1 **Comparison of Niskin vs. *in situ* approaches for analysis of gene expression in deep**

2 **Mediterranean Sea water samples**

3

4 V. P. Edgcomb[1*], C. Taylor[2*] M. G. Pachiadaki[1], S. Honjo[3], I. Engstrom[3], M. Yakimov[4],

5

6 [1]Woods Hole Oceanographic Institution, Department of Geology and Geophysics, Woods Hole,

7 MA, USA

8 [2]Woods Hole Oceanographic Institution, Department of Biology, Woods Hole, MA, USA

9 [3]McLane Research Laboratories, Inc., East Falmouth, MA 02536, USA

10 [4]CNR – Institute for Coastal Marine Environment, Messina, Italy

11

12

13 Corresponding and Co-First Authors:

14 *V.P. Edgcomb, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, Tel: +01 508-

15 289-3734, Fax: +01 508-457-2192, Email: vedgcomb@whoi.edu and C. Taylor, Woods Hole

16 Oceanographic Institution, Woods Hole, MA 02543, Tel: +01 508-289-2354, Email:

17 ctaylor@whoi.edu

18

ABSTRACT

Obtaining an accurate picture of microbial processes occurring *in situ* is essential for our understanding of marine biogeochemical cycles of global importance. Water samples are typically collected at depth and returned to the sea surface for processing and downstream experiments. Metatranscriptome analysis is one powerful approach for investigating metabolic activities of microorganisms in their habitat and which can be informative for determining responses of microbiota to disturbances such as the Deepwater Horizon oil spill. For studies of microbial processes occurring in the deep sea, however, sample handling, pressure, and other changes during sample recovery can subject microorganisms to physiological changes that alter the expression profile of labile messenger RNA. Here we report a comparison of gene expression profiles for whole microbial communities in a bathypelagic water column sample collected in the Eastern Mediterranean Sea using Niskin bottle sample collection and a new water column sampler for studies of marine microbial ecology, the Microbial Sampler – *In Situ* Incubation Device (MS-SID). For some taxa, gene expression profiles from samples collected and preserved *in situ* were significantly different from potentially more stressful Niskin sampling and preservation on deck. Some categories of transcribed genes also appear to be affected by sample handling more than others. This suggests that for future studies of marine microbial ecology, particularly targeting deep sea samples, an *in situ* sample collection and preservation approach should be considered.

Keywords:, metatranscriptomics, Microbial Sampler – *In Situ* Incubation Device, pressure effects

42

43 **INTRODUCTION**

44 Microbial metabolic activities are the basis of almost every major biogeochemical cycle 45 in the oceans, and as the research community transitions away from purely descriptive studies of 46 marine microbes to system-level investigations of community activity and responses to changing 47 environmental conditions, it is imperative that we obtain less biased samples for those studies. 48 As a consequence of the fact that the majority of microorganisms are not amenable to existing 49 cultivation approaches, many marine microbiologists and microbial ecologists have embraced 50 culture-independent methods. Metatranscriptomics, or the isolation and sequencing of messenger 51 RNA (mRNA) from an environmental sample, is one powerful method currently used for linking 52 diversity with activity, and for examining microbial activities in response to changing conditions. 53 Metatranscriptomics provides an overview of (at a minimum) the most highly expressed genes in 54 a sample. These transcripts inform about the metabolic pathways that are utilized by microbiota 55 in that sample at the time of sample preservation, and specific proteins that were expressed. 56 Enabled by recent advances in high-throughput sequencing technologies and bioinformatics for 57 processing datasets that can contain tens of millions of reads, metatranscriptomics has become 58 one of the most powerful tools for examining microbial community activities.

59 The metatranscriptomics approach has been used successfully to examine gene 60 expression in varied marine habitats. Examples include deep subsurface sediments (Orsi, *et al.*, 61 2013), the North Pacific Subtropical gyre (Frias-Lopez, *et al.*, 2008, Poretsky, *et al.*, 2009), 62 eastern tropical South Pacific oxygen minimum zone (Ulloa, *et al.*, 2012), coastal waters 63 (Hollibaugh, *et al.*, 2010, Gifford, *et al.*, 2011), hydrothermal vent plumes (Li, *et al.*, 2013), 64 microcosm experiments on mixed water layers from the NE Pacific Ocean (Marchetti, *et al.*, 65 2012). Metatranscriptomics was also recently used to study microbial responses to the Deep

66 Water Horizon oil spill. Mason et al. (2012) found a rapid increased expression of genes
67 associated with motility, chemotaxis, and aliphatic hydrocarbon degradation originating from
68 members of the *Oceanospirallales* in hydrocarbon plume samples (Mason et al. 2012). Rivers et
69 al. (2013) also applied metatranscriptomics to show increased activity (primarily associated with
70 methane- and petroleum-degrading Gammaproteobacteria) within specific metabolic pathways
71 for the degradation of alkanes, aromatic compounds and methane following the Deepwater
72 Horizon spill. Metatranscriptomic studies in the future will be enhanced by sampling
73 technologies that allow us to minimize potential artifacts that can be introduced due to sample
74 handling.

75 For studies of gene expression it is necessary to minimize time between sample collection
76 and chemical preservation. Historically, oceanographers studying marine microbiota have relied
77 on ship-based hydrocasting operations whereby water samples from various depths in the ocean
78 are brought to the surface via Niskin rosette samplers for shipboard water processing. This
79 approach is likely not appropriate for mRNA-based investigations due to the typically large and
80 variable lapses in time and accompanying physicochemical shifts samples are exposed to
81 between collection and preservation. When working in the near-surface (top 1-200m) up to 30
82 minutes may pass between when a water sample is collected and when it is returned to the
83 surface, preserved and processed in the ship's laboratory. In addition, samples are exposed to
84 pressure, and potentially temperature and redox changes before preservation. Pressure changes,
85 potential physicochemical changes, and separation in time between sample collection and
86 processing are exacerbated when working in the deep sea and/or collecting water samples from
87 low-oxygen or anoxic zones. Impacts on transcription by microbes captured in water samples are
88 also likely to vary between taxa depending fragility of cell structures, and on their strategies (or

89 lack thereof) for responding to such changes. Given that average lifetimes of prokaryotic
90 transcripts can be on the order of several minutes (Wang, *et al.*, 2002, Andersson, *et al.*, 2006,
91 Steglich, *et al.*, 2010), the gene expression profiles of microorganisms can potentially be altered
92 significantly. While delays in the preservation of DNA and rRNA (often used as a phylogenetic
93 identifier of viable organisms) is less susceptible to such biases because of their significantly
94 longer half-lives, delayed preservation of Niskin samples can still be an issue if cell integrity is
95 lost due to unnatural and changing conditions during transport from the ocean to the ship's
96 laboratory. Variable and unknown fractions of genetic material from lysed cells can be lost
97 during filtration. This problem is particularly severe for microbial eukaryotes and potentially
98 compounded when sampling greater depths (Edgcomb, *et al.*, 2011).

99 Numerous technical approaches have been undertaken for microbial sampling of the sea
100 and include hydrowire deployed devices (Zobell, 1941, Nikin, 1962, Lewis, *et al.*, 1963), devices
101 that minimize exogenous contamination (e.g. Jannasch & Maddux, 1967, Taylor, *et al.*, 2006),
102 samplers that preserve the conditions of the deep sea (e.g. Jannasch, *et al.*, 1973, Jannasch &
103 Wirsen, 1977, Tabor, *et al.*, 1981, Bianchi, *et al.*, 1999), a sampler that can preserve a whole
104 water sample *in situ* from 120m depth (Feike, *et al.*, 2012), samplers that collect hydrothermal
105 vent fluids (e.g., Malahoff, *et al.*, 2002, Phillips, *et al.*, 2003, Taylor, *et al.*, 2006), AUV
106 (autonomous underwater vehicle) based water samplers (Bird, *et al.*, 2007, Ryan, *et al.*, 2010)
107 those that conduct *in situ* molecular analyses, such as the Environmental Sample Processor (ESP)
108 (Scholin, *et al.*, 2006, Roman, *et al.*, 2007, Scholin, 2010), assess phytoplankton assemblages via
109 imaging flow cytometry (Olson & Sosik, 2007, Sosik & Olson, 2007) and that sample remote
110 biospheres (Carsey, *et al.*, 2000, French, *et al.*, 2001, Blake & Price, 2002, Siegert, *et al.*, 2003,
111 Cardell, *et al.*, 2004). A common limitation of available instrumentation for *in situ* preservation

112 of deep-sea samples, or that might be adapted for this purpose, is the restriction to one or a few
113 samples of limited volume. Sample replication is desirable when examining microbial diversity
114 and activities, and when working in the deep sea, greater total volumes are often required due to
115 low cell densities. Additionally, when working in the mesopelagic and bathypelagic zones, wire
116 time is often precious, and the ability to sample multiple depths is an advantage.

117 Differences in microbial gene expression have been observed between samples preserved
118 *in situ* vs. those recovered to the deck prior to preservation from suboxic samples collected from
119 70-120m depth in the Baltic Sea (Feike*, et al.*, 2012). Additionally, we have documented that
120 changing physicochemical conditions during Niskin sampling can cause lysis of some microbial
121 eukaryotes (Edgcomb*, et al.*, 2011). Obtaining accurate information on *in situ* microbial
122 activities is of fundamental importance to understanding microbially-driven ocean processes and
123 responses of microbiota (and the major biogeochemical cycles that they mediate) to global
124 climate change. To date, no data exist comparing the profile of microbial community gene
125 expression in the deep sea using *in situ* vs. conventional Niskin-based approaches. Here we
126 compare the profile of community gene expression by microbiota in waters from 2222m depth in
127 the Eastern Mediterranean Sea using conventional Niskin rosette collection vs. *in situ* filtering
128 and sample preservation performed using a newly-developed oceanographic instrument for
129 marine microbiological studies, the Microbial Sampler – Submersible Incubation Device (MS-
130 SID) that allows for collection and *in situ* preservation of up to 48 filtered or whole water
131 samples during a single hydrocasting operation. While not an ecological study of microbial
132 activities at this location, the aim of this work was to conduct a general comparison of
133 transcriptome results obtained using both methods, and to analyze the reproducibility of
134 biological replicates collected sequentially using the MS-SID.

135

136 **MATERIALS AND METHODS**

137 *Study site:* The Ionian Sea extends from the Sicily Strait to the Cretan passage, in the Eastern
138 Mediterranean Sea, which is characterized by an eastward progression of increasingly
139 oligotrophic conditions (Sarmiento*, et al.*, 1988, Danovaro*, et al.*, 1999, Thingstad*, et al.*, 2005).
140 The study was conducted using samples collected at a site named KM3 (36° 29'98''N, 15°
141 39'97''E) from 2222m water depth in September, 2012 using the R/V *Urania* of the Italian
142 National Research Council (CNR).

143

144 *Niskin bottle water collection*: Water was collected using 12L Niskin bottles mounted on a
145 General Oceanics rosette sampler equipped with conductivity-temperature and depth (CTD) and
146 pressure sensors. Dissolved oxygen was measured with a SBE oxygen sensor mounted on the
147 CTD, and nutrient concentrations were determined previously at this site using a nutrient auto-
148 analyzer (La Cono*, et al.*, 2010). After transferring water from Niskin bottles to a large sterile
149 carboy, 30 liters of water were pumped through a 0.22µm Sterivex filter cartridge using a
150 peristaltic pump operating around 125 ml/min containing a Durapore filter (Millipore, Millford,
151 MA, USA), which was immediately filled with RNAlater (Life Technologies Inc., Grand Island,
152 NY, USA) and frozen at -80°C until extraction.

153

154 *Use of the MS-SID:* Water samples from the same depth and on the same day were also collected
155 and preserved *in situ* using the MS-SID equipped with a CTD, two turbidity sensors, and an
156 oxygen optode (Figure 1). C. Taylor and McLane Research Laboratories developed automated
157 micro-laboratories for conducting multiple tracer incubation studies during cabled or free-

158 drifting deployments (Taylor & Doherty, 1990, Taylor, *et al.*, 1993, Taylor & Howes, 1994).
159 This technology was recently modified by C. Taylor, V. Edgcomb, and McLane Research
160 Laboratories into an instrument (Figure 1) that conducts *in situ* tracer incubations in combination
161 with *in situ* microbial sampling and preservation. The modular MS-SID (Figure 1) consists of a
162 2L syringe-like incubation chamber, a 50-port Fluidic Distribution Valve, a micro- gear pump
163 for microbial sampling, a tracer injector, 48 Fixation Filter Units (Figure 1 inset) for collection &
164 *in situ* preservation, a High Range CTD (Neil Brown Ocean Sensors, Inc., Falmouth MA, USA),
165 Aanderaa oxygen optode (Aanderaa Data Instruments, Inc., Attleboro, MA, USA), 2 NTURTD
166 0-124 turbidity sensors (WET Labs, Inc., Philomath, OR, USA), and real-time bi-directional
167 communication, and electronic control. The MS-SID possesses a Synchronous Digital Subscriber
168 Line (SDSL) data link (Swartz, *et al.*, 2012) that multiplexes digital signals from up to 5 sensors
169 at once, and bi-directionally transmits signals to and from the instrument for triggering "adaptive
170 sampling" operations via ordinary conducting hydro-wire. This allowed us to precisely position
171 the instrument at the target depth in the water column and then to trigger sampling for our
172 transcriptome studies.

173 The MS-SID collected three consecutive and separate samples for this analysis from the
174 exact same depth as the Niskin rosette based on pressure readings from both instruments, and
175 were collected within 3 hours of the Niskin collections. All samples were filtered *in situ* at 125
176 ml/min through a 47mm 0.2 µm Durapore (Millipore, USA) filter that, upon cessation of
177 filtration was within 10-20 seconds flooded with the preservative RNAlater following filtration.
178 The three filters collected 4L, 3L and 3.4L of water, respectively. Upon retrieval of the
179 instrument to the ship's deck, the Fixation Filter Units were disassembled, and the filters and
180 associated RNAlater solutions were transferred aseptically to three separate sterile cryovials, and

181 frozen at -80°C until extraction. The Sterivex capsule (from Niskin bottle collection) was also
182 stored at -80°C until extraction.
183
184 *Extraction of RNA:* For each of the 3 MS-SID filters total RNA was extracted separately from
185 the filter and RNAlater in order to capture the mRNA from any cells that may have lysed during
186 sample storage. The filters were transferred to 2ml tubes and homogenized without beads for 20
187 seconds at speed 4.0m s^{-1} on a FastPrep®-24 homogenizer (MP Biomedicals, Solon, OH, USA)
188 in 600 µl of RLT Plus buffer (All Prep DNA/RNA Mini Kit, Qiagen). An equal volume of 70%
189 ethanol was added and the manufacturer's instructions were followed for remaining steps. The
190 final volume of extracted RNA was 50 µl in distilled deionized water. The RNAlater from each
191 filter's original tube was transferred to a 50 ml centrifuge tube. Four ml of RLT buffer (RNeasy
192 Midi Kit, Qiagen, USA) were added and gently mixed. An equal volume of 70% ethanol was
193 added and mixed. Total RNA was then extracted using the RNeasy Midi Kit protocol beginning
194 at step 3 and contained in 500 µl of distilled deionized water. Since some cell lysis may have
195 occurred between sample preservation *in situ* and recovery of the filter from each Fixation Filter
196 Unit, we also retained and extracted RNA from the RNAlater contained within the Fixation Filter
197 Unit. For each filter separately, this RNAlater was transferred to a 50 ml centrifuge tube
198 (~15ml). Approximately 7.5 ml of Buffer RLT were added and the samples were run through the
199 All Prep DNA spin column (Qiagen, USA). An equal volume of 70% ethanol was mixed with
200 the sample and total RNA was then extracted using the RNeasy Midi Kit beginning at step 3 and
201 contained in 500µl of distilled, deionized water. The combined extractions for each sample were
202 cleaned and precipitated following the instructions in the MEGAclear kit protocol (Life
203 Technologies, Grand Island NY, USA). The RNA pellet was suspended in 100 µl of distilled,

204 deionized water, and DNAsed for 1 hour at 37°C (Turbo DNAse, Life Techologies, Grand Island
205 NY, USA). Finally all samples were purified using the MegaClear kit one final time.

206 RNA from the Sterivex capsule was extracted in an identical fashion, with the exception
207 that RLT buffer (Qiagen, USA) was initially added directly into the cartridge with the RNAlater
208 already inside the cartridge. The buffer and RNAlater were then collected into a single 50 ml
209 centrifuge tube. An additional 2 ml of RLT buffer (final volume same as for the MS-SID
210 samples) were added to the filter cartridge, and the collection steps were repeated. All extraction
211 protocols were performed as described above for the MS-SID samples. The final suspension
212 volume was 10 µl for all samples. To verify the absence of genomic DNA in the RNA extracts,
213 one µl was used as a template for PCR using 40 cycles and the bacterial primers 8F and 1492R
214 (Edwards*, et al.*, 1989, Stackebrandt & Liesack, 1993) . RNA was quantified using a Qubit®2.0
215 fluorometer (Life Technologies, USA). cDNA was synthesized using the Ovation RNA-Seq
216 system V2 (NuGEN, San Carlos, CA, USA) following the manufacturer's protocol, and sent for
217 transcriptome library sequencing. One lane of Illumina HiSeq 2x100bp was requested for each of
218 the four samples.

219

220 *Analysis of transcriptome data:* Quality trimming of the reads (minimum quality score 28,
221 minimum read length 94 bp and no ambiguous nucleotides) as well as read assembly into contigs
222 and mapping of reads to contigs were performed using CLC Genomics Workbench 6.0, CLCBio,
223 Cambridge, MA, USA). The Rapid Analysis of Multiple Metagenomes with a Clustering and
224 Annotation Pipeline (RAMMCAP) (Weizhong, 2009) was used to assign contigs to clusters of
225 orthologous gene (COG) families, gene ontologies (GO), and protein families (Pfam).
226 Taxonomic assignments of contigs were made using PhymmBL (Brady & Salzberg, 2009),

227 incorporating all available fungal and protist genomes in public databases. The total number of
228 annotated reads assigned to different COG families for each dataset was expressed as a
229 percentage of the total annotated reads for each dataset.

230 The degree of variation in gene expression profiles observed between each of the three
231 replicates collected and preserved *in situ* using the MS-SID was assessed by comparing each of
232 the three replicate expression profiles to each of the other two replicate profiles using R version
233 3.0.2 (http://www.r-project.org), the *DEGseq* package
234 (http://bioconductor.org/packages/release/bioc/manuals/DEGseq/man/DEGseq.pdf) and a MA-
235 plot-based method with Random Sampling (Wang*, et al.*, 2010), as implemented within R. This
236 approach was also used to assess the variation in expression profile obtained with the single
237 Niskin sample relative to the three replicated samples that were preserved *in situ* with the MS-
238 SID. Data for each sample were first normalized to the total number of reads within each
239 taxonomic group being analyzed. Within *DEGseq* we calculated a MA-plot with random
240 sampling to compare our three replicate MS-SID samples to our Niskin sample. This method
241 assigns a score to each gene transcript on the basis of its differential expression relative to the
242 standard deviation of repeated measurements and permutations of these repeated measurements
243 are used to estimate the false discovery rate (Wang*, et al.*, 2009).

244

245 **RESULTS**

246 *Study site:* At 2222m depth, the water temperature was 13.81°C, and salinity was 38.723 PSU.
247 Oxygen concentration was 204µM, representative of the oxygen profiles between the near-
248 surface and 2222m at this site. Nitrate and phosphate concentrations were measured previously at
249 this site (La Cono et al. 2010), showing an increasing trend with depth, reaching 3.75 +/- 0.54

250 and 0.13 +/- 0.02 µmol L^{-1}, respectively at 3010m depth. Nitrite was constant at 0.04 +/- 0.01
251 µmol $L^{-1}$from the sea surface to 3010 m.
252
253 *Transcriptome library sequencing*: For each of the 4 samples (one Niskin 30L water sample and
254 3 MS-SID samples) the number of recovered reads, reads remaining after trimming, number of
255 contigs formed, annotated reads, and annotated contigs can be found in Table 1. 78.3 million
256 sequence reads were obtained from the Niskin sample. From the 4L, 3L, and 3.4L MS-SID
257 samples, 122.4, 96.1, and 118.9 million sequence reads were obtained, respectively. After
258 trimming sequences for low-quality and short reads, we obtained 37.9, 66.3, 51.5, and 63.9
259 million reads from the Niskin, MS-SID 4L, MS-SID 3L, and MS-SID 3.4L samples,
260 respectively. Of the total reads that passed quality control, 8.9 million were possible to annotate
261 from the Niskin sample, and these formed 11.8 thousand annotated contigs. From the MS-SID
262 replicate samples, 26.6, 25.0, and 27.0 million reads were annotated, and these formed 9.7, 8.5,
263 and 6.1 thousand annotated contigs. Data are deposited in CAMERA
264 (https://portal.camera.calit2.net) and the Short Read Archive (SRP042349).
265
266 *Transcriptome library analyses*. The majority of annotated reads for both MS-SID and Niskin
267 samples affiliated with the Gammaproteobacteria (47% and 38% for MS-SID and Niskin
268 samples, respectively, data not shown). Percent representation of COG categories (out of total
269 reads recovered for each library) for gammaproteobacterial transcripts recovered from the Niskin
270 sample and from each of the three MS-SID replicate samples is shown in Figure 2 for all COG
271 categories representing >0.001% of total reads in either library. For COG categories T (signal
272 transduction mechanisms), M (cell wall, membrane, and envelope biogenesis), H (coenzyme

273 transport and metabolism), F (nucleotide transport and metabolism), C (energy production and
274 conversion), L (replication, recombination and repair), P (inorganic ion transport and
275 metabolism), and E (amino acid transport and metabolism), percent representation out of total
276 reads for Gammaproteobacteria was higher by ~1-2% in the MS-SID samples. In the case of
277 COGs V (defense mechanisms), U (intracellular trafficking, secretion, and vesicular transport), Z
278 (cytoskeleton), Q (secondary metabolites biosynthesis, transport and catabolism), N (cell
279 motility), K (transcription), J (translation, ribosomal structure and biogenesis), I (lipid transport
280 and metabolism), G (carbohydrate transport and metabolism), D (cell cycle control, cell division,
281 chromosome partitioning), expression levels as a percentage of total annotated reads for each
282 library were approximately the same (within approximately half a percent) in the Niskin and MS-
283 SID samples. For COG categories O (posttranslational modification, protein turnover,
284 chaperones), and A (RNA processing and modification), expression levels were slightly higher in
285 the Niskin sample than in each of the MS-SID samples. The COG categories with the most
286 abundant annotated transcripts for Gammaproteobacteria included amino acid transport and
287 metabolism (4.8-7.3% of total reads in each of the 4 libraries), inorganic ion transport and
288 metabolism (3.5-6.1%), replication, recombination and repair (3.5-6.4%), and energy production
289 and conversion (4.9-6.4%) (Figure 2).

290 The degree of variation in gene expression profiles, defined here as the relative
291 abundance of annotated transcripts, between the three MS-SID replicates was assessed using the
292 *DEGseq* package, which revealed 9.2-9.8% variation in pairwise expression profiles (relative
293 transcript abundance) between replicates. Differentially expressed transcripts included
294 peptidases, three RNA methylases, ATPases, and a NCAIR synthetase. MA plots for pairwise

295 comparisons of the three replicates depicting the $\log_2$ fold change in gene expression vs. the
296 mean of normalized counts showed a balanced distribution of the contigs (data not shown).

297 For comparisons of Niskin to MS-SID profiles for protists and Fungi, the data for each of
298 the MS-SID libraries were combined, representing a total of 10.4L filtered seawater. All
299 together, annotated reads assigned to eukaryotes represented a small fraction of the
300 metatranscriptome recovered in the MS-SID and Niskin datasets (1.3 and 1.0%, respectively).
301 The percent representation of GOC categories for transcripts assigned to protists (Figure 3)
302 reveal some differences from the results for Gammaproteobacteria. In most cases, the relative
303 abundance of transcripts (within each of the categories) was higher in the MS-SID dataset than in
304 the Niskin dataset. These categories include defense mechanisms (0.12 vs. 0.02%), signal
305 transduction mechanisms (0.04 vs. 0.01%), secondary metabolites biosynthesis, transport and
306 catabolism (0.06 vs. 0.01%), carbohydrate transport and metabolism (0.09 vs. 0.01%), nucleotide
307 transport and metabolism (0.03 vs. ~0%), energy production and conversion (0.08 vs. 0.03%),
308 replication, recombination and repair (0.29 vs. 0.08%), inorganic ion transport and metabolism
309 (0.22 vs. 0.06%), and amino acid transport and metabolism (0.07 vs. 0.03%). There were some
310 exceptions, however. Protists exhibited higher expression of genes in the Niskin sample
311 associated with posttranslational modification, protein turnover, and chaperones (higher by
312 0.12%), cell wall, membrane and envelope biosynthesis (by 0.1%), transcription (by 0.18%),
313 cytoskeleton (by 0.02%), cellular trafficking, secretion, and vesicular transport (by 0.03%),
314 translation, ribosomal structure and biogenesis (by 0.35%), and coenzyme transport and
315 metabolism (by 0.03%).

316 Fungal transcripts represented 1.8 and 0.3% of the MS-SID and Niskin datasets,
317 respectively. For Fungi, transcript relative abundances for almost every COG category were

consistently higher by 0.02 to 0.3 in the combined MS-SID vs. the Niskin library (Supplementary Figure 1). Only for the two COG categories, transcription and cytoskeleton, were expression levels higher (0.015 and 0.005%, respectively) in the Niskin sample.

Analysis of variation in relative transcript abundance obtained with the Niskin sample vs. the three replicated samples that were preserved *in situ* with the MS-SID was performed for both the dominant bacterial group, the Gammaproteobacteria, and the dominant eukaryotic group, Fungi, using the *DEGseq* package (Wang, *et al.*, 2009). These analyses revealed that 75% of the fungal (n=319) and 99% of gammaproteobacterial transcripts (n=1713) showed differential expression ($p<0.001$). Supplementary Figure 2A and B show the differentially expressed transcripts for Gammaproteobacteria and Fungi on MA-plots for the MS-SID vs. Niskin comparisons, with red dots representing transcripts with significantly different ($p<0.001$) expression between the two types of samples. Supplementary Figure 2C and D show boxplots of the read counts per gene for libraries prepared using the two sampling methods for Gammaproteobacteria and Fungi, respectively. When transcribed genes within COG category O (posttranslational modification, protein turnover, chaperones) were compared for Niskin and the three MS-SID samples, 86% of the reads assigned to the 118 transcripts detected, were differentially expressed ($p<0.001$).

DISCUSSION

Profiles of gene expression obtained for these water column samples from 2222m in the Eastern Mediterranean Sea using *in situ* preservation of water samples vs. Niskin bottle collection were not the same. While it can not be entirely ruled out that some differences were due to the use of membrane filters in the MS-SID sampler vs. Sterivex filters aboard the ship where the Niskin

341 waters were processed, or due to differences in total volumes filtered (10.4L total for MS-SID
342 and 30L for Niskin), it is intuitive to consider that increased handling and physicochemical
343 alterations prior to preservation likely alters the profile of community activities. Sample
344 collection approaches that rely on returning water samples to the surface prior to processing and
345 preservation may be completely appropriate for many types of investigations, however our
346 results and the results of Feike et al. who conducted a similar experiment using water samples
347 from 70-120m depth in the Baltic Sea (Feike*, et al.*, 2012) suggest that for examinations of
348 microbial activities, *in situ* preservation is desirable. This may be increasingly important for
349 deeper water samples.

350 Obtaining an accurate snapshot of *in situ* microbial activities requires collecting and
351 preserving samples in a fashion that minimizes detectable environmental perturbations. All
352 existing water-sampling technologies that involve drawing water through intake plumbing into a
353 chamber or bag or drawing water through a filter have the potential to introduce artifacts
354 resulting from detected perturbations that cause microbes to alter transcription of particular types
355 of genes. Nonetheless, the primary advantage of *in situ* approaches is that they minimize artifacts
356 associated with sample handling, and allow for almost immediate preservation of sensitive
357 molecules. Furthermore, moored *in situ* technologies are useful for studies of temporal variations
358 in microbial communities. For example, Ottesen et al. (2011) used a moored ESP to
359 automatically filter and preserve (in RNAlater) near-surface water samples from Monterey Bay
360 at four time points over the course of a day. Using this technology to capture and preserve RNA
361 allowed these researchers to track changes in community composition and gene expression over
362 this timeframe. This was expanded upon by Ottesen and colleagues (2013) who used the moored
363 ESP to detect coordinated gene expression between microbial groups over a 2 day timeframe at

364 23m depth off the coast of California, possibly in response to specific environmental cues.
365 Preserving samples *in situ* in the deep ocean prior to returning them to the sea surface has the
366 obvious advantage of avoiding potentially significant pressure and other physicochemical
367 changes that can occur during sample retrieval and prior to processing/preservation in the ship
368 laboratory. Preserving a whole water sample immediately upon collection by drawing it into a
369 bag preloaded with preservative (as can be performed by the MS-SID) or into a container that is
370 immediately injected with preservative (Feike*, et al.*, 2012) accomplishes this goal. However,
371 this approach limits the volume that one can collect. In some cases, such as for water samples
372 where cell densities are low, it is desirable to collect information (e.g., RNA) from a much
373 greater volume of water via filtration. During filtration and prior to preservation, cells impinged
374 on filters earlier in the filtration routine undoubtedly experience more significant environmental
375 changes than those that are captured toward the end of filtration. This is a concern not only for
376 Niskin-based sample collection, but also for *in situ* filtration technologies such as the MS-SID. In
377 contrast to waters collected using *in situ* filtration and preservation methods, all cells on filters
378 from Niskin water samples have undergone depressurization and have spent variable lengths of
379 time on deck prior to filtration. This is not the case for *in situ* preserved filters, where until
380 impinged on a filter the cells are in their natural environment. The composite mRNA pools
381 within the filtered organisms reflect an average residence time on the filter surface that is
382 approximately one-half the total filtration time (e.g., ~10 min for our 20 min filtrations), which
383 additionally enhances the fidelity of the samples collected. While the MS-SID will capture *in situ*
384 gene expression patterns with improved fidelity, there may be some genes that reflect the stresses
385 of being impinged on a filter during filtration. In all cases, artifacts introduced into expression
386 profiles by filtration can be minimized by shortening filtration times.

The three MS-SID samples for which we present data represent biological replicates, however we note that they were collected sequentially, and not concurrently, from the same water feature. Figure 2 shows that these replicates captured essentially the same profile of relative gene expression between different COG categories for Gammaproteobacteria, and this was true of all other taxonomic groups (data not shown). This is supported by analysis of variation in relative transcript abundance between the three MS-SID replicates for all taxa using the *DEGseq* package, which revealed 9.2-9.8% of genes were differentially expressed between replicates (<p=0.05 levels). *DEGseq* is suspected to overestimate differentially expressed genes between samples (Guo et al. 2013), so the amount of differential expression between replicates might be lower than this.. The 9.2-9.8% variation between replicates is still much lower than the percent of differentially expressed genes detected using the same method comparing the Niskin and the average MS-SID overall gene expression profiles (discussed below). MA plots for pairwise comparisons of the three replicates depicting the $\log_2$ fold change in gene expression vs. the mean of normalized counts showed a balanced distribution of contigs between the replicates (data not shown). Although many metatranscriptome studies of marine samples published to date have relied upon single samples from selected habitats (e.g., Frias-Lopez*, et al.*, 2008, Poretsky, *et al.*, 2009, Feike*, et al.*, 2012, Orsi*, et al.*, 2013), we would have preferred to analyze replicate samples, however replicate Niskin samples from a separate cast were not available for this study. Having biological replicates for the Niskin collection would have helped to confirm that some differences in observed profiles in the MS-SID and Niskin samples were not due to temporal variation. For this deep-sea location, however, we doubt that the 3 hours difference in sample collection between the methods could account for observed differences, although this possibility cannot be entirely ruled out.

410 It is noteworthy that the Niskin sample produced 64-81% of the number of initial reads
411 obtained from the 3 MS-SID samples, and 59-73% after quality control trimming, in spite of
412 identical library preparation protocols, equal sequencing efforts and the fact that the Niskin filter
413 represented 30L of filtered water compared to 3-4L of water for each of the MS-SID samples.
414 This lower number of initial reads however, assembled into roughly twice as many contigs.
415 Approximately 23% of Niskin initial reads were successfully annotated, in comparison to 49, 40,
416 and 42% of the initial reads for the 3 MS-SID replicates, and 29% of Niskin contigs were
417 successfully annotated vs. an average of 44% of MS-SID contigs (Table 1). The N50 lengths for
418 the MS-SID contigs were slightly higher than the Niskin sample (473, 447, and 428 vs. 424 bp),
419 however this does not likely explain differences in contig formation or annotation. We
420 hypothesize that some aspect of Niskin water collection and processing (pressure changes,
421 increased time delays between sample collection and processing, or other unknown
422 physicochemical alterations to the sample) reduced the community complexity in this sample
423 (possible cell lysis, cell death) for some taxa. Lower complexity could lead to greater contig
424 formation, and if remaining complexity favored taxa that are less represented in public sequence
425 databases, then this may explain the lower percentage of annotated contigs for the Niskin sample.
426 Pressure changes and sample handling are likely to have differential impacts on cell
427 integrity and gene transcription of different taxa, and some cell activities are also likely to be
428 more sensitive to perturbations than others. This was observed in our datasets. For example, for
429 Gammaproteobacteria, which appear to be one of the dominant bacterial groups in our samples,
430 certain COG categories appear to show a greater difference than others in gene expression
431 between the Niskin and MS-SID data sets (Figure 2). However, a test for differential expression
432 revealed that almost all unique transcripts annotated to Gammaproteobacteria in Niskin vs. the

433 three MS-SID datasets were expressed differentially ($p<0.001$). This suggests that for
434 investigations of specific metabolic pathways within a particular COG category, that *in situ*
435 fixation may provide a different picture than Niskin sampling.

436 While the percentages of reads associated with certain functional categories, such as
437 genes associated with the 3-hydroxypropionate/4-hydroxybutyrate carbon fixation cycle were
438 detected at approximately equal levels (0.27% of Niskin library and 0.25-0.39% of the MS-SID
439 libraries) for many other transcribed genes, expression in Niskin vs. MS-SID libraries varied.
440 Although it was not the objective of this study proof-of-concept study to closely examine
441 specific microbial activities (ecological function) within different COG categories of expressed
442 genes in these samples, a closer examination of specific transcripts and taxonomic affiliation of
443 those transcripts within COG O (posttranslational modification, protein turnover, and
444 chaperones) provides insight into the types of variation underlying observed differences in
445 overall results. Transcripts annotated to COG category O represented 2.8% (annotated to 29
446 taxa) and 1.1% (annotated to 18 taxa) of total transcripts in each library for Niskin and MS-SID
447 samples, respectively. Transcripts in the Niskin library were annotated to 10 taxonomic groups
448 that were not also detected in the MS-SID samples; Bacillariophycea, Chlamydiae, Chlorobia,
449 Chloroflexi, Choanoflagellata, Crenarchaeota, Tenericutes, Thaumarchaeota, Thermotogae, and
450 Verrucomicrobia (results presented at different taxonomic levels due to variance in annotation
451 detail). This suggests that *in situ* filtration and preservation may have reduced stresses
452 experienced by some taxa. Looking within the COG O transcripts annotated to protozoa and
453 metazoa provides insight into observed variations between Niskin and MS-SID samples, and
454 illustrates that sensitivity to sample handling is variable between taxa.

455 Among transcript types that were unique to the Niskin library, 5478, 4858, 529, and
456 4103 reads were detected for molecular chaperones of choanoflagellates, ciliates,
457 Bacillariophyta, and metazoa, respectively. Also unique to the Niskin library was 62 and 1301
458 reads annotated to molecular chaperone HSP90 of choanoflagellates and various metazoan,
459 respectively. Heat shock proteins such as HSP90 assist the proper folding of proteins under cell
460 stress. Expression of transcripts for the protein ubiquitin was detected in both libraries, but at
461 higher levels (1786 vs. 42 reads) in Niskin vs. MS-SID samples. Ubiquitin is a small protein
462 associated with post-translational modification that is expressed in many eukaryotes and can
463 among other functions, signal for the degradation of a damaged protein (Glickman and
464 Ciechanover 2002). Some types of COG O transcripts were detected only in the MS-SID
465 samples, and these were all affiliated with various metazoan (echinoderms, nematodes). These
466 included metazoan transcripts for thioredoxin reductase (511 reads), FKBP-type peptidyl prolyl
467 cis-trans isomerase (190 reads), transcripts annotated to disulfide bond chaperone activity (345
468 reads), an ATPase with chaperone activity (345 reads), and a predicted metalloendopeptidase
469 associated with peptide breakdown (8729 reads). Transcriptional responses to sample handling
470 also differed between taxa. Illustrating this, in comparison to protists (Figure 3), Fungi
471 (Supplementary Figure 1) exhibit greater differences in percentage representation within many
472 COG categories between MS-SID and Niskin samples.

473 Increase in hydrostatic pressure is known to affect a wide range of cellular processes of
474 microorganisms, including DNA structure and function, membrane synthesis and repair,
475 cytochrome formation, cell division, enzyme function, and to affect certain taxa more than others
476 (see review by Bartlett (2002) and Ishii et al. (2004)). In studies of *Escherichia coli*, cellular
477 processes including motility, substrate transport, cell division, growth, DNA replication,

478 translation, transcription, and viability are repressed by increases in pressure (Bartlett, 2002).
479 Less is known about decompression effects on microorganisms. We observed lower expression
480 levels (as a percentage of total annotated reads) in the Niskin sample for many COG categories
481 for Fungi (Supplementary Figure 1), and decompression effects are one possible explanation for
482 this. In comparison, for protists, some categories of transcripts appear to be suppressed and
483 others, such as COG M (cell wall/membrane/envelope biogenesis), enhanced. In the case of
484 COG M, this may indicate protists increase transcription of genes associated with cell membrane
485 repair in response to decompression. Pressure is the only stressor known to simultaneously
486 induce a wide range of both heat shock and cold shock proteins, likely in response to
487 destabilization of protein quaternary and tertiary structures, and their simultaneous induction in
488 *E. coli* may represent an attempt to repair the damaging effects of elevated pressure on
489 membrane integrity, translation processes, and macromolecule stability (Bartlett, 2002). Heat
490 shock proteins are known to be induced in piezophiles such as the deep-sea piezophilic
491 hyperthermophile *Thermococcus barophilus* upon decompression (Marteinsson, *et al.*, 1999).
492 While an increase in percent transcription within COG category O (posttranslational
493 modification, protein turnover, chaperones) was not observed in the Niskin sample relative to the
494 MS-SID samples for Fungi (Supplementary Figure 1), an increase was observed for
495 Gammaproteobacteria (Figure 2), protozoa (Figure 3), and many other taxa (data not shown).
496 When the number of reads assigned to the 118 unique types of transcripts detected in our
497 samples within this COG category in the Niskin sample were compared to those detected in the
498 three MS-SID samples, 86% of them showed evidence of differential levels of expression
499 ($p<0.001$). Given that high-pressure activation of gene expression is known among piezophilic

500 bacteria (Bartlett, 2002), it stands to reason that transcription of many types of genes may also be
501 affected by decompression for potentially, a wide range of taxa.

502 Lipid membranes are also known to be sensitive to pressure effects since lipids are highly
503 compressible (Weber & Drickamer, 1983), and during decompression, the bacterium *Colwellia*
504 responded by forming intracellular vesicles and releasing membrane fragments into the medium,
505 followed by cell lysis (Chastain & Yayanos, 1991). Our observations of differential transcription
506 (repression or enhancement) between Niskin and MS-SID samples of genes associated with lipid
507 transport and metabolism, intracellular trafficking, secretion, and vesicular transport, and cell
508 wall, membrane, and envelope biogenesis may be indicative of such decompression responses
509 for some taxa.

510 Free-living (dispersed in the water column), and detritus-associated protozoa are now
511 recognized as important components of marine microbial communities (Caron, 1991). While
512 piezotolerant protists (flagellates, ciliates, and amoebae) from the bathypelagic have been
513 observed using microscopic analysis of cultures grown under high pressure or preserved samples
514 (Patterson*, et al.*, 1993, Atkins*, et al.*, 2000, Edgcomb*, et al.*, 2011), less is known about the
515 effects of pressure changes on protozoa than for Bacteria and Archaea. Numbers of species of
516 free-living and particle-associated protists declines from the photic zone into the bathypelagic,
517 and this decline is thought to be attributed to reductions in diversity and quantities of available
518 food, and potentially also due to pressure effects (Patterson*, et al.*, 1993). Flagellates are known
519 to have variable sensitivities to changes in pressure, and even species within genera can vary
520 significantly in their responses (Turley & Carstens, 1991). This implies that significant
521 depressurization may cause some protist taxa to lyse during recovery if cells are not preserved *in*
522 *situ*. Supporting this notion, in studies of anoxic waters from 900m depth in the Cariaco Basin,

523 Venezuela, it was necessary to preserve water samples *in situ* for fluorescence and scanning
524 electron microscopy in order to more accurately assess protist abundances (Edgcomb*, et al.*,
525 2011). Deep ocean microorganisms are known to alter the phospholipid composition of their
526 membranes to maintain adequate membrane fluidity under higher pressures (Yano*, et al.*, 1998).
527 It is conceivable that for an unknown fraction of the microbial community from mesopelagic and
528 bathypelagic realms the higher membrane fluidity upon depressurization may make them more
529 susceptible to lysing or leaking. Depressurization during sample recovery not only has the
530 potential to alter gene expression from what was occurring in the sampled habitat, but to bias
531 against the recovery of an unknown fraction of the community from any sample, due to loss of
532 cell integrity for some taxa.

533

534 **CONCLUSIONS**

535

536 Collectively, the potential effects of depressurization during sample retrieval and on deck sample
537 handling during filtration have the potential to confound our ability to gather an accurate
538 impression of *in situ* microbial metabolic activities. Results of this study suggest that such effects
539 may be non-trivial when sampling deep-sea habitats, for certain taxa, and for certain categories
540 of transcribed genes. Further comparisons of gene expression in a variety of marine water
541 column and sedimentary habitats using different sampling methodologies are warranted.
542 Understanding the mechanisms underpinning our observed pressure and/or sample handing
543 effects on different types of microorganisms will require analyses of the integrity and/or changes
544 in specific enzymes catalyzing apparently pressure-sensitive processes in organisms preserved *in*
545 *situ* vs. exposed to significant pressure changes and increased sample handling.

546

547 AUTHOR CONTRIBUTIONS

548

549 VE, CT, and MP designed experiments together with SH. CT and VE together with SH, IE and
550 McLane Research Laboratories personnel developed the MS-SID. VE, CT, MP, and MY
551 conducted experiments, and VE and MP analyzed results.

552

553 ACKNOWLEDGMENTS

554

555 We would like to thank the captains and crews of the R/V *Urania* for their hard work to assure
556 the success of our sampling objectives, E. Leadbetter and D. Patterson for helpful discussions on
557 the data and their interpretation, R. Schmitt for generous help with repairs to the High Range
558 CTD, G. LaSpada and F. Smedile for assistance with sampling, and L. Wang for helpful
559 discussion on analytical approaches using DEGexp. In addition to technical and engineering
560 contributions, we thank Y. Honjo and McLane Research Laboratories for assistance with funding
561 the transcriptome library analyses. This research was funded by NSF OCE-1061774 to VE and
562 CT, NSF DBI-0424599 to CT and NSF OCE-0849578 to VE and colleague J. Bernhard. Cruise
563 participation was partially supported by Deutsche Forschungsgemeinschaft (DFG) grant
564 STO414/10-1 to T. Stoeck.

565

566 TABLE AND FIGURE LEGENDS

567

568 Table 1. Summary of transcriptome sequencing, quality control, and number of annotated reads
569 (based on the COG database) for the Niskin water sample and the three replicate MS-SID
570 samples.
571
572 Figure 1. Microbial Sampler-Submersible Incubation Device (MS-SID). Panel A, schematic of
573 the instrument illustrating major modular components. Each Fixation Filter Unit (lower left
574 inset) possesses a reservoir for containing an appropriate (denser than water) chemical
575 preservative (in our study RNAlater). During filtration, preservative loss is prevented by a
576 poppet that seals the access hole shown. Upon cessation of filtration, the poppet settles by
577 gravity, opening the reservoir, allowing preservative to flow onto the filter surface by
578 convection. Panel B, deployment of SID-ISMS for collection of microbial samples from
579 Mediterranean Sea for grazing studies.
580
581 Figure 2. Percentage of total annotated reads assigned to Gammaproteobacteria in each COG
582 category for the Niskin sample (black) and the three replicate MS-SID samples (shades of grey).
583
584 Figure 3. Percentage of total annotated reads assigned to protist taxa in each COG category for
585 the Niskin sample (black) and the combined MS-SID samples (grey).
586
587 Supplementary Figure 1. Percentage of total annotated reads assigned to protist taxa in each
588 COG category for the Niskin sample (black) and the pooled MS-SID samples (grey).
589

590 Supplementary Figure 2. *DEGseq* analysis comparing differential expression between the three
591 replicate MS-SID transcript libraries vs. the Niskin library. A) MA-plot B) Boxplot of read
592 counts (log2) for each gene in the MS-SID vs. Niskin libraries. Y axis depicts the $\log_2$ fold
593 change in gene expression (M) and X axis depicts the mean of normalized counts (A). Red data
594 points are gene transcripts with significant differences ($p<0.001$) in expression levels.

595

596 **REFERENCES**

597

598 Andersson AF, Lundgren M, Eriksson S, Rosenlund M, Bernander R & Nilsson P (2006) Global
599 analysis of mRNA stability in the archaeon Sulfolobus. *Genome Biology* **7**: -.
600 Atkins MS, Teske AP & Anderson OR (2000) A survey of flagellate diversity at four deep-sea
601 hydrothermal vents in the Eastern Pacific Ocean using structural and molecular
602 approaches. *J Eukaryot Microbiol* **47**: 400-411.
603 Bartlett DH (2002) Pressure effects on in vivo microbial processes. *Biochimica et Biophysica*
604 *Acta* **1595**: 367-381.
605 Bianchi A, Garcin J & Tholosan O (1999) A high-ressure serial sampler to measure microbial
606 activity in the deep sea. *Deep-Sea Res I* **46**: 2029-2142.
607 Bird LE, Sherman AD & Ryan JP (2007) Development of an active, large volume, discrete
608 seawater sampler for autonomous underwater vehicles. ed.^eds.), p.^pp. Vancouver,
609 Canada.
610 Blake EW & Price B (2002) A proposed sterile sampling system for Antarctic subglacial lakes.
611 *Memoirs of the Institute for Polar Research* **Special Issue 56**: 253-263.
612 Brady A & Salzberg SL (2009) Phymm and PhymmBL: Metagenomic phylogenetic
613 classification with interpolated markov models. *Nature Methods* **6**: 673-676.
614 Cardell G, Hecht MH, Carsey FD, Engelhardt H, Fisher D, Terrell C & Thompson J (2004) The
615 subsurface ice probe (sipr): a low-power thermal probe for the martian polar layered
616 deposits. *35th Lunar and Planetary Science Conference, 15-19 March 2004, League City,*
617 *Texas, Abstract no. 2041*.
618 Caron DA (1991) Heterotrophic flagellates associated with sedimenting detritus. *In The Biology*
619 *of Free-living Heterotrophic Flagellates* (ed. D.J. Patterson and J. Larsen). pp. 77-92.
620 Oxford: Oxford University Press.
621 Carsey FD, Chen G-S, Cutts J*, et al.* (2000) Exploring Europa's ocean: a challenge for marine
622 technology of this century. *Mar Technol Soc J* **33**: 5-12.
623 Chastain RA & Yayanos AA (1991) Ultrastructural changes in an obligately barophilic marine
624 bacterium after decompression. *Appl Environ Microbiol* **57**: 1489-1497.
625 Danovaro R, Dell'Anno A, Pusceddu A & Fabiano M (1999) Nucleic acid concentrations (DNA,
626 RNA) in the continental and deep-sea sediments of the eastern Mediterranean:

627 relationships with seasonally varying organic inputs and bacterial dynamics. *Deep-Sea*
628 *Res.* **46**: 1077-1094.
629 Edgcomb V, Orsi W, Taylor GT, Vdacny P, Taylor C, Suarez P & Epstein S (2011)
630 Commentary: In situ fixation provides insights into the diversity of marine protists. *ISME*
631 *J*.
632 Edgcomb VP, Orsi W, Breiner H-W, Stock A, Filker S, Yakimov MM & Stoeck T (2011) Novel
633 kinetoplastids associated with hypersaline anoxic lakes in the Eastern Mediterranean
634 deep-sea. *Deep-Sea Res I* **58**: 1040-1048.
635 Edwards U, Rogall T, Blocker H, Emde M & Bottger EC (1989) Isolation and direct complete
636 nucleotide determination of entire genes: characterization of a gene coding for 16S
637 ribosomal RNA. *Nuc Acids Res* **17**: 7843-7853.
638 Feike J, Jurgens K, Hollibaugh JT, Kruger S, Jost G & Labrenz M (2012) Measuring unbiased
639 metatranscriptomics in suboxic waters of the central Baltic Sea using a new *in situ*
640 fixation system. *The ISME J* **6**: 461-470.
641 French LC, Anderson FS, Carsey FD, Green JR, Lane AL & Zimmerman WF (2001) Subsurface
642 exploration technologies and strategies for Europa. Forum on Innovative Approaches to
643 Outer Planetary Exploration 2001-2020. 30.
644 Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW & DeLong EF
645 (2008) Microbial community gene expression in ocean surface waters. *PNAS* **105**: 3805-
646 3810.
647 Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisolm SW & al. e (2008)
648 Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci USA*
649 **105**: 3805-3810.
650 Gifford SM, Sharma S, Rinta-Kanto JM & Moran MA (2011) Quantitative analysis of a deeply
651 sequenced marine microbial metatranscriptome. *ISME J* **5**: 461-472.
652 Glickman MH, Ciechanover A (April 2002). "The ubiquitin-proteasome proteolytic pathway:
653 destruction for the sake of construction". *Physiol. Rev.* **82** (2): 373–428.
654 Guo Y, Li C-I, Ye F, Shyr Y (2013) Evaluation of read count based RNAseq analysis methods.
655 *BMC Genomics* **14**(Suppl 8):52.
656 Hollibaugh JT, Gifford SM, Sharma S, Bano N & Moran MA (2010) Metatranscriptomic
657 analysis of ammonia-oxidizing organisms in an estuarine bacterioplankton assemblage.
658 *ISME J* **5**: 866-878.
659 Ishii A, Sato T, Wachi M, Nagai K & Kato C (2004) Effects of high hydrostatic pressure on
660 bacterial cytoskeleton FtsZ polymers *in vivo* and *in vitro*. *Microbiology* **150**: 1965-1972.
661 Jannasch HW & Maddux WS (1967) A note on bacteriological sampling in seawater. *J Mar Res*
662 **25**: 185-189.
663 Jannasch HW & Wirsen CO (1977) Retrieval of concentrated and undecompressed microbial
664 populations from the deep sea. *Appl Environ Microbiol* **33**: 642-646.
665 Jannasch HW, Wirsen CO & Winget CL (1973) A bacteriological pressure-retaining deep-sea
666 sampler. *Limnol and Oceanogr* **20**: 661-664.
667 La Cono V, Smedile F, Ferrer M, Golyshin PN, Giuliano L & Yakimov MM (2010) Genomic
668 signatures of fifth autotrophic carbon assimilation pathway in bathypelagic
669 *Crenarchaeota*. *Microbial Biotechnology* **3**: 595-606.
670 Lewis WM, McNail OD & Summerfelt RC (1963) A device for taking water samples in sterile
671 bottles at various depths. *Ecology* **44**: 171-173.

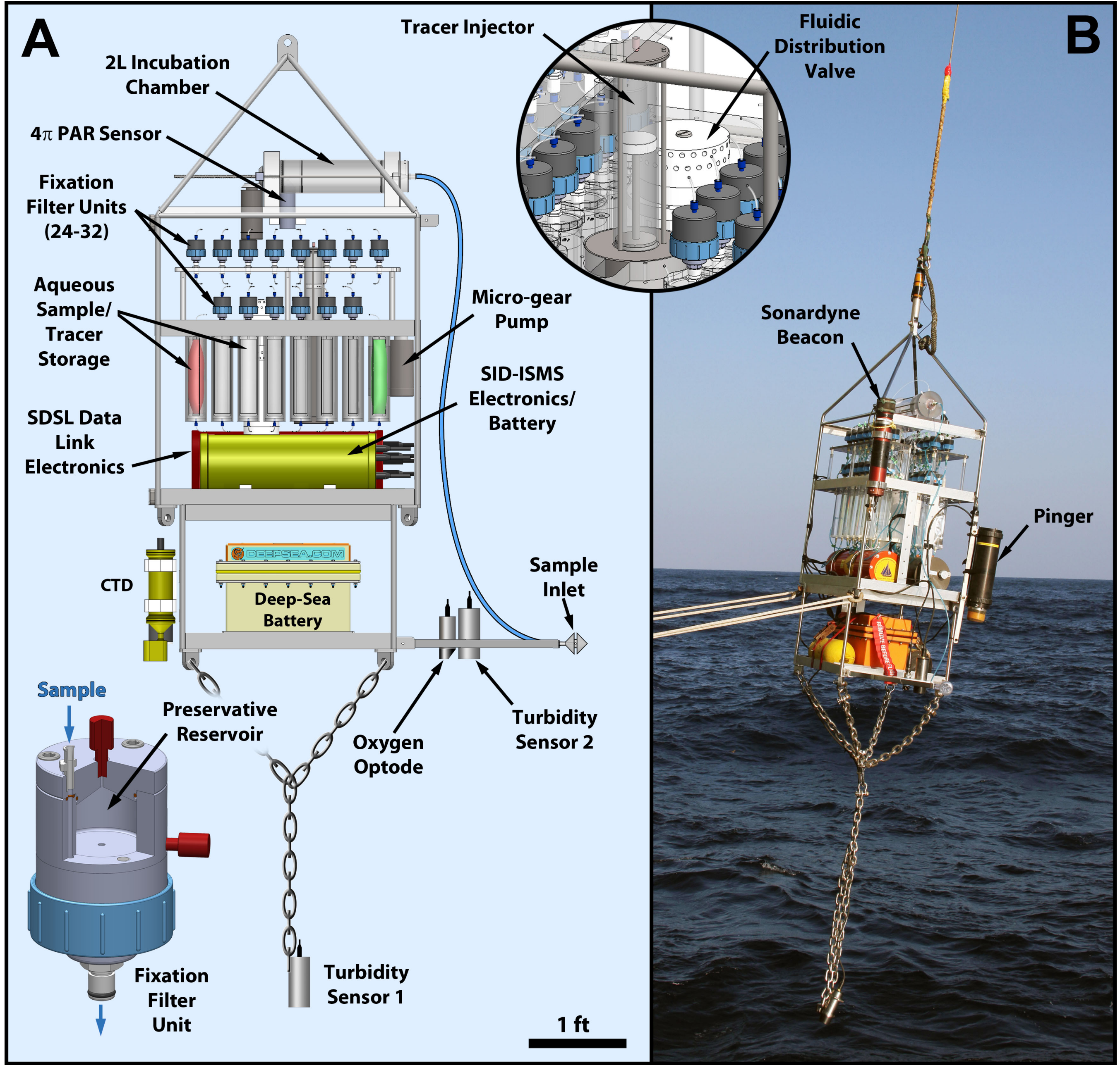
672 Li M, Jain S, Baker BJ, Taylor C & Dick GJ (2013) Novel hydrocarbon monooxygenase genes
673 in the metatranscriptome of a natural deep-sea hydrocarbon plume. *Environ Microbiol*.
674 Malahoff A, Gregory T, Bossuyt A, Donachie S & Alam M (2002) A seamless system for the
675 collection and cultivation of extremophiles from deep-ocean hydrothermal vents. *IEEE J*
676 *Oceanic Engineer* **27**: 862-869.
677 Marchetti A, Schruth DM, Durkin CA*, et al.* (2012) Comparative metatranscriptomics identifies
678 molecular bases for the physiological responses of phytoplankton to varying iron
679 availability. *Proc Natl Acad Sci USA*.
680 Marteinsson VT, Reysenbach A-L, Birrien J-L & Prieur D (1999) A stress protein is induced in
681 the deep-sea barophilic hyperthermophile *Thermococcus barophilus* when grown under
682 atmospheric pressure. *Extremophiles* **3**: 277-282.
683 Mason OU, Hazen TC, Borglin S, Chain PSG, Dubinsky EA, Fortney JL, Han J, Holman H-YN,
684 Hultman J, Lamendella R, Mackelprang R, Malfatti S, Tom LM, Tringe SG, Woyke T,
685 Zhou J, Rubin EM, & Jansson JK (2012) Metagenome, metatranscriptome and single-
686 cell sequencing reveal microbial response to Deepwater Horizon oil spill. *The ISME J*
687 6:1715-1727.
688 Nikin SJ (1962) A water sampler for microbiological studies. *Deep-Sea Res.* **9**: 501-503.
689 Olson RJ & Sosik HM (2007) A submersible imaging-in-flow instrument to analyze nano- and
690 microplankton: Imaging FlowCytobot. *Limnol and Oceanogr: Methods* **5**: 195-203.
691 Orsi W, Edgcomb VP, Christman GD & Biddle JF (2013) Gene expression in the deep
692 biosphere. *Nature*.
693 Patterson DJ, Nygaard K, Steinberg G & Turley CM (1993) Heterotrophic flagellates and other
694 protists associated with oceanic detritus throughout the water column in the mid North
695 Atlantic. *J Mar Biol Ass U.K.* **73**: 67-95.
696 Ottesen EA, Young CR, Eppley JM, Ryan JP, Chavez FP, Scholin CA, DeLong EF (2013)
697 Pattern and synchrony of gene expression among sympatric marine microbial
698 populaitons. *PNAS* **110**:E488-E497, doi: 10.1073/pnas.1222099110.
699 Ottesen EA, Martin R 3rd, Preston CM, Young CR, Ryan JP, Scholin CA, DeLong EF (2011)
700 Metatranscriptomic analysis of autonomously collected and preserved marine
701 bacterioplankton. *ISME J* **5**:1881-95.
702 Phillips H, Wells LE, Johnson II RV, Elliott S & Deming JW (2003) LAREDO: a new
703 instrument for sampling and in situ incubation of deep-sea hydrothermal vent fluids.
704 *Deep-Sea Res I* **50**: 1375-1387.
705 Poretsky RS, Hewson I, Sun S, Allen AE, Zehr JP & Moran MA (2009) Comparative day/night
706 metatranscriptomic analysis of microbial communities in the North Pacific subtropical
707 gyre. *Environ Microbiol* **11**: 1358-1375.
708 Rivers AR, Sharma S, Tringe SG, Martin J, Joye SB & Moran MA (2013) Transcriptional
709 response of bathypelagic marine bacterioplankton to the Deepwater Horizon oil spill. *The*
710 *ISME J* 7:2315-2329.
711 Roman BR, Scholin C, Jensen S*, et al.* (2007) Controlling a robotic marine water sampler with
712 the Ruby scripting language. *J Assoc Lab Automat* **12**: 56-61.
713 Ryan JP, Johnson SB, Sherman A*, et al.* (2010) Mobile autonomous process sampling within
714 coastal ocean observing systems. *Limnol and Oceanogr* **8**: 394-402.
715 Sarmiento JL, Herbert T & Toggweiler JR (1988) Mediterranean nutrient balance and episodes
716 of anoxia. *Global Biogeochemical Cycles* **2**: 427-444.

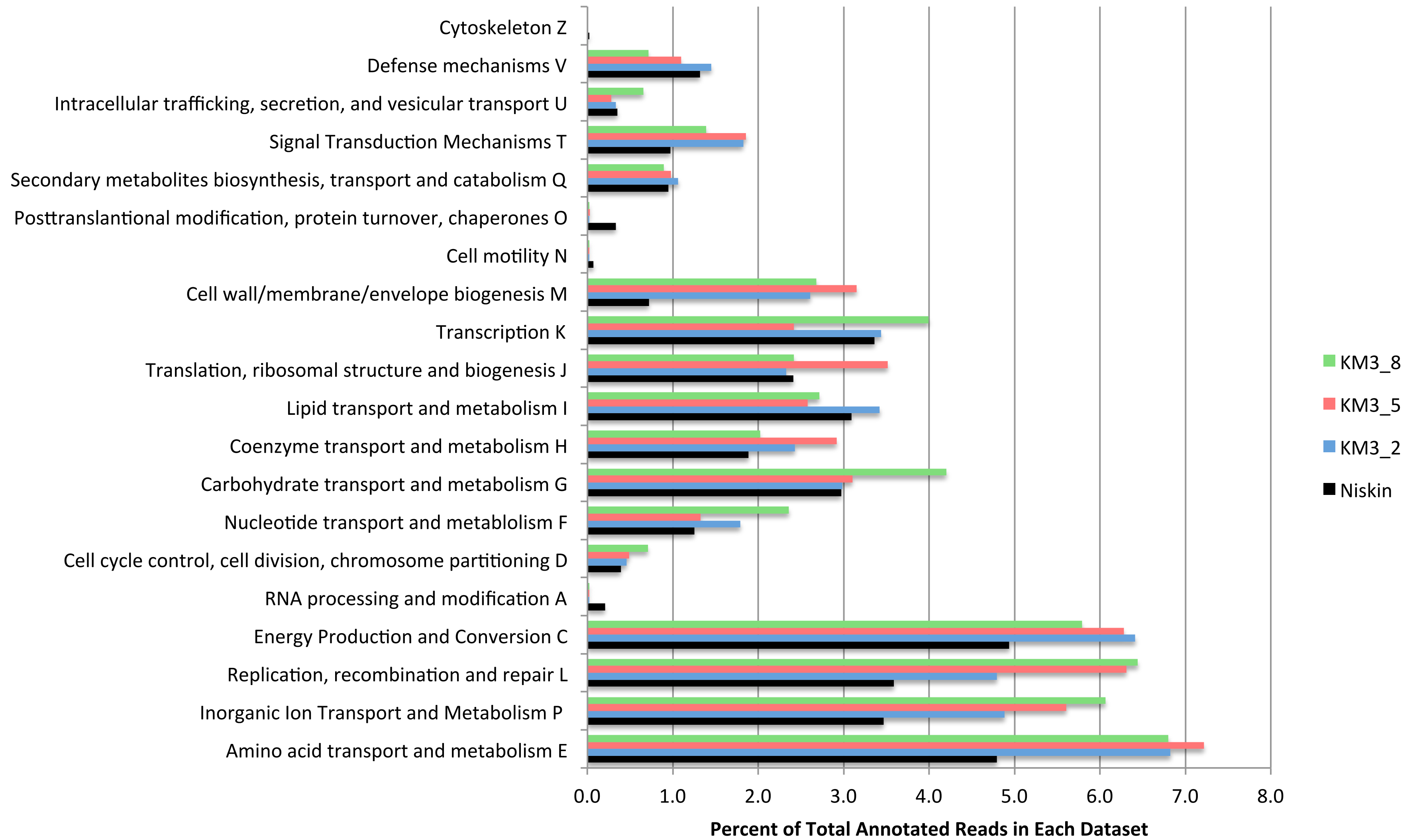
717 Scholin C, Jensen S, Roman B*, et al.* (2006) The Environmental Sample Processor (ESP) - An
718 automated robotic device for detecting microorganisms remotely using molecular probe
719 technology. *Proceedings, OCEANS 2006 MTS/IEEE Conference, Boston, MA. Marine*
720 *Technology Society, Columbia MD*.
721 Scholin CA (2010) What are "ecogenomic sensors?" A review and thoughts for the future. *Ocean*
722 *Sci* **6**: 51-60.
723 Siegert MJ, Tranater M, Ellis-Evans JC, Priscu JC & Lyons WB (2003) The hydrochemistry of
724 Lake Vostok and the potential for life in Antarctic subglacial lakes. *Hydrological Proc*
725 **17**: 795-814.
726 Sosik HM & Olson RJ (2007) Automated taxonomic classification of phytoplankton sampled
727 with imaging-in-flow cytometry. *Limnol and Oceanogr: Methods* **5**: 204-216.
728 Stackebrandt E & Liesack W (1993) *Nucleic acids and classification*. Academic Press, London,
729 England.
730 Steglich C, Lindell D, Futschik M, Rector T, Steen R & Chisolm SW (2010) Short RNA half-
731 lives in the slow-growing marine cyanobacterium *Prochlorococcus*. *Genome Biology* **11**.
732 Swartz M, Torres D, Liberatore S & Millard R (2012) WHOI SDSL Data Link Project - Ethernet
733 Telemetry through sea cables. *J Atmos Oceanic Technol* **in press**.
734 Tabor PS, Ohwada K & Colwell RR (1981) Filterable marine bacteria found in the deep sea:
735 distribution, taxonomy, and response to starvation. *Microb Ecol* **7**: 67-83.
736 Taylor C & Howes B (1994) Effect of sampling frequency on measurements of seasonal primary
737 production and oxygen status in near-shore coastal ecosystems. *Marine Ecology Progress*
738 *Series* **108**: 193-203.
739 Taylor C, Howes B & Doherty K (1993) Automated instrumentation for time-series
740 measurement of primary production and nutrient status in production platform-accessible
741 environments. *Marine Technology Society Journal* **27**: 32-44.
742 Taylor C, Doherty KW, Molyneaux SJ*, et al.* (2006) Autonomous Microbial Sampler (AMS), a
743 device for the uncontaminated collection of multiple microbial samples from submarine
744 hydrothermal vents and other aquatic environments. *Deep-Sea Res I* **56**: 1266-1283.
745 Taylor CD & Doherty KW (1990) Autonomous instrumentation for the in situ measurement of
746 primary production and other microbial rate processes. *Deep-Sea Res* **37**: 343-358.
747 Thingstad TF, Krom MD, Mantoura RFC*, et al.* (2005) Nature of phosphorus limitation in the
748 ultraoligotrohic Eastern Mediterranean. *Science* **309**: 1068-1071.
749 Turley CM & Carstens M (1991) Pressure tolerance of oceanic flagellates: implications for
750 remineralization of organic matter. *Deep-Sea Res* **39**: 403-413.
751 Ulloa O, Canfield DE, DeLong EF, Letelier RM & Stewart FJ (2012) Microbial oceanography of
752 anoxic oxygen minimum zones. *Proc Natl Acad Sci USA* **109**: 15996-16003.
753 Wang L, Feng Z, Wang X, Wang X & Zhang X (2009) DEGseq: an R package for identifying
754 differentially expressed genes from RNA-seq data. *Bioinformatics* **26**: 136-148.
755 Wang LK, Feng Z, Wang X & Zhang X (2010) DEGseq: an R package for identifying
756 differentially expressed genes from RNA-seq data. *Bioinformatics* **26**: 136-138.
757 Wang Y, Liu CL, Storey JD, Tibshirani RJ, Herschlag D & Brown PO (2002) Precision and
758 functional specificity in mRNA decay. *Proc Natl Acad Sci USA* **99**: 5860-5865.
759 Weber G & Drickamer HG (1983) The effect of high pressure upon proteins and other
760 biomolecules. *Q Rev Biophys* **16**: 89-112.
761 Weizhong L (2009) Analysis and comparison of very large metagenomes with fast clustering and
762 functional annotation. *BMC Bioinformatics* **10**: 359.

763 Yano Y, Nakayama A, Ishihara K & Saito H (1998) Adaptive changes in membrane lipids of
764 barophilic bacteria in response to changes in growth pressure. *Appl Environ Microbiol*
765 **64**: 479-485.
766 Zobell CE (1941) Apparatus for collecting water samples from different depths for
767 bacteriological analysis. *J Mar Res* **4**: 173-188.

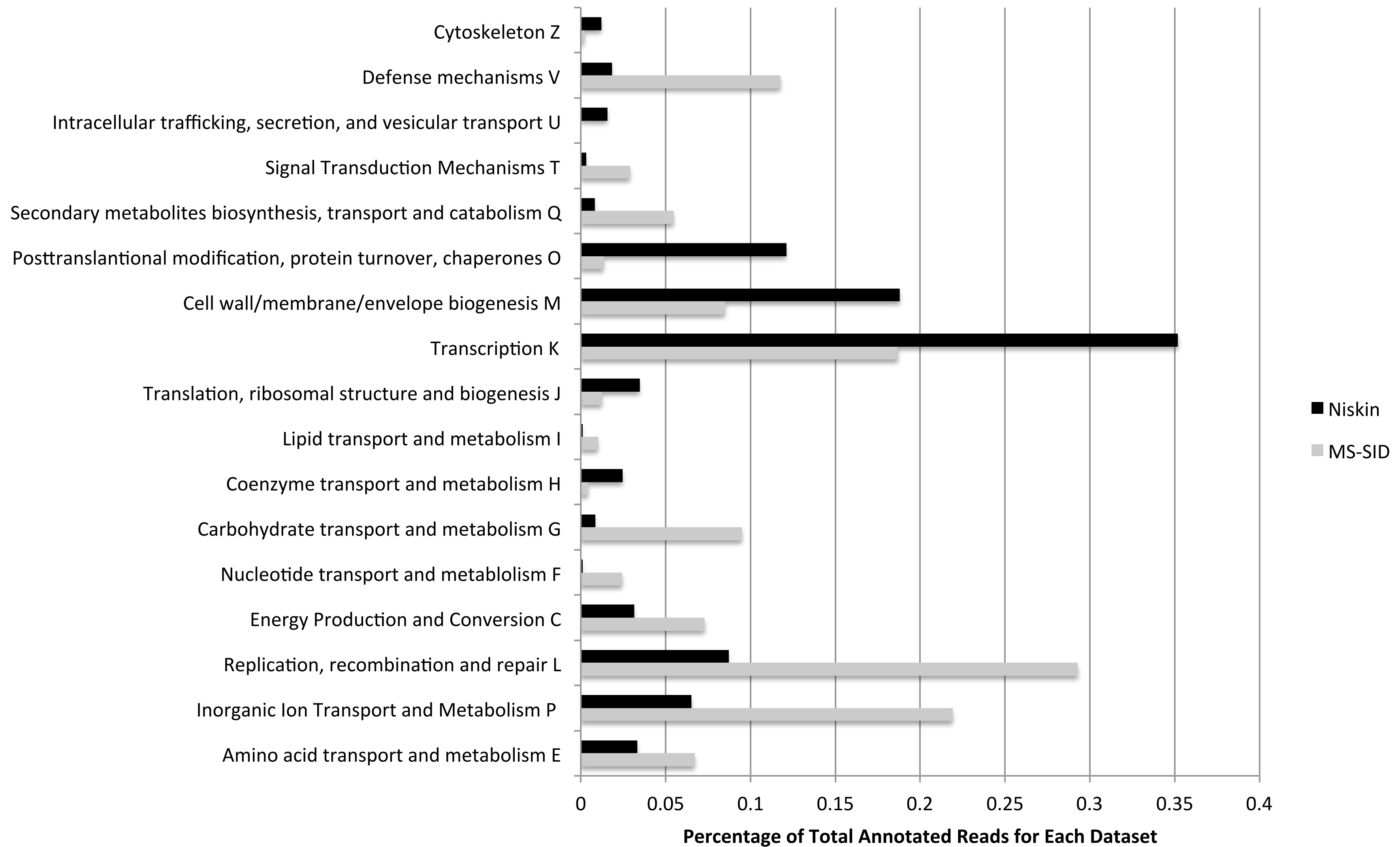
768

769


A
2L Incubation Chamber
4π PAR Sensor
Fixation Filter Units (24-32)
Aqueous Sample/ Tracer Storage
SDSL Data Link Electronics
Tracer Injector
Micro-gear Pump
SID-ISMS Electronics/ Battery
CTD
DEEPSEA.COM
Deep-Sea Battery
Sample Inlet
Sample
Preservative Reservoir
Oxygen Optode
Turbidity Sensor 2
Fixation Filter Unit
Turbidity Sensor 1
1 ft
B
Fluidic Distribution Valve
Sonardyne Beacon
Pinger



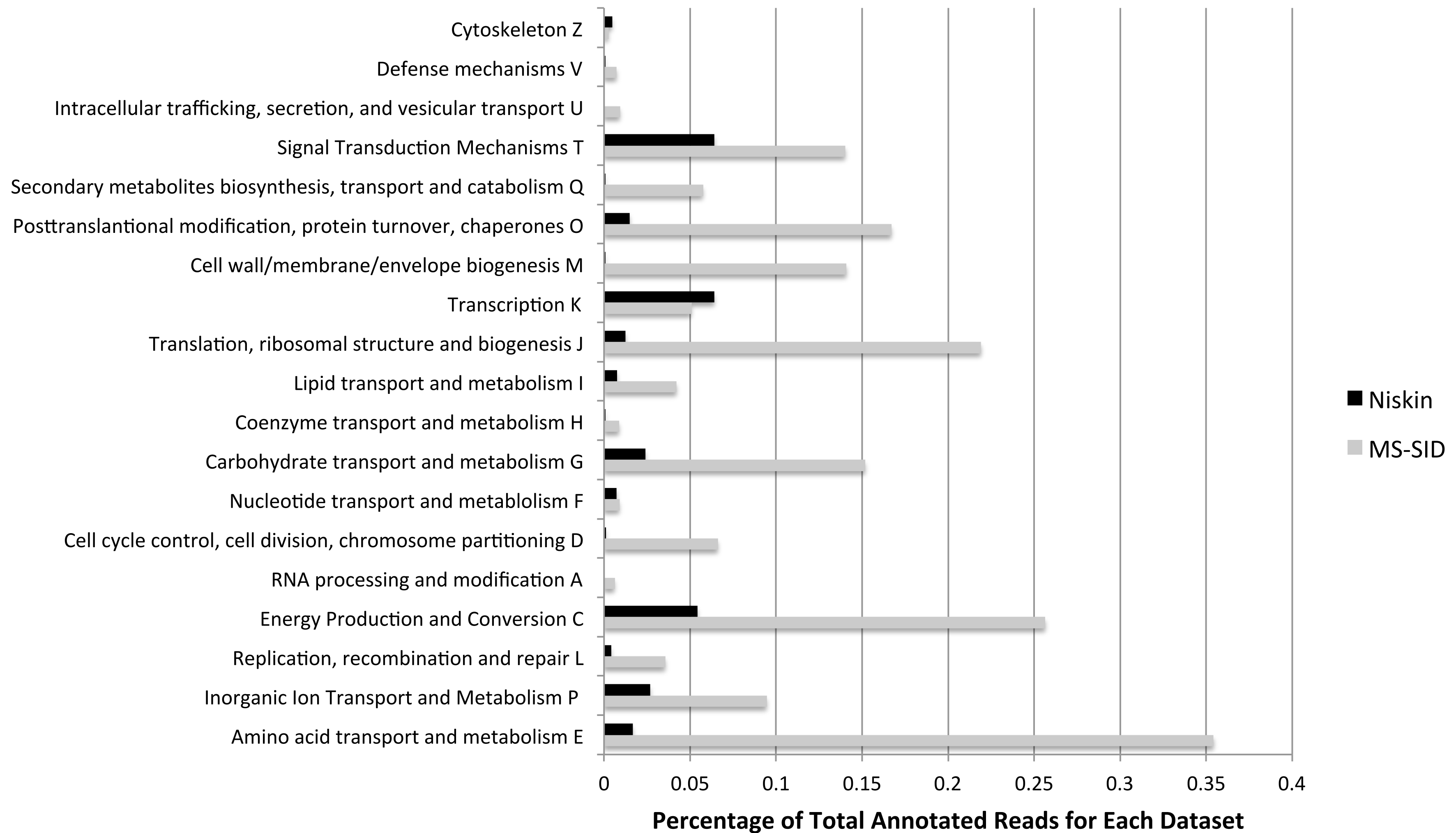
% Representation of COG Categories for Gammaproteobacterial Transcripts Recovered From the Niskin Water Sample and Each of the Three MS-SID Replicate Water Samples
Cytoskeleton Z
Defense mechanisms V
Intracellular trafficking, secretion, and vesicular transport U
Signal Transduction Mechanisms T
Secondary metabolites biosynthesis, transport and catabolism Q
Posttranslantional modification, protein turnover, chaperones O
Cell motility N
Cell wall/membrane/envelope biogenesis M
Transcription K
Translation, ribosomal structure and biogenesis J
Lipid transport and metabolism I
Coenzyme transport and metabolism H
Carbohydrate transport and metabolism G
Nucleotide transport and metablolism F
Cell cycle control, cell division, chromosome partitioning D
RNA processing and modification A
Energy Production and Conversion C
Replication, recombination and repair L
Inorganic Ion Transport and Metabolism P
Amino acid transport and metabolism E
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
Percent of Total Annotated Reads in Each Dataset
KM3_8
KM3_5
KM3_2
Niskin

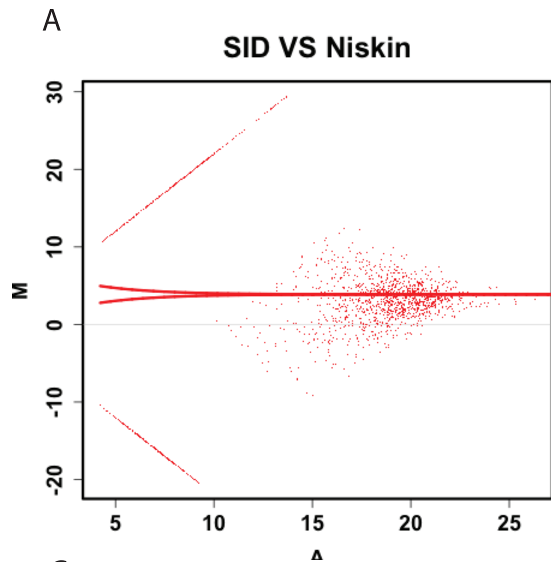


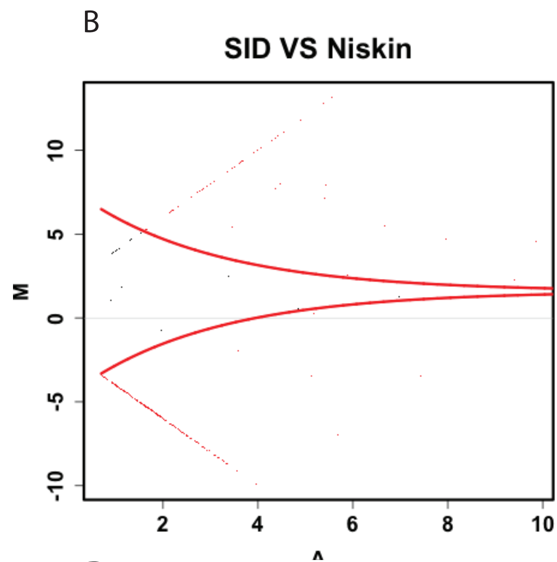
Percent Representation of COG Categories for Protist Transcripts
Cytoskeleton Z
Defense mechanisms V
Intracellular trafficking, secretion, and vesicular transport U
Signal Transduction Mechanisms T
Secondary metabolites biosynthesis, transport and catabolism Q
Posttranslantional modification, protein turnover, chaperones O
Cell wall/membrane/envelope biogenesis M
Transcription K
Translation, ribosomal structure and biogenesis J
Lipid transport and metabolism I
Coenzyme transport and metabolism H
Carbohydrate transport and metabolism G
Nucleotide transport and metablolism F
Energy Production and Conversion C
Replication, recombination and repair L
Inorganic Ion Transport and Metabolism P
Amino acid transport and metabolism E
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
Percentage of Total Annotated Reads for Each Dataset
Niskin
MS-SID

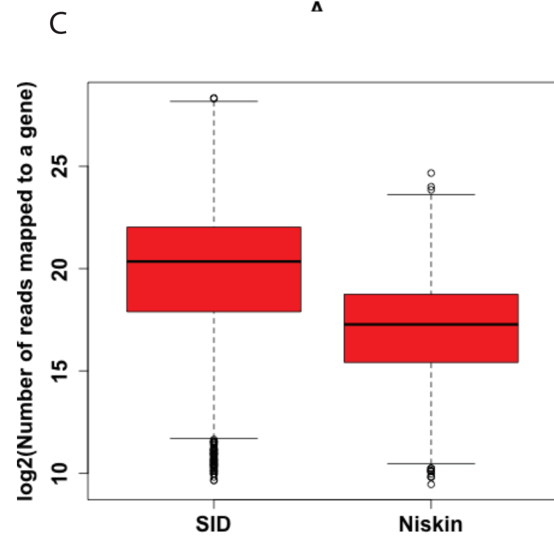
Table 1. Overall sequencing results for each sample collected from 2222m depth.

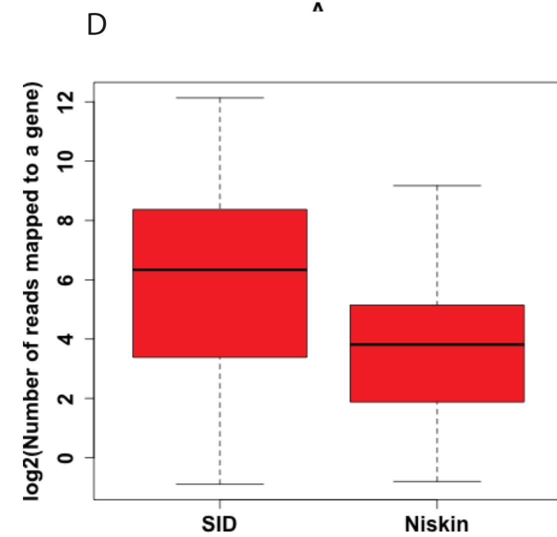
Sample	**Volume Filtered (L)**	**Initial Reads**	**Initial Reads After Trimming**	**Contigs**	**Total Annotated Reads**	**Total Annotated Contigs**
Niskin	30	78,307,034	37,947,773	41,372	8,968,832	11,888
MS-SID_1	4	122,435,482	65,237,000	20,973	26,658,601	9,731
MS-SID_2	3	96,136,502	51,529,890	18,606	25,003,471	8,542
MS-SID_3	3.4	118,922,488	63,968,753	17,000	27,057,430	6,153


Percent Representation of COG Categories for Fungal Transcripts
Cytoskeleton Z
Defense mechanisms V
Intracellular trafficking, secretion, and vesicular transport U
Signal Transduction Mechanisms T
Secondary metabolites biosynthesis, transport and catabolism Q
Posttranslantional modification, protein turnover, chaperones O
Cell wall/membrane/envelope biogenesis M
Transcription K
Translation, ribosomal structure and biogenesis J
Lipid transport and metabolism I
Coenzyme transport and metabolism H
Carbohydrate transport and metabolism G
Nucleotide transport and metablolism F
Cell cycle control, cell division, chromosome partitioning D
RNA processing and modification A
Energy Production and Conversion C
Replication, recombination and repair L
Inorganic Ion Transport and Metabolism P
Amino acid transport and metabolism E
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
Percentage of Total Annotated Reads for Each Dataset
Niskin
MS-SID


A
SID VS Niskin
M
30
20
10
0
-10
-20
5
10
15
20
25


B
SID VS Niskin
M
10
5
0
-5
-10
2
4
6
8
10


C
log2(Number of reads mapped to a gene)
25
20
15
10
SID
Niskin


D
log2(Number of reads mapped to a gene)
12
10
8
6
4
2
0
SID
Niskin