectrometry

217 conditions were set, limiting the number of chromatographic dimensions to one (1D), the total

218 length of the chromatographic run, the amount of protein injected (as proteolytic digests), and a

219 single mass spectrometry injection rather than gas phase fraction approaches (Table S4). Each

220 laboratory group's specific approach is summarized in the supplemental methods, with

221 extraction in Table S5, and chromatography and mass spectrometry equipment and parameters

222 in Tables S6 and S7. While there are more sophisticated methods such as two-dimensional

223 (2D) chromatography and gas phase fractionations that have been demonstrated to provide

224 deeper metaproteomes (McIlvin and Saito, 2021), these often require specialized equipment

225 and/or additional instrument time. As a result, the study constraints were provided to ensure a

226 single simple method that all labs could utilize. Laboratories were invited to submit additional

227 data from more complex analytical setups if they first completed the 1D analyses. Methods used

228 for the informatics intercomparison study are also presented within the Supplemental Materials.

229

230 *2.4 Compilation, analysis, and re-analysis of laboratory data submissions*

231   Results from individual laboratories' data submissions were analyzed in two ways as

232 shown in the flowchart of Figure 1a. First, submitted processed data reports (i.e. PSMs,

233 taxonomic, functional annotations) were compiled and interpreted. Second, raw data files (i.e.

234 spectra directly from instruments) from each group were put through a single informatic pipeline

235 using SEQUEST HT/Percolator within Proteome Discoverer (Version 2.2.0.388, Thermo

236   Scientific) and Scaffold (Version 5.2.1, Proteome Software) to isolate variability associated with

237   bioinformatic processing. This re-analysis (*single pipeline re-analysis* hereon) allowed detailed

238   cross-comparisons of laboratory practices to assess the influence of the extraction and mass

239   spectrometry components. Specific parameters of the latter included: parent and fragment

240   tolerances of 10ppm and 0.02 Da, respectively, with fixed and variable modifications of +57 on

241   C (fixed), and +16 on M and +42 on Peptide N-Terminal (variable). Peptide and protein FDRs

242   were set to lower than 1.0% using a decoy database, with 1 minimum peptide per protein, and

243   the resulting peptide FDR was 0.1%. The database used for PSMs was

244   Intercal_ORFs_prodigal_metagenemark.fasta based on the metagenomic sequencing

245   described above with 197,824 protein entries. The protein in this re-analysis was conducted

246   within Scaffold using total spectral counts and allowing single peptides to be attributed to

247   proteins. In addition to the total number of protein identifications, the number of protein groups

248   identified by Scaffold was also provided. Each protein group represented proteins identified with

249   identical peptides, collapsed into a single protein entry with the highest probability and number

250   of spectral counts.

251

252   *2.5 Data analysis methods*

253   Several analyses were conducted using data from the single pipeline re-analysis. First,

254   pairwise comparisons of protein identifications were conducted using spectral abundance

255   reports produced in Scaffold, and loaded, analyzed and visualized in MATLAB (MathWorks Inc).

256   Two-way (independent) linear regressions were conducted using the script linfit.m. R^2 on the

257   seven datasets were averaged and their standard deviation calculated for shared proteins in

258   each dataset. Second, a Sørensen similarity (Sørensen, 1948) was calculated where a matrix

259   was generated that consisted of the unique proteins or peptides identified across all technical

260   replicates from the various labs with the relative abundance per replicate (% contribution of

261   each protein/peptide per technical replicate total). The Bray-Curtis dissimilarity pairwise distance

262   was calculated on this matrix using Python and the SciPy library (v. 1.4.1, (Virtanen et al.,

263   2020)) and then 1 – Bray-Curtis dissimilarity was calculated across the matrix to generate the

264   Sørensen pairwise similarity across all replicates. The resulting similarities per replicate were

265   clustered and visualized using the clustermap function in the Seaborn library (v. 0.10.0,

266   (Waskom, 2021)). Third, shared peptides and proteins were visualized using Upset plots, using

267   the R package UpSetR (Conway et al., 2017) to determine the number of unique peptide

268   sequences and annotated proteins in intersecting sets between all labs, all permutations of lab

269   subsets, and all lab pairs.

270

271   **3. Results**

272   *3.1 Experimental Design*

273   This ocean metaproteomic intercomparison consisted of two major activities: a

274   laboratory component, where independent labs processed identical ocean samples

275   simultaneously collected from the North Atlantic Ocean (Fig. 1a, see Section 2.1), and a

276   subsequent informatic component. Participating institutions and persons at those institutions are

277   listed in Table S1, with all participants also listed as co-authors. Both arms of the study were

278   conducted under blinded conditions, where correspondence with participants was conducted by

279   an individual not involved in either study, and submitted results and data were anonymized prior

280   to sharing with the consortium. The laboratory study involved two biomass-laden filter slices

281   being sent to each participating group for protein extraction, mass spectrometry, and informatic

282   analyses (see below). The informatic effort was independent of the laboratory effort and

283   involved the distribution and informatic analysis of two metaproteomic raw data files generated

284   from samples also from the North Atlantic Ocean upper water column (20m and 120m depths,

285   see Section 2.1). These files were distributed after labs had submitted their laboratory extracted

286   raw data files. The raw files from the informatic study were distinct from the samples used in the

287   laboratory intercomparison study to avoid any biases from groups that analyzed those samples

288   previously. Submitted results from both components were anonymized and assigned three-digit

289   lab identifiers generated randomly with laboratory and informatic results from the same lab

290   being assigned distinct identifiers.

291   We report results for two study components: Part 1 (Section 3.2) involves the data

292   generation intercomparison of distributed subsamples from the North Atlantic Ocean (Fig. 1;

293   Section 2.1). Part 2 (Section 3.3) was an informatic intercomparison, where metaproteomic raw

294   files were shared with participants and processed results were submitted. Both components

295   were conducted as blinded studies, where each dataset was assigned a three digit randomly

296   generated identifier, with those identifiers used throughout the Results and Discussion.

297

298   *3.2 Mass Spectrometry Data Generation Intercomparison*

299   Nine laboratories submitted raw and processed datasets from the analysis of the

300   distributed Atlantic Ocean field samples (Table S1). The processed data submissions were

301   heterogeneous in output formats, statistical approaches, and parameter definitions. Because of

302   the challenges of comparing data derived from different types of statistical approaches used for

303   peptide and protein identification and inference, as well as the varying output formats from

304   various software packages, the user-generated data submissions were difficult to compile and

305   compare, resulting in variability in the number of identifications depending on the statistical

306   approaches and thresholds applied. These results are further discussed in the Supplemental

307   Section (Figure S1, Table S8). Despite these challenges, an average of 7142 +/- 2074 peptides

308   were identified across the pairwise comparisons (Figure S1c), implying a consistency of peptide

309   identifications across participants. The variability in proteome depth reflected the combination of

310   differing parameters employed by software and laboratory approaches.

311   To remove this variability associated with user-selected informatic pipelines, a single

312   pipeline re-analysis of the submitted raw mass spectral data was conducted. Raw data files

313   were processed together within a single informatic pipeline consisting of SEQUEST-HT,

314   Percolator, and Scaffold software and evaluated to a false discovery rate threshold of < 1% (see

315   see Section 2.4). Two datasets were found to have had issues during extraction and analysis

316   that affected the results in both processed and raw data (Labs 593 and 811; Table S8). Notably

317   these two laboratories differed from the others in that they did not use SDS as a protein

318   solubilizing detergent (Table S5). This likely resulted in inefficient extraction of the bacteria that

319   dominated the sample biomass (e.g. picocyanobacteria and *Pelagibacter*) embedded within the

320   membrane filter slices. Further examination showed polyethylene glycol contamination of one

321   dataset (Lab 811) and low yield from sample processing and extraction from the other (Lab

322   593). As a result, those datasets were not included in the single pipeline re-analysis. The

323   standardized pipeline included calculations of shared peptides and proteins, quantitative

324   comparisons, and consistency of taxonomic and functional results.

325   The total number of peptide and protein identifications and PSMs in the single informatic

326   pipeline analysis varied by laboratory (Table S9), with unique peptides ranging by more than a

327   factor of 3 from 3,354 to 16,500, and with 27,346 total unique peptides identified across

328   laboratories. This variability was likely due to different extraction, chromatographic, and mass

329   spectrometry approaches used by each laboratory, resulting in a varying depth of

330   metaproteomic results. Yet, as with the user-submitted results, there was considerable overlap

331   in identifications between all datasets. An intersection analysis found the numerous shared

332   peptides between all combinations of laboratories, with 1,395 peptides shared between all

333   seven laboratory datasets (Figure 2a). Laboratories with deeper proteomes shared numerous

334   peptides, for example the two laboratories with the most discovered unique peptides shared

335   ~3000 peptides between them. They also had the largest numbers of peptides that were not

336   found by any other labs (3617 and 2819, respectively). The fourth largest intersection size

337   (1395) represented the unique peptides discovered by all labs. Beyond that there were 12

338   different groupings of peptides that were shared among at least four laboratories. Consistent

339   with this, 3-way Venn diagrams of labs 135, 209 and 438 had an intersection of 2398 peptides,

340   labs 652, 729, and 774 showed 3016 peptides, and labs 127, 135, and 309 shared 2304

341   peptides (Figure 2d).

342   A similar analysis was conducted at the protein level, where the number of proteins

343   identified for each sample based on peptide mapping to the metagenome database (see

344   Section 2. Methods), identified 8,043 total unique proteins across all seven laboratories and

345   1,056 proteins shared amongst those laboratories as shown in the 7-way Venn diagram (Figure

346   2c). Three-way Venn diagram comparisons among labs 135, 209 and 438 had an intersection of

347   1,254 proteins, and labs 652, 729, and 774 shared 1,925 proteins (data not shown).

348   Optional deeper metaproteome results were submitted by three laboratories using either

349   a long gradient of 12 hours or 2 dimensional chromatographic methods (Table S10). The

350   number of discovered peptide and protein identifications were higher in each case, with as

351   many as 18477 unique peptides and 7765 protein identifications from an online 2-dimensional

352   chromatographic analysis from a 5 μg single injection.

353   The mapping of identified peptides to protein sequences forms the basis for protein

354   identifications in the form of DDA bottom-up proteomics employed here. The relationship

355   between peptides and protein identification was explored in Figure 3 and found to be correlated

356   by two-way linear regression with R^2 values of 0.97 and 0.98 for total protein identifications and

357   protein groups, respectively. This indicated that with deeper depth of analysis by some

358   laboratories, there was no fall off in the increase in protein identifications that might be attributed

359   to additional peptides mapping to already discovered protein sequences. In addition, the 2D and

360   long gradient additional analyses conducted by several laboratories fell upon this line consistent

361   with this "more peptides – more proteins" observation, implying more room for improvements in

362   depth of metaproteomic analyses.

363   A quantitative analysis of spectral counts from the wet lab re-analysis showed broad

364   coherence among the seven laboratories. Pairwise comparisons of protein spectral counts were

365   conducted for each of the seven labs against the other six (visualized in a 7x7 matrix, with

366   duplicate comparisons removed (e.g., A vs B and B vs A)), where each data point reflects the

367   spectral counts for a protein shared between laboratories (Figure 4a). When a dataset was

368   compared with itself a unity line of datapoints was observed along the diagonal axis. Two-way

369   linear regressions were conducted on each of these pairwise comparisons. The slopes ranged

370   from 0.33 to 5.5 (Figure S2), implying a varying dynamic range in spectral counts across

371   laboratories, likely due to variations in instrument parameterizations selected by each

372   laboratory. The coefficient of determination R^2 values from 0.43 to 0.73 showed coherence

373   among results for these large metaproteomic datasets (Figure 4b). To provide a sense of

374   coherence of each laboratory to the others, the R^2 values of a lab against the other six

375   laboratories were averaged and the standard deviation calculated. All of these average R^2

376   values were higher than 0.5, which showed overall quantitative consistency despite the size and

377   complexity of these datasets (Figure 4d).

378   A comparative taxonomic and functional analysis was also conducted using a single

379   informatic pipeline (see metagenomic sequencing methods for annotation pipeline). Lowest

380   common ancestor (LCA) analysis of peptides identified from datasets from seven laboratories

381   showed consistent patterns of taxonomic distribution using the MetaTryp package (Figure 5a;

382   (Saunders et al., 2020). Cyanobacteria and alphaproteobacteria were the top two taxonomic

383   groups in all laboratory submissions, consistent with the abundant picocyanobacteria

384   *Prochlocococcus* and the heterotrophic bacterium *Pelagibacter ubique* known to be dominant

385   components of the Sargasso Sea ecosystem (Sowell et al., 2009; Malmstrom et al., 2010).

386   Similarly, KEGG Orthology group (KO) analysis of those datasets also showed highly similar

387   patterns of protein distributions across laboratories (Figure 5b). Notably the PstS phosphate

388   transporter protein was the most abundant functional protein in all datasets, consistent with

389   observations of phosphorus stress in the North Atlantic oligotrophic gyre and its biosynthesis in

390   marine cyanobacteria (Scanlan et al., 1997; Coleman and Chisholm, 2010; Ustick et al., 2021).

391   These findings demonstrate the reproducibility in the primary functional and taxonomic

392   conclusions from the metaproteome datasets. Finally, a Sørensen similarity analysis of the

393   1,000 proteins with highest spectral counts revealed 70–80% similarities between most

394   laboratory groups in the data re-analysis (Figure 6). Similarity analyses on the full dataset (with

395   all peptides and proteins) revealed lower similarity at the peptide level than the protein level,

396   implying variability in peptide identification is ameliorated as it is aggregated to the protein level

397   (Figure S3).

398

399   *3.3. Informatic Data Analysis Intercomparison*

400   Two metaproteomic raw files were provided to intercomparison participants and were

401   searched with each laboratory's preferred database searching informatic pipeline. The samples

402   that generated the data for these files were collected by autonomous AUV *Clio* during a single

403   dive at the Bermuda Atlantic Time-series Study Station (Breier et al., 2020), and were distinct

404   from the samples associated with the laboratory intercomparison component. However, they

405   were also from the North Atlantic Ocean, allowing the same metagenomic database to be used.

406   This informatic study component was not launched until after the laboratory-based

407   intercomparison submission deadline to avoid influencing that part of the study by sharing

408   similar raw data. Samples were named Ocean 8 and Ocean 11 and were taken from 120 m and

409   20 m depths, respectively.

410   The informatic intercomparison involved 10 laboratories utilizing 8 different software

411   pipelines including the PSM search engines: SEQUEST, X!Tandem, MaxQuant, MSGF+,

412   Mascot, MSFragger, and PEAKS (Table S11). As with the user supplied laboratory results, the

413   results were challenging to compile due to different types of data outputs, approaches used in

414   protein inference, and statistical approaches applied within each pipeline. Unique peptide

415   discoveries served as a useful base unit of comparison that were less subject to these

416   comparison challenges. The number of peptides ranged from 1724 to 6369 in Ocean 8 and

417   3019 to 8288 in Ocean 11 (Figure 7; Table S11). The differences in the number of peptides was

418   likely due to parameters used in software, for example, laboratory 932 had the lowest number of

419   peptides identified in both samples, but also used a highly stringent 99.9% probability cutoff that

420   likely influenced this result.

421

422   **4. Discussion**

423   *4.1 Assessment of Ocean Metaproteomics Reproducibility*

424   Given the relatively recent establishment of ocean metaproteomic techniques as well as

425   their methodological complexity, intercomparisons of methods are important in demonstrating

426   the suitability of metaproteomic analyses in ocean ecological and biogeochemistry studies.

427   Synthesizing the results of the laboratory and mass spectrometry blinded intercomparison study

428   (Section 3.2) processed with a single informatic pipeline (Section 2.4), we observed consistent

429   reproducibility with regards to three attributes of ocean metaproteomics analyses: 1) the identity

430   of discovered peptides and proteins (Fig. 2), 2) their relative quantitation (Figs. 4 and 6), and 3)

431   the taxonomic and functional assignments within intercompared samples (Fig 5). With over

432   1000 proteins identified across seven laboratories and Sørensen similarity indexes typically

433   higher than 70–80% (Fig. 6), the results unambiguously demonstrate consistent detection and

434   quantitation of major proteins in the sample. Together these results provide confidence that

435   multiple laboratories can generate reproducible results describing the major proteome

436   composition of ocean microbiome samples, and in doing so can assess the functional

437   composition and biogeochemical significance of these complex microbial communities.

438   *4.2 Metrics in metaproteomics: Core versus rare "long tail" proteins*

17

https://doi.org/10.5194/egusphere-2023-3148
Preprint. Discussion started: 16 January 2024
© Author(s) 2024. CC BY 4.0 License.

439   While abundant proteins were consistently detected across seven laboratories'

440   submissions, there was substantial variability in the less abundant proteins (Fig. 2). This

441   simultaneous consistency in abundant proteins and diversity in rare proteins (and their

442   respective peptide constituents) was likely a result of several factors in the study design and

443   execution. First, the intercomparison experimental design stipulated 1D chromatography in

444   order to provide straightforward comparisons that all laboratories could accomplish. This

445   contributed to study consistency, but also resulted in lesser proteome depth compared to more

446   elaborate methods such as 2D chromatography and gas phase fractionation commonly in use.

447   Second, the sample complexity of ocean metaproteomes has been shown to be enormous, with

448   a far greater number of low abundance peptides present than HeLa human cell lines (Saito et

449   al., 2019). The combined effect of these factors meant that, while laboratories were able to

450   detect abundant proteins consistently, there was considerable stochasticity associated with the

451   detection of less abundant peptides resulting in a long tail of discovered lower abundance

452   proteins. This is evident in Figure 8, where most of the 1063 proteins across seven laboratories

453   in the re-analysis were in the upper half of proteins when ranked by abundance.

454   Mass spectrometer settings such as dynamic exclusion, chromatography conditions, and

455   variation in sample preparation methods all likely contributed to this stochastic variability in rare

456   peptide detection among laboratories. Moreover, while all participating laboratories used

457   Thermo orbitrap mass spectrometers, there were seven variants of instrument model, including

458   some with Tribrid multiple detector capability (Table S6). While testing other mass spectrometry

459   platforms is of interest, this trend of community orbitrap usage in this study is consistent with the

460   broader proteomics community, where currently 9 of the top 10 instruments used in

461   ProteomeXchange consortium repository data submissions utilize orbitraps as of the manuscript

462   submission date (Deutsch et al., 2019). When conducting analysis of environmental samples,

463   choices can be made about instrument setup and parameters based on the scientific objectives,

464 for example if maximal proteome depth or robust quantitation while using a discovery approach

465 is desired. Future intercalibration efforts enlisting more sensitive metaproteomic methods such

466 as 2D-chromatography (McIlvin and Saito, 2021), more sensitive instruments (Stewart et al.,

467 2023), and other emerging methods can greatly improve detection and quantitation of rarer

468 proteins in metaproteomes, allowing exploration of the depths of state-of-the-art capabilities

469 rather than our present emphasis on interlaboratory consistency. Moreover, the development

470 and adoption of best practices in sample collection, extraction, chromatographic separation,

471 mass spectrometry analyses, and informatic approaches will contribute to interlaboratory

472 consistency.

473   Despite the inter-laboratory variability in the detected sets of rarer peptides and proteins,

474 we interpret these to be largely robust identifications. The stringent 0.1% peptide-level FDR

475 threshold we use here is determined by scoring decoys: reverse sequenced peptides that are

476 not in our samples. Peptide assignments to these decoys model the score distribution of all

477 incorrect peptide-spectrum matches (PSMs) in our study such that FDRs can be estimated in an

478 unbiased way for each laboratory. However, these estimates are complicated by subtle

479 sequence diversity within a population's proteome, which is typically not considered by

480 proteomics software designed to analyze single species (Schiebenhoefer et al., 2019). This

481 diversity within metaproteomic samples results in the presence of highly similar peptides with

482 nearly identical precursor masses that produce many of the same b- and y-ions, and this

483 similarity is not well modeled by decoy peptides. The influence of microdiversity on

484 metaproteomics FDR estimation using strain-specific proteogenomic databases is an important

485 area of future exploration (Wilmes et al., 2008).

486 *4.3 Informatics Intercomparison Assessment*

487   The discovery of peptide constituents of proteins within a complex ocean metaproteomic

488   matrix was successful across all software packages tested (Figure 7), where the metric for

489   success is a comparable number of peptide identifications. This is a notable finding due to the

490   highly complex mass spectra, large number of chimeric peaks present (Saito et al., 2019), and

491   large database sizes involved in ocean metaproteomes. To our knowledge, some of these

492   software packages had not yet been applied to ocean metaproteomes. There was also

493   variability associated with the stringency of statistical parameters employed, which points to the

494   challenges in assembling datasets from multiple laboratories with different depth of proteome

495   identifications.

496   Despite the success of this intercomparison component across software packages, there

497   is likely considerable room for improvement in the future. As mentioned previously, ocean

498   samples are highly complex and there are likely additional peptides that remain unidentified

499   using current technology, due to low intensity peaks and co-elution with other peptides resulting

500   in the chimeric spectra. Significant improvements in depth of analysis can be achieved through

501   increased chromatographic sample separation and optimized (or alternative) mass spectrometry

502   data acquisition strategies. Yet there is room for informatic improvements as well: most DDA

503   database searching algorithms are unable to identify multiple peptides within a single

504   fragmentation spectrum. Moreover, when in DDA collection mode mass spectrometry software

505   typically does not isolate and fragment peptides that cannot be assigned a charge state, which

506   is a common occurrence for the low abundance peaks within ocean samples. As a result, there

507   is considerable room for improvements in informatic pipelines to discover additional peptides.

508   Although the application of data independent approaches (DIA) to oceanographic

509   metaproteomics analysis is currently limited (Morris et al., 2010), the systematic nature of ion

510   selection and fragmentation allows for a greater number of low abundant peptides to be

511   quantified. By avoiding the need to select precursor ions for fragmentation, DIA methods have

512   the promise to identify some of these rarer peptides, assuming enough ions can be isolated to

513   produce robust MS2 spectra, as the wider isolation windows often used in DIA will dilute

514   precursor ions within ion traps.

515   *4.4 Future Efforts in Ocean Metaproteomic Intercomparisons and Intercalibrations*

516   As the first interlaboratory ocean metaproteomics study, we chose to describe this study

517   as an intercomparison rather than an intercalibration and it served as a vehicle with which to

518   assess the extent of reproducibility. Future intercalibration efforts could aim to further assess

519   and improve upon the level of accuracy, reproducibility, and standardization of ocean

520   metaproteome measurements. As mentioned above, development of best practices associated

521   with sample collection, extraction, and analysis would be valuable, while also encouraging

522   methodological improvements and backward compatibility through the use of reference

523   samples. Alternative modes of data collection could also be tested in future interlaboratory

524   comparisons, including parallel reaction monitoring mode (PRM), multiple reaction monitoring

525   mode (MRM), quantification using isotopic labeling or tagging, and DIA methods. PRM and

526   MRM methods allow sensitive targeted measurements of absolute quantities of peptides (e.g.

527   copies per liter of seawater in the ocean context). As many 'omics methodologies applied in

528   environmental settings operate in relative abundance modes, adding the ability to measure

529   absolute quantities would be particularly valuable for comparisons of environments across

530   space and time. Targeted metaproteomic methods have been deployed in marine studies using

531   stable isotope labeled peptides for calibration, achieving femtomoles per liter of seawater

532   estimates of transporters, regulatory proteins, and enzymes (Saito et al., 2020; Bertrand et al.,

533   2013; Saito et al., 2014, 2015; Joy-Warren et al., 2022; Wu et al., 2019). These methods are not

534   yet widely adopted, but with growing interest could be deployed to other laboratories and

535   incorporated into future iterations of intercomparison and intercalibration studies. DIA also has

536   great potential in ocean metaproteome studies and is increasingly being deployed in laboratory

537   and field studies of marine systems. Similar to this DDA intercomparison, the methodological

538   and informatic challenges of DIA could be explored during intercomparisons of analyses of

539   ocean samples. Finally, as mentioned above, all participants of this study used orbitrap mass

540   spectrometers for DDA submissions, but new instrumentation such as trapped ion mobility

541   spectrometry time of flight mass spectrometers (timsTOF) may be applied to ocean

542   metaproteome analyses and would be important to intercompare with orbitrap platforms.

543   As noted above, there were also challenges in collating and comparing data outputs

544   from various software, as well as variation in how those programs conducted protein inference.

545   For example, peptide-level data from different research groups were reported as either

546   unmodified peptide sequences or as various peptide analytes (where modifications and charges

547   states were included with the peptide sequence), making compilation of peptide reports difficult.

548   Similarly, at the protein level reported proteins could be counted either before or after protein

549   grouping, e.g. applying Occam's-razor logic to peptide groupings into proteins – the former

550   reflecting the set of all proteins in the database that could be in the sample, the latter the

551   minimum set required to explain the peptide data. Such issues will also contribute to challenges

552   in integration and assembly of data from different laboratories for large ocean datasets. While

553   best practices for metadata and data types have been described by the community (Saito et al.,

554   2019), there continues to be a need for standardization of data output formats for

555   metaproteomic results, similar to the metadata standard recently put forward in the human

556   proteome field (Dai et al., 2021).

557   *4.5 Metaproteomics in Global Ocean Surveys*

558   Understanding how the oceans are responding to the rapid changes driven by human

559   alteration of ecosystems is a high priority. Ocean and environmental sciences have a long

560   history of chemical measurements that are critical to assessing ecosystems and climatic

22

561   change. Such measurements have been straightforward for discrete measurements, such as

562   temperature, pH, chlorophyll, phosphate, dissolved iron and numerous other variables. When

563   collected over large spatial (ocean basin) or temporal (seasonal or decadal spans) scales, these

564   datasets have been powerful in identifying major (both cyclical and secular) changes. 'Omics'

565   measurements represent a more complex data type where each discrete sample can generate

566   thousands (if not more) of units of information. This study demonstrates the power and potential

567   for collaborative metaproteomics studies to identify key functional molecules and relate them to

568   their taxonomic microbial sources within the microbiome from multiple lab groups. Moreover,

569   multi-lab metaproteomics results in vastly enhanced identification of low abundance proteins

570   that are not identified by all research groups. Such low abundance proteins can be more likely

571   to change in abundance with changing environmental conditions and nutrient limitations,

572   resulting in a more nuanced and richer investigation of marine microbial ecology and

573   biogeochemistry with collaborative metaproteomics research. The implementation of such

574   voluminous data is beginning to be applied on larger scales and holds great promise in

575   improving not only our understanding of the functioning of the current system, but also the way

576   we assess how environments are changing with continued human perturbations.

577   Intercomparison and intercalibration are critical activities to undertake in order to allow

578   comparison of 'omics results across time and space dimensions. With major programs

579   underway and being envisioned such as the BioGEOTRACES, AtlantECO, Bio-GO-SHIP, and

580   BioGeoSCAPES efforts, the imperative for such intercalibration has grown and the need for best

581   practices is urgent. This Ocean Metaproteomic Intercomparison study is a valuable step in

582   assessing metaproteomic capabilities across a number of international laboratories,

583   demonstrating a clear consistency in measurement capability, while also pointing to the

584   potential for continued community development of metaproteomic capacity and technology.

585

586 *Author Contributions* **:** MAS and MRM obtained OCB workshop support and drafted the

587 experimental design with feedback from BN, MJ, and DL acting as the Advisory Committee. SC,

588 JH, DL, GJV, and JKS conducted the metagenomic analyses and assembly. JKS, MAS, MMB,

589 MRM, and RM conducted data analysis and visualization. MRM, MAS, JAB, MVJ, and RJ

590 conducted sample collection and/or AUV Clio operations. MAS, JKS, MRM, EMB, SC, JRC, TG,

591 JH, RLH, PJ, MJ, RK, HK, DL, JSPM, EM, SM, DMM, JN, BN, JJ, MD, GJH, RG, RM, BLN, MP,

592 SP, AR, ER, BS, TVDB, JRW, HZ, and ZZ contributed mass spectrometry data and/or

593 informatics data for the intercomparison. JKS anonymized data submissions and conducted

594 follow-up correspondence about methods. The manuscript was drafted by MAS and all authors

595 contributed to the writing and editing.

596

597 *Data and Code Availability :* The raw files for this project summarized in Table S3 are available

598 at ProteomeXchange and PRIDE repository with the dataset identifier PXD043218 and

599 10.6019/PXD043218. Access for reviewers is available using the username**:**

600 reviewer_pxd043218@ebi.ac.uk and password**:** uSxVkRza, and

601 reviewer_pxd044234@ebi.ac.uk and password: Evvgced0 Co-located information about these

602 datasets are available at the Biological and Chemical Data Management Office under project

603 765945 (https://www.bco-dmo.org/project/765945) and at the BATS page (https://www.bco-

604 dmo.org/project/2124). The metagenomic reads are listed under Bioproject Accession:

605 PRJNA932835; SRA submission: SUB12819843, available at link:

606 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA932835. The code for for upset visualization is

607 available at: https://maggimars.github.io/protein/PeptideUpSetR.html.

608

609 *Competing Interests* - The authors declare no competing financial interests.

610   *Supplemental Materials* - Methods for the informatic intercomparison study are available in the

611   Supplemental Methods. Supplemental Informational is available as Tables S1-S11, and Figures

612   S1-S3.

630

631

632

633 **References**

634 Bender, S. J., Moran, D. M., McIlvin, M. R., Zheng, H., McCrow, J. P., Badger, J., DiTullio, G.
635 R., Allen, A. E., and Saito, M. A.: Colony formation in Phaeocystis antarctica: connecting
636 molecular mechanisms with iron biogeochemistry, Biogeosciences, 15, 4923–4942, 2018.

637 Bergauer, K., Fernandez-Guerra, A., Garcia, J. A., Sprenger, R. R., Stepanauskas, R.,
638 Pachiadaki, M. G., Jensen, O. N., and Herndl, G. J.: Organic matter processing by microbial
639 communities throughout the Atlantic water column as revealed by metaproteomics, Proceedings
640 of the National Academy of Sciences, 115, E400–E408, 2018.

641 Bertrand, E. M., Moran, D. M., McIlvin, M. R., Hoffman, J. M., Allen, A. E., and Saito, M. A.:
642 Methionine synthase interreplacement in diatom cultures and communities: Implications for the
643 persistence of B12 use by eukaryotic phytoplankton, Limnology and Oceanography, 58, 1431–
644 1450, 2013.

645 Breier, J. A., Jakuba, M. V., Saito, M. A., Dick, G. J., Grim, S. L., Chan, E. W., McIlvin, M. R.,
646 Moran, D. M., Alanis, B. A., and Allen, A. E.: Revealing ocean-scale biochemical structure with a
647 deep-diving vertical profiling autonomous vehicle, Science Robotics, 5, eabc7104, 2020.

648 Buchfink, B., Xie, C., and Huson, D. H.: Fast and sensitive protein alignment using DIAMOND,
649 Nature methods, 12, 59–60, 2015.

650 Cohen, N. R., McIlvin, M. R., Moran, D. M., Held, N. A., Saunders, J. K., Hawco, N. J.,
651 Brosnahan, M., DiTullio, G. R., Lamborg, C., and McCrow, J. P.: Dinoflagellates alter their
652 carbon and nutrient metabolic strategies across environmental gradients in the central Pacific
653 Ocean, Nature Microbiology, 6, 173–186, 2021.

654 Cohen, N. R., Krinos, A. I., Kell, R. M., Chmiel, R. J., Moran, D. M., McIlvin, M. R., Lopez, P. Z.,
655 Barth, A., Stone, J., Alanis, B. A., Chan, E. W., Breier, J. A., Jakuba, M. V., Johnson, R.,
656 Alexander, H., and Saito, M. A.: Microeukaryote metabolism across the western North Atlantic
657 Ocean revealed through autonomous underwater profiling, Ecology,
658 https://doi.org/10.1101/2023.11.20.567900, 2023.

659 Coleman, M. L. and Chisholm, S. W.: Ecosystem-specific selection pressures revealed through
660 comparative population genomics, Proceedings of the National Academy of Sciences, 107,
661 18634–18639, 2010.

662 Conway, J. R., Lex, A., and Gehlenborg, N.: UpSetR: an R package for the visualization of
663 intersecting sets and their properties, Bioinformatics, 2017.

664 Dai, C., Füllgrabe, A., Pfeuffer, J., Solovyeva, E. M., Deng, J., Moreno, P., Kamatchinathan, S.,
665 Kundu, D. J., George, N., and Fexova, S.: A proteomics sample metadata representation for
666 multiomics integration and big data analysis, Nature communications, 12, 1–8, 2021.

667 Deutsch, E. W., Bandeira, N., Sharma, V., Perez-Riverol, Y., Carver, J. J., Kundu, D. J., García-
668 Seisdedos, D., Jarnuczak, A. F., Hewapathirana, S., Pullman, B. S., Wertz, J., Sun, Z., Kawano,
669 S., Okuda, S., Watanabe, Y., Hermjakob, H., MacLean, B., MacCoss, M. J., Zhu, Y., Ishihama,
670 Y., and Vizcaíno, J. A.: The ProteomeXchange consortium in 2020: enabling 'big data'
671 approaches in proteomics, Nucleic Acids Research, gkz984, https://doi.org/10.1093/nar/gkz984,
672 2019.

Fuchsman, C. A., Palevsky, H. I., Widner, B., Duffy, M., Carlson, M. C., Neibauer, J. A., Mulholland, M. R., Keil, R. G., Devol, A. H., and Rocap, G.: Cyanobacteria and cyanophage contributions to carbon and nitrogen cycling in an oligotrophic oxygen-deficient zone, The ISME journal, 13, 2714–2726, 2019.

Georges, A. A., El-Swais, H., Craig, S. E., Li, W. K., and Walsh, D. A.: Metaproteomic analysis of a winter to spring succession in coastal northwest Atlantic Ocean microbial plankton, The ISME journal, 8, 1301–1313, 2014.

Hawley, A. K., Brewer, H. M., Norbeck, A. D., Paša-Tolić, L., and Hallam, S. J.: Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes, Proceedings of the National Academy of Sciences, 111, 11395–11400, 2014.

Held, N. A., Sutherland, K. M., Webb, E. A., McIlvin, M. R., Cohen, N. R., Devaux, A. J., Hutchins, D. A., Waterbury, J. B., Hansel, C. M., and Saito, M. A.: Mechanisms and heterogeneity of in situ mineral processing by the marine nitrogen fixer Trichodesmium revealed by single-colony metaproteomics, ISME Communications, 1, 1–9, 2021.

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J.: Prodigal: prokaryotic gene recognition and translation initiation site identification, BMC bioinformatics, 11, 1–11, 2010.

Jagtap, P. D., Hoopmann, M. R., Neely, B. A., Harvey, A., Käll, L., Perez-Riverol, Y., Abajorga, M. K., Thomas, J. A., Weintraub, S. T., and Palmblad, M.: The Association of Biomolecular Resource FacilitiesProteome Informatics Research Group Study on Metaproteomics(iPRG-2020), J Biomol Tech, 34, 3fc1f5fe.a058bad4, https://doi.org/10.7171/3fc1f5fe.a058bad4, 2023.

Joy-Warren, H. L., Alderkamp, A.-C., van Dijken, G. L., J Jabre, L., Bertrand, E. M., Baldonado, E. N., Glickman, M. W., Lewis, K. M., Middag, R., and Seyitmuhammedov, K.: Springtime phytoplankton responses to light and iron availability along the western Antarctic Peninsula, Limnology and Oceanography, 67, 800–815, 2022.

Kanehisa, M., Sato, Y., and Morishima, K.: BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences, Journal of molecular biology, 428, 726–731, 2016.

Kleiner, M., Thorson, E., Sharp, C. E., Dong, X., Liu, D., Li, C., and Strous, M.: Assessing species biomass contributions in microbial communities via metaproteomics, Nature communications, 8, 1–14, 2017.

Leary, D. H., Li, R. W., Hamdan, L. J., Hervey IV, W. J., Lebedev, N., Wang, Z., Deschamps, J. R., Kusterbeck, A. W., and Vora, G. J.: Integrated metagenomic and metaproteomic analyses of marine biofilm communities, Biofouling, 30, 1211–1223, 2014.

Malmstrom, R. R., Coe, A., Kettler, G. C., Martiny, A. C., Frias-Lopez, J., Zinser, E. R., and Chisholm, S. W.: Temporal dynamics of Prochlorococcus ecotypes in the Atlantic and Pacific oceans, The ISME journal, 4, 1252–1264, 2010.

McCain, J. S. P. and Bertrand, E. M.: Prediction and consequences of cofragmentation in metaproteomics, Journal of proteome research, 18, 3555–3566, 2019.

27

712   McCain, J. S. P., Allen, A. E., and Bertrand, E. M.: Proteomic traits vary across taxa in a coastal
713   Antarctic phytoplankton bloom, The ISME journal, 16, 569–579, 2022.

714   McIlvin, M. R. and Saito, M. A.: Online Nanoflow Two-Dimension Comprehensive Active
715   Modulation Reversed Phase–Reversed Phase Liquid Chromatography High-Resolution Mass
716   Spectrometry for Metaproteomics of Environmental and Microbiome Samples, Journal of
717   proteome research, 20, 4589–4597, 2021.

718   Mikan, M. P., Harvey, H. R., Timmins-Schiffman, E., Riffle, M., May, D. H., Salter, I., Noble, W.
719   S., and Nunn, B. L.: Metaproteomics reveal that rapid perturbations in organic matter prioritize
720   functional restructuring over taxonomy in western Arctic Ocean microbiomes, The ISME journal,
721   14, 39–52, 2020.

722   Moore, E. K., Nunn, B. L., Goodlett, D. R., and Harvey, H. R.: Identifying and tracking proteins
723   through the marine water column: Insights into the inputs and preservation mechanisms of
724   protein in sediments, Geochimica et cosmochimica acta, 83, 324–359, 2012.

725   Morris, R. M., Nunn, B. L., Frazar, C., Goodlett, D. R., Ting, Y. S., and Rocap, G.: Comparative
726   metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy
727   transduction, The ISME journal, 4, 673–685, 2010.

728   Mueller, R. S. and Pan, C.: Sample handling and mass spectrometry for microbial
729   metaproteomic analyses, in: Methods in Enzymology, vol. 531, Elsevier, 289–303, 2013.

730   Piehowski, P. D., Petyuk, V. A., Orton, D. J., Xie, F., Moore, R. J., Ramirez-Restrepo, M., Engel,
731   A., Lieberman, A. P., Albin, R. L., and Camp, D. G.: Sources of technical variability in
732   quantitative LC–MS proteomics: human brain tissue sample analysis, Journal of proteome
733   research, 12, 2128–2137, 2013.

734   Pietilä, S., Suomi, T., and Elo, L. L.: Introducing untargeted data-independent acquisition for
735   metaproteomics of complex microbial samples, ISME COMMUN., 2, 51,
736   https://doi.org/10.1038/s43705-022-00137-0, 2022.

737   Ram, R. J., VerBerkmoes, N. C., Thelen, M. P., Tyson, G. W., Baker, B. J., Blake, R. C., Shah,
738   M., Hettich, R. L., and Banfield, J. F.: Community proteomics of a natural microbial biofilm,
739   Science, 308, 1915–1920, 2005.

740   Saito, M. A., McIlvin, M. R., Moran, D. M., Goepfert, T. J., DiTullio, G. R., Post, A. F., and
741   Lamborg, C. H.: Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by
742   protein biomarkers, Science, 345, 1173–1177, 2014.

743   Saito, M. A., Dorsk, A., Post, A. F., McIlvin, M. R., Rappé, M. S., DiTullio, G. R., and Moran, D.
744   M.: Needles in the blue sea: Sub-species specificity in targeted protein biomarker analyses
745   within the vast oceanic microbial metaproteome, Proteomics, 15, 3521–3531, 2015.

746   Saito, M. A., Bertrand, E. M., Duffy, M. E., Gaylord, D. A., Held, N. A., Hervey IV, W. J., Hettich,
747   R. L., Jagtap, P. D., Janech, M. G., and Kinkade, D. B.: Progress and challenges in ocean
748   metaproteomics and proposed best practices for data sharing, Journal of proteome research,
749   18, 1461–1476, 2019.

750 Saito, M. A., McIlvin, M. R., Moran, D. M., Santoro, A. E., Dupont, C. L., Rafter, P. A., Saunders,
751 J. K., Kaul, D., Lamborg, C. H., and Westley, M.: Abundant nitrite-oxidizing metalloenzymes in
752 the mesopelagic zone of the tropical Pacific Ocean, Nature Geoscience, 13, 355–362, 2020.

753 Saunders, J. K., Gaylord, D. A., Held, N. A., Symmonds, N., Dupont, C. L., Shepherd, A.,
754 Kinkade, D. B., and Saito, M. A.: METATRYP v 2.0: Metaproteomic least common ancestor
755 analysis for taxonomic inference using specialized sequence assemblies—standalone software
756 and web servers for marine microorganisms and coronaviruses, Journal of proteome research,
757 19, 4718–4729, 2020.

758 Scanlan, D. J., Silman, N. J., Donald, K. M., Wilson, W. H., Carr, N. G., Joint, I., and Mann, N.
759 H.: An immunological approach to detect phosphate stress in populations and single cells of
760 photosynthetic picoplankton, Applied and environmental microbiology, 63, 2411–2420, 1997.

761 Schiebenhoefer, H., Van Den Bossche, T., Fuchs, S., Renard, B. Y., Muth, T., and Martens, L.:
762 Challenges and promise at the interface of metaproteomics and genomics: an overview of
763 recent progress in metaproteogenomic data analysis, Expert Review of Proteomics, 16, 375–
764 390, 2019.

765 Sørensen, T.: A method of establishing groups of equal amplitude in plant sociology based on
766 similarity of species and its application to analyses of the vegetation on Danish common.,
767 Kongelige Danske Videnskabernes Selskab, 5, 1–34, 1948.

768 Sowell, S. M., Wilhelm, L. J., Norbeck, A. D., Lipton, M. S., Nicora, C. D., Barofsky, D. F.,
769 Carlson, C. A., Smith, R. D., and Giovanonni, S. J.: Transport functions dominate the SAR11
770 metaproteome at low-nutrient extremes in the Sargasso Sea, The ISME journal, 3, 93–105,
771 2009.

772 Stewart, H. I., Grinfeld, D., Giannakopulos, A., Petzoldt, J., Shanley, T., Garland, M., Denisov,
773 E., Peterson, A. C., Damoc, E., Zeller, M., Arrey, T. N., Pashkova, A., Renuse, S., Hakimi, A.,
774 Kühn, A., Biel, M., Kreutzmann, A., Hagedorn, B., Colonius, I., Schütz, A., Stefes, A., Dwivedi,
775 A., Mourad, D., Hoek, M., Reitemeier, B., Cochems, P., Kholomeev, A., Ostermann, R., Quiring,
776 G., Ochmann, M., Möhring, S., Wagner, A., Petker, A., Kanngiesser, S., Wiedemeyer, M.,
777 Balschun, W., Hermanson, D., Zabrouskov, V., Makarov, A. A., and Hock, C.: Parallelized
778 Acquisition of Orbitrap and Astral Analyzers Enables High-Throughput Quantitative Analysis,
779 Anal. Chem., 95, 15656–15664, https://doi.org/10.1021/acs.analchem.3c02856, 2023.

780 Tagliabue, A.: 'Oceans are hugely complex': modelling marine microbes is key to climate
781 forecasts, Nature, 623, 250–252, https://doi.org/10.1038/d41586-023-03425-4, 2023.

782 Timmins-Schiffman, E., May, D. H., Mikan, M., Riffle, M., Frazar, C., Harvey, H. R., Noble, W.
783 S., and Nunn, B. L.: Critical decisions in metaproteomics: achieving high confidence protein
784 annotations in a sea of unknowns, The ISME journal, 11, 309–314, 2017.

785 Ustick, L. J., Larkin, A. A., Garcia, C. A., Garcia, N. S., Brock, M. L., Lee, J. A., Wiseman, N. A.,
786 Moore, J. K., and Martiny, A. C.: Metagenomic analysis reveals global-scale patterns of ocean
787 nutrient limitation, Science, 372, 287–291, 2021.

788 Van Den Bossche, T., Kunath, B. J., Schallert, K., Schäpe, S. S., Abraham, P. E., Armengaud,
789 J., Arntzen, M. Ø., Bassignani, A., Benndorf, D., and Fuchs, S.: Critical Assessment of

790   MetaProteome Investigation (CAMPI): a multi-laboratory comparison of established workflows,
791   Nature communications, 12, 1–15, 2021.

792   Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D.,
793   Burovski, E., Peterson, P., Weckesser, W., and Bright, J.: SciPy 1.0: fundamental algorithms for
794   scientific computing in Python, Nature methods, 17, 261–272, 2020.

795   Waskom, M. L.: Seaborn: statistical data visualization, Journal of Open Source Software, 6,
796   3021, 2021.

797   Williams, T. J., Long, E., Evans, F., DeMaere, M. Z., Lauro, F. M., Raftery, M. J., Ducklow, H.,
798   Grzymski, J. J., Murray, A. E., and Cavicchioli, R.: A metaproteomic assessment of winter and
799   summer bacterioplankton from Antarctic Peninsula coastal surface waters, The ISME journal, 6,
800   1883–1900, 2012.

801   Wilmes, P. and Bond, P. L.: Metaproteomics: studying functional gene expression in microbial
802   ecosystems, Trends in microbiology, 14, 92–97, 2006.

803   Wilmes, P., Andersson, A. F., Lefsrud, M. G., Wexler, M., Shah, M., Zhang, B., Hettich, R. L.,
804   Bond, P. L., VerBerkmoes, N. C., and Banfield, J. F.: Community proteogenomics highlights
805   microbial strain-variant protein expression within activated sludge performing enhanced
806   biological phosphorus removal, The ISME journal, 2, 853–864, 2008.

807   Wu, M., McCain, J. S. P., Rowland, E., Middag, R., Sandgren, M., Allen, A. E., and Bertrand, E.
808   M.: Manganese and iron deficiency in Southern Ocean Phaeocystis antarctica populations
809   revealed through taxon-specific protein indicators, Nature communications, 10, 1–10, 2019.

810
811

812 **Figure Captions**

813 **Figure 1.** Ocean metaproteomics intercomparison experimental design and sample collection.

814 a) The laboratory component (left) consisted of collection of field samples, 1-dimensional (1D)

815 chromatographic separation followed by data dependent analysis (DDA) uniformly employing

816 orbitrap mass spectrometers analyses by participating laboratories and submission of raw and

817 processed data. The informatic (right) component consisted of distribution of two 1D-DDA files,

818 peptide-to-spectrum matching (PSMs), and submission and compilation of results. b) Size-

819 fractionated sample collection on 3.0 μm pore-size filter followed by a 0.2 μm pore-size Supor

820 filter, and the 0.2–3.0 μm size fraction was used for the intercomparison study. c) Two horizontal

821 *in-situ* McLane pumps were bracketed together with two Mini-MULVS filter head units each and

822 deployment on synthetic line. d) The four 142 mm filters were sliced into eighths (inset) and two

823 slices were distributed to each participating laboratory.

824

825 **Figure 2.** Shared peptides and proteins between laboratory groups using laboratory

826 submissions processed through a single informatics re-analysis pipeline. a) Total number of

827 discovered unique peptides varied by more than three-fold among seven laboratory groups

828 (horizontal bars) due to varying extraction and analytical schemes (FDR 0.1%). The number of

829 intersections between datasets across all seven datasets was 1395 (fourth blue bar from left),

830 and various sets of intersections of peptides were observed amongst the data. b) Total number

831 of discovered proteins (FDR < 1%) varied more than four-fold from 1586 to 6221 among labs

832 (horizontal bars). Intersections between datasets across all seven laboratories was 1056, with

833 various sets of intersections of proteins observed, similar to the peptides. c) 7-way Venn

834 diagrams of shared unique peptides between laboratories showed 1056 shared peptides

835 between the 7 laboratories. d) 3-way Venn diagrams showed 2398, 2304, and 3016 shared

836 unique peptides between laboratories.

837

Figure 3. Comparison of unique peptides and discovered proteins. Comparison as total protein

identifications and protein groups from the single pipeline re-analysis based on submissions

from 9 laboratories. Increasing sample depth is linear with mapping to proteins, (R^2 of 0.97 and

0.98 for total protein IDs and protein groups, respectively, with slopes of 0.37 and 33) implying

that additional peptide discovery leads to proportionally more protein discovery, and that protein

discovery has not yet begun to saturate with more peptides mapping to each protein. Because

simple 1D analyses were stipulated in the intercomparison experimental design, peptide and

protein discovery was correspondingly limited in depth.

846

Figure 4. Quantitative comparison of intercomparison results. a) Pairwise comparisons of

uantitative abundance across six laboratories in units of spectral counts (comparisons with itself

show unison diagonals). b) R^2 values from pairwise linear regressions. d) Total proteins

identified in each laboratory. d) Average of each laboratory's R^2 values from pairwise regression

with the other six laboratories (error bars are standard deviation). In all cases average R^2 values

are higher than 0.5.

853

Figure 5. Taxonomic and functional analysis of metaproteomic intercomparison. a) Percent

spectral counts by taxonomy was similar across laboratories and technical replicates within

laboratories. The sample was dominated by cyanobacteria and alphaproteobacteria,

corresponding primarily to *Prochlorococcus* and *Pelagibacter*, respectively. b) Percent spectral

counts per Kegg Ontology group showed the functional diversity of the sample.

859

Figure 6. Quantitative Sørensen similarity analysis. Analysis of top 1000 proteins (~75% of all

proteins) showed 70–80% similarity between most laboratory groups. Technical triplicates for

each laboratory group are shown.

863

864   **Figure 7.** Intercomparison of informatic pipelines among laboratories. Unique peptide

865   identifications for sample Ocean 8 from 120m depth (a) and Ocean 11 from 20m depth (b), both

866   from the North Atlantic Ocean (Table S3), using a variety of pipelines and PSM algorithms.

867

868   **Figure 8.** Variability in discovered proteins between laboratories occurs in lower abundance

869   proteins. Top 7 panels: Abundance of proteins as percentage of total protein spectral counts

870   within each laboratory (y-axis is percentage), with proteins on the x-axis shown by ranked

871   abundance as the sum of spectral counts across all laboratories. Almost all proteins fall below

872   1% of spectral counts within the sample, and deeper proteomes have lower percentages due to

873   sharing of percent spectral counts across more discovered proteins. Bottom panel: Shared

874   proteins were found early within the long-tail of discovered proteins: the 1056 proteins shared

875   between all laboratory groups are almost all found to the left side indicating their higher

876   abundance in all seven datasets. Scale is binary in the seventh panel indicating presence in 7

877   labs or not.

878

879

880   Figure 1.

881

882

883

34

884   Figure 2.

885

886

a

c

887

888

889

d

890

895

896

897

898   Figure 3

899

900

901

902

903

904

905

906

907   Figure 4.

908

909

910

911

912

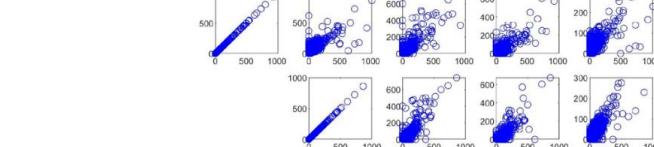913

914   Figure 5.

915

916

917

918

919

920

921
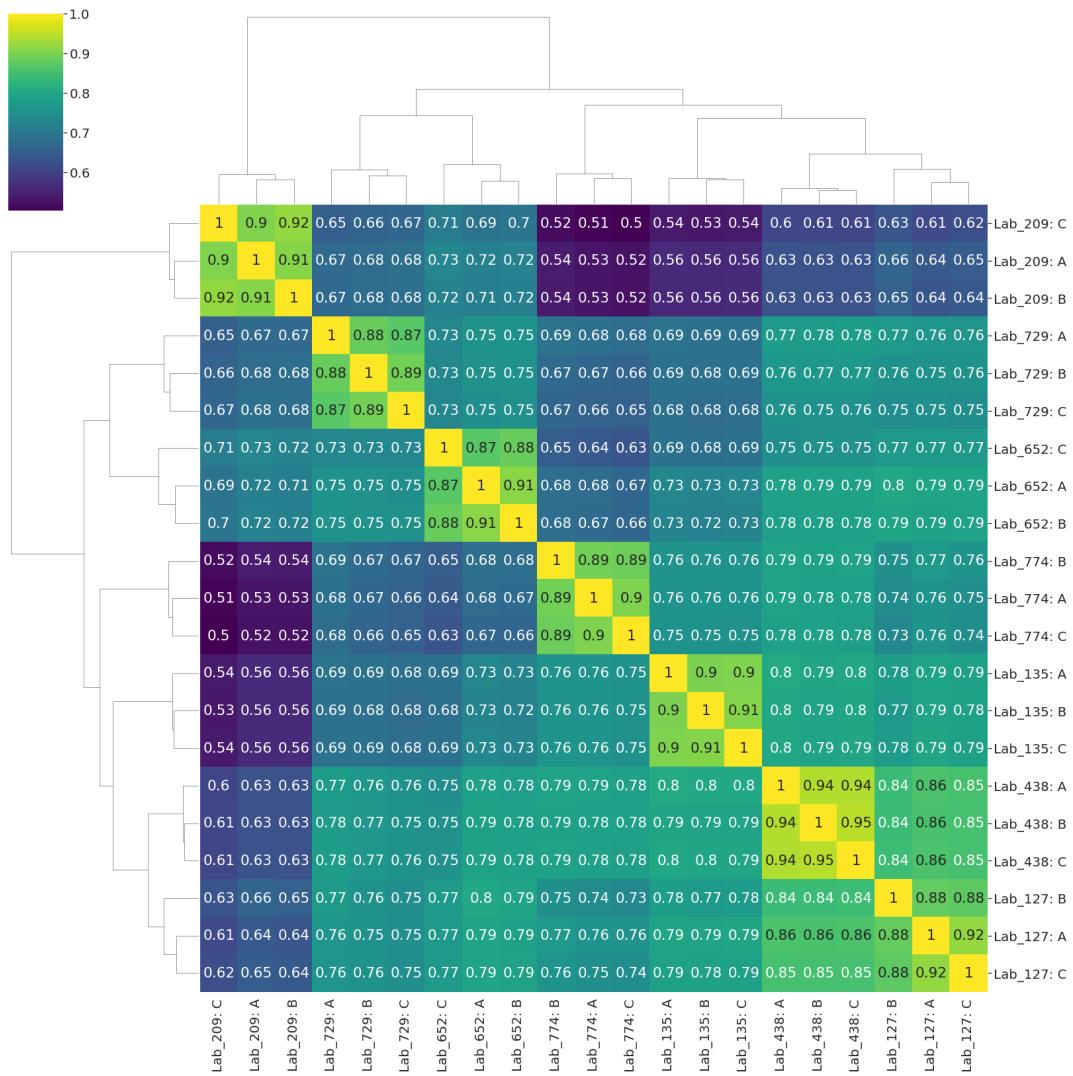
922

923

924

925

926

927   Figure 6.

928

929

930

931

932

933

934

935   Figure 7.

936

937

938

a

b

939

940

941   Figure 8.

942