Comparison of Niskin vs. in situ approaches for analysis of gene expression in deep Mediterranean Sea water samples

V. P. Edgcomb¹*, C. Taylor²*, M. G. Pachiadaki¹, S. Honjo³, I. Engstrom³, M. Yakimov⁴,

¹Woods Hole Oceanographic Institution, Department of Geology and Geophysics, Woods Hole, MA, USA
²Woods Hole Oceanographic Institution, Department of Biology, Woods Hole, MA, USA
³McLane Research Laboratories, Inc., East Falmouth, MA 02536, USA
⁴CNR – Institute for Coastal Marine Environment, Messina, Italy

Corresponding and Co-First Authors:
*V.P. Edgcomb, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, Tel: +01 508-289-3734, Fax: +01 508-457-2192, Email: vedgcomb@whoi.edu and C. Taylor, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, Tel: +01 508-289-2354, Email: ctaylor@whoi.edu
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
Microbial metabolic activities are the basis of almost every major biogeochemical cycle in the oceans, and as the research community transitions away from purely descriptive studies of marine microbes to system-level investigations of community activity and responses to changing environmental conditions, it is imperative that we obtain less biased samples for those studies. As a consequence of the fact that the majority of microorganisms are not amenable to existing cultivation approaches, many marine microbiologists and microbial ecologists have embraced culture-independent methods. Metatranscriptomics, or the isolation and sequencing of messenger RNA (mRNA) from an environmental sample, is one powerful method currently used for linking diversity with activity, and for examining microbial activities in response to changing conditions. Metatranscriptomics provides an overview of (at a minimum) the most highly expressed genes in a sample. These transcripts inform about the metabolic pathways that are utilized by microbiota in that sample at the time of sample preservation, and specific proteins that were expressed. Enabled by recent advances in high-throughput sequencing technologies and bioinformatics for processing datasets that can contain tens of millions of reads, metatranscriptomics has become one of the most powerful tools for examining microbial community activities.

The metatranscriptomics approach has been used successfully to examine gene expression in varied marine habitats. Examples include deep subsurface sediments (Orsi, et al., 2013), the North Pacific Subtropical gyre (Frias-Lopez, et al., 2008, Poretsky, et al., 2009), eastern tropical South Pacific oxygen minimum zone (Ulloa, et al., 2012), coastal waters (Hollibaugh, et al., 2010, Gifford, et al., 2011), hydrothermal vent plumes (Li, et al., 2013), microcosm experiments on mixed water layers from the NE Pacific Ocean (Marchetti, et al., 2012). Metatranscriptomics was also recently used to study microbial responses to the Deep
Water Horizon oil spill. Mason et al. (2012) found a rapid increased expression of genes associated with motility, chemotaxis, and aliphatic hydrocarbon degradation originating from members of the *Oceanospirallales* in hydrocarbon plume samples (Mason et al. 2012). Rivers et al. (2013) also applied metatranscriptomics to show increased activity (primarily associated with methane- and petroleum-degrading Gammaproteobacteria) within specific metabolic pathways for the degradation of alkanes, aromatic compounds and methane following the Deepwater Horizon spill. Metatranscriptomic studies in the future will be enhanced by sampling technologies that allow us to minimize potential artifacts that can be introduced due to sample handling.

For studies of gene expression it is necessary to minimize time between sample collection and chemical preservation. Historically, oceanographers studying marine microbiota have relied on ship-based hydrocasting operations whereby water samples from various depths in the ocean are brought to the surface via Niskin rosette samplers for shipboard water processing. This approach is likely not appropriate for mRNA-based investigations due to the typically large and variable lapses in time and accompanying physicochemical shifts samples are exposed to between collection and preservation. When working in the near-surface (top 1-200m) up to 30 minutes may pass between when a water sample is collected and when it is returned to the surface, preserved and processed in the ship’s laboratory. In addition, samples are exposed to pressure, and potentially temperature and redox changes before preservation. Pressure changes, potential physicochemical changes, and separation in time between sample collection and processing are exacerbated when working in the deep sea and/or collecting water samples from low-oxygen or anoxic zones. Impacts on transcription by microbes captured in water samples are also likely to vary between taxa depending fragility of cell structures, and on their strategies (or
lack thereof) for responding to such changes. Given that average lifetimes of prokaryotic transcripts can be on the order of several minutes (Wang, et al., 2002, Andersson, et al., 2006, Steglich, et al., 2010), the gene expression profiles of microorganisms can potentially be altered significantly. While delays in the preservation of DNA and rRNA (often used as a phylogenetic identifier of viable organisms) is less susceptible to such biases because of their significantly longer half-lives, delayed preservation of Niskin samples can still be an issue if cell integrity is lost due to unnatural and changing conditions during transport from the ocean to the ship's laboratory. Variable and unknown fractions of genetic material from lysed cells can be lost during filtration. This problem is particularly severe for microbial eukaryotes and potentially compounded when sampling greater depths (Edgcomb, et al., 2011).

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

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

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

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

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

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

The MS-SID collected three consecutive and separate samples for this analysis from the exact same depth as the Niskin rosette based on pressure readings from both instruments, and were collected within 3 hours of the Niskin collections. All samples were filtered *in situ* at 125 ml/min through a 47mm 0.2 µm Durapore (Millipore, USA) filter that, upon cessation of filtration was within 10-20 seconds flooded with the preservative RNAlater following filtration. The three filters collected 4L, 3L and 3.4L of water, respectively. Upon retrieval of the instrument to the ship’s deck, the Fixation Filter Units were disassembled, and the filters and associated RNAlater solutions were transferred aseptically to three separate sterile cryovials, and
frozen at -80°C until extraction. The Sterivex capsule (from Niskin bottle collection) was also stored at -80°C until extraction.

Extraction of RNA: For each of the 3 MS-SID filters total RNA was extracted separately from the filter and RNAlater in order to capture the mRNA from any cells that may have lysed during sample storage. The filters were transferred to 2ml tubes and homogenized without beads for 20 seconds at speed 4.0m s⁻¹ on a FastPrep®-24 homogenizer (MP Biomedicals, Solon, OH, USA) in 600 µl of RLT Plus buffer (All Prep DNA/RNA Mini Kit, Qiagen). An equal volume of 70% ethanol was added and the manufacturer’s instructions were followed for remaining steps. The final volume of extracted RNA was 50 µl in distilled deionized water. The RNAlater from each filter’s original tube was transferred to a 50 ml centrifuge tube. Four ml of RLT buffer (RNeasy Midi Kit, Qiagen, USA) were added and gently mixed. An equal volume of 70% ethanol was added and mixed. Total RNA was then extracted using the RNeasy Midi Kit protocol beginning at step 3 and contained in 500 µl of distilled deionized water. Since some cell lysis may have occurred between sample preservation in situ and recovery of the filter from each Fixation Filter Unit, we also retained and extracted RNA from the RNAlater contained within the Fixation Filter Unit. For each filter separately, this RNAlater was transferred to a 50 ml centrifuge tube (~15ml). Approximately 7.5 ml of Buffer RLT were added and the samples were run through the All Prep DNA spin column (Qiagen, USA). An equal volume of 70% ethanol was mixed with the sample and total RNA was then extracted using the RNeasy Midi Kit beginning at step 3 and contained in 500µl of distilled, deionized water. The combined extractions for each sample were cleaned and precipitated following the instructions in the MEGAclear kit protocol (Life Technologies, Grand Island NY, USA). The RNA pellet was suspended in 100 µl of distilled,
deionized water, and DNAsed for 1 hour at 37°C (Turbo DNase, Life Technologies, Grand Island
NY, USA). Finally all samples were purified using the MegaClear kit one final time.

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

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

Taxonomic assignments of contigs were made using PhymmBL (Brady & Salzberg, 2009),
incorporating all available fungal and protist genomes in public databases. The total number of annotated reads assigned to different COG families for each dataset was expressed as a percentage of the total annotated reads for each dataset.

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

**RESULTS**

*Study site:* At 2222m depth, the water temperature was 13.81°C, and salinity was 38.723 PSU. Oxygen concentration was 204µM, representative of the oxygen profiles between the near-surface and 2222m at this site. Nitrate and phosphate concentrations were measured previously at this site (La Cono *et al.* 2010), showing an increasing trend with depth, reaching 3.75 +/- 0.54
and 0.13 +/- 0.02 µmol L^{-1}, respectively at 3010m depth. Nitrite was constant at 0.04 +/- 0.01 µmol L^{-1} from the sea surface to 3010 m.

Transcriptome library sequencing: For each of the 4 samples (one Niskin 30L water sample and 3 MS-SID samples) the number of recovered reads, reads remaining after trimming, number of contigs formed, annotated reads, and annotated contigs can be found in Table 1. 78.3 million sequence reads were obtained from the Niskin sample. From the 4L, 3L, and 3.4L MS-SID samples, 122.4, 96.1, and 118.9 million sequence reads were obtained, respectively. After trimming sequences for low-quality and short reads, we obtained 37.9, 66.3, 51.5, and 63.9 million reads from the Niskin, MS-SID 4L, MS-SID 3L, and MS-SID 3.4L samples, respectively. Of the total reads that passed quality control, 8.9 million were possible to annotate from the Niskin sample, and these formed 11.8 thousand annotated contigs. From the MS-SID replicate samples, 26.6, 25.0, and 27.0 million reads were annotated, and these formed 9.7, 8.5, and 6.1 thousand annotated contigs. Data are deposited in CAMERA (https://portal.camera.calit2.net) and the Short Read Archive (SRP042349).

Transcriptome library analyses. The majority of annotated reads for both MS-SID and Niskin samples affiliated with the Gammaproteobacteria (47% and 38% for MS-SID and Niskin samples, respectively, data not shown). Percent representation of COG categories (out of total reads recovered for each library) for gammaproteobacterial transcripts recovered from the Niskin sample and from each of the three MS-SID replicate samples is shown in Figure 2 for all COG categories representing >0.001% of total reads in either library. For COG categories T (signal transduction mechanisms), M (cell wall, membrane, and envelope biogenesis), H (coenzyme
transport and metabolism), F (nucleotide transport and metabolism), C (energy production and
conversion), L (replication, recombination and repair), P (inorganic ion transport and
metabolism), and E (amino acid transport and metabolism), percent representation out of total
reads for Gammaproteobacteria was higher by ~1-2% in the MS-SID samples. In the case of
COGs V (defense mechanisms), U (intracellular trafficking, secretion, and vesicular transport), Z
(cytoskeleton), Q (secondary metabolites biosynthesis, transport and catabolism), N (cell
motility), K (transcription), J (translation, ribosomal structure and biogenesis), I (lipid transport
and metabolism), G (carbohydrate transport and metabolism), D (cell cycle control, cell division,
chromosome partitioning), expression levels as a percentage of total annotated reads for each
library were approximately the same (within approximately half a percent) in the Niskin and MS-
SID samples. For COG categories O (posttranslational modification, protein turnover,
chaperones), and A (RNA processing and modification), expression levels were slightly higher in
the Niskin sample than in each of the MS-SID samples. The COG categories with the most
abundant annotated transcripts for Gammaproteobacteria included amino acid transport and
metabolism (4.8-7.3% of total reads in each of the 4 libraries), inorganic ion transport and
metabolism (3.5-6.1%), replication, recombination and repair (3.5-6.4%), and energy production
and conversion (4.9-6.4%) (Figure 2).

The degree of variation in gene expression profiles, defined here as the relative
abundance of annotated transcripts, between the three MS-SID replicates was assessed using the
DEGseq package, which revealed 9.2-9.8% variation in pairwise expression profiles (relative
transcript abundance) between replicates. Differentially expressed transcripts included
peptidases, three RNA methylases, ATPases, and a NCAIR synthetase. MA plots for pairwise
comparisons of the three replicates depicting the log₂ fold change in gene expression vs. the
mean of normalized counts showed a balanced distribution of the contigs (data not shown).

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

Fungal transcripts represented 1.8 and 0.3% of the MS-SID and Niskin datasets,
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
waters were processed, or due to differences in total volumes filtered (10.4L total for MS-SID and 30L for Niskin), it is intuitive to consider that increased handling and physicochemical alterations prior to preservation likely alters the profile of community activities. Sample collection approaches that rely on returning water samples to the surface prior to processing and preservation may be completely appropriate for many types of investigations, however our results and the results of Feike et al. who conducted a similar experiment using water samples from 70-120m depth in the Baltic Sea (Feike, et al., 2012) suggest that for examinations of microbial activities, in situ preservation is desirable. This may be increasingly important for deeper water samples.

Obtaining an accurate snapshot of in situ microbial activities requires collecting and preserving samples in a fashion that minimizes detectable environmental perturbations. All existing water-sampling technologies that involve drawing water through intake plumbing into a chamber or bag or drawing water through a filter have the potential to introduce artifacts resulting from detected perturbations that cause microbes to alter transcription of particular types of genes. Nonetheless, the primary advantage of in situ approaches is that they minimize artifacts associated with sample handling, and allow for almost immediate preservation of sensitive molecules. Furthermore, moored in situ technologies are useful for studies of temporal variations in microbial communities. For example, Ottesen et al. (2011) used a moored ESP to automatically filter and preserve (in RNAlater) near-surface water samples from Monterey Bay at four time points over the course of a day. Using this technology to capture and preserve RNA allowed these researchers to track changes in community composition and gene expression over this timeframe. This was expanded upon by Ottesen and colleagues (2013) who used the moored ESP to detect coordinated gene expression between microbial groups over a 2 day timeframe at
23m depth off the coast of California, possibly in response to specific environmental cues.

Preserving samples *in situ* in the deep ocean prior to returning them to the sea surface has the obvious advantage of avoiding potentially significant pressure and other physicochemical changes that can occur during sample retrieval and prior to processing/preservation in the ship laboratory. Preserving a whole water sample immediately upon collection by drawing it into a bag preloaded with preservative (as can be performed by the MS-SID) or into a container that is immediately injected with preservative (Feike, *et al.*, 2012) accomplishes this goal. However, this approach limits the volume that one can collect. In some cases, such as for water samples where cell densities are low, it is desirable to collect information (e.g., RNA) from a much greater volume of water via filtration. During filtration and prior to preservation, cells impinged on filters earlier in the filtration routine undoubtedly experience more significant environmental changes than those that are captured toward the end of filtration. This is a concern not only for Niskin-based sample collection, but also for *in situ* filtration technologies such as the MS-SID. In contrast to waters collected using *in situ* filtration and preservation methods, all cells on filters from Niskin water samples have undergone depressurization and have spent variable lengths of time on deck prior to filtration. This is not the case for *in situ* preserved filters, where until impinged on a filter the cells are in their natural environment. The composite mRNA pools within the filtered organisms reflect an average residence time on the filter surface that is approximately one-half the total filtration time (e.g., ~10 min for our 20 min filtrations), which additionally enhances the fidelity of the samples collected. While the MS-SID will capture *in situ* gene expression patterns with improved fidelity, there may be some genes that reflect the stresses of being impinged on a filter during filtration. In all cases, artifacts introduced into expression 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 log₂ 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.
It is noteworthy that the Niskin sample produced 64-81% of the number of initial reads obtained from the 3 MS-SID samples, and 59-73% after quality control trimming, in spite of identical library preparation protocols, equal sequencing efforts and the fact that the Niskin filter represented 30L of filtered water compared to 3-4L of water for each of the MS-SID samples. This lower number of initial reads however, assembled into roughly twice as many contigs. Approximately 23% of Niskin initial reads were successfully annotated, in comparison to 49, 40, and 42% of the initial reads for the 3 MS-SID replicates, and 29% of Niskin contigs were successfully annotated vs. an average of 44% of MS-SID contigs (Table 1). The N50 lengths for the MS-SID contigs were slightly higher than the Niskin sample (473, 447, and 428 vs. 424 bp), however this does not likely explain differences in contig formation or annotation. We hypothesize that some aspect of Niskin water collection and processing (pressure changes, increased time delays between sample collection and processing, or other unknown physicochemical alterations to the sample) reduced the community complexity in this sample (possible cell lysis, cell death) for some taxa. Lower complexity could lead to greater contig formation, and if remaining complexity favored taxa that are less represented in public sequence databases, then this may explain the lower percentage of annotated contigs for the Niskin sample. Pressure changes and sample handling are likely to have differential impacts on cell integrity and gene transcription of different taxa, and some cell activities are also likely to be more sensitive to perturbations than others. This was observed in our datasets. For example, for Gammaproteobacteria, which appear to be one of the dominant bacterial groups in our samples, certain COG categories appear to show a greater difference than others in gene expression between the Niskin and MS-SID data sets (Figure 2). However, a test for differential expression revealed that almost all unique transcripts annotated to Gammaproteobacteria in Niskin vs. the
three MS-SID datasets were expressed differentially (p<0.001). This suggests that for investigations of specific metabolic pathways within a particular COG category, that in situ fixation may provide a different picture than Niskin sampling.

While the percentages of reads associated with certain functional categories, such as genes associated with the 3-hydroxypropionate/4-hydroxybutyrate carbon fixation cycle were detected at approximately equal levels (0.27% of Niskin library and 0.25-0.39% of the MS-SID libraries) for many other transcribed genes, expression in Niskin vs. MS-SID libraries varied. Although it was not the objective of this study proof-of-concept study to closely examine specific microbial activities (ecological function) within different COG categories of expressed genes in these samples, a closer examination of specific transcripts and taxonomic affiliation of those transcripts within COG O (posttranslational modification, protein turnover, and chaperones) provides insight into the types of variation underlying observed differences in overall results. Transcripts annotated to COG category O represented 2.8% (annotated to 29 taxa) and 1.1% (annotated to 18 taxa) of total transcripts in each library for Niskin and MS-SID samples, respectively. Transcripts in the Niskin library were annotated to 10 taxonomic groups that were not also detected in the MS-SID samples; Bacillariophyceae, Chlamydiae, Chlorobia, Chloroflexi, Choanoflagellata, Crenarchaeota, Tenericutes, Thaumarchaeota, Thermotogae, and Verrucomicrobia (results presented at different taxonomic levels due to variance in annotation detail). This suggests that in situ filtration and preservation may have reduced stresses experienced by some taxa. Looking within the COG O transcripts annotated to protozoa and metazoa provides insight into observed variations between Niskin and MS-SID samples, and illustrates that sensitivity to sample handling is variable between taxa.
Among transcript types that were unique to the Niskin library, 5478, 4858, 529, and 4103 reads were detected for molecular chaperones of choanoflagellates, ciliates, Bacillariophyta, and metazoa, respectively. Also unique to the Niskin library was 62 and 1301 reads annotated to molecular chaperone HSP90 of choanoflagellates and various metazoan, respectively. Heat shock proteins such as HSP90 assist the proper folding of proteins under cell stress. Expression of transcripts for the protein ubiquitin was detected in both libraries, but at higher levels (1786 vs. 42 reads) in Niskin vs. MS-SID samples. Ubiquitin is a small protein associated with post-translational modification that is expressed in many eukaryotes and can among other functions, signal for the degradation of a damaged protein (Glickman and Ciechanover 2002). Some types of COG O transcripts were detected only in the MS-SID samples, and these were all affiliated with various metazoan (echinoderms, nematodes). These included metazoan transcripts for thioredoxin reductase (511 reads), FKBP-type peptidyl prolyl cis-trans isomerase (190 reads), transcripts annotated to disulfide bond chaperone activity (345 reads), an ATPase with chaperone activity (345 reads), and a predicted metalloendopeptidase associated with peptide breakdown (8729 reads). Transcriptional responses to sample handling also differed between taxa. Illustrating this, in comparison to protists (Figure 3), Fungi (Supplementary Figure 1) exhibit greater differences in percentage representation within many COG categories between MS-SID and Niskin samples.

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

Less is known about decompression effects on microorganisms. We observed lower expression levels (as a percentage of total annotated reads) in the Niskin sample for many COG categories for Fungi (Supplementary Figure 1), and decompression effects are one possible explanation for this. In comparison, for protists, some categories of transcripts appear to be suppressed and others, such as COG M (cell wall/membrane/envelope biogenesis), enhanced. In the case of COG M, this may indicate protists increase transcription of genes associated with cell membrane repair in response to decompression. Pressure is the only stressor known to simultaneously induce a wide range of both heat shock and cold shock proteins, likely in response to destabilization of protein quaternary and tertiary structures, and their simultaneous induction in E. coli may represent an attempt to repair the damaging effects of elevated pressure on membrane integrity, translation processes, and macromolecule stability (Bartlett, 2002). Heat shock proteins are known to be induced in piezophiles such as the deep-sea piezophilic hyperthermophile Thermococcus barophilus upon decompression (Marteinsson, et al., 1999).

While an increase in percent transcription within COG category O (posttranslational modification, protein turnover, chaperones) was not observed in the Niskin sample relative to the MS-SID samples for Fungi (Supplementary Figure 1), an increase was observed for Gammaproteobacteria (Figure 2), protozoa (Figure 3), and many other taxa (data not shown). When the number of reads assigned to the 118 unique types of transcripts detected in our samples within this COG category in the Niskin sample were compared to those detected in the three MS-SID samples, 86% of them showed evidence of differential levels of expression (p<0.001). Given that high-pressure activation of gene expression is known among piezophilic
bacteria (Bartlett, 2002), it stands to reason that transcription of many types of genes may also be affected by decompression for potentially, a wide range of taxa.

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

Free-living (dispersed in the water column), and detritus-associated protozoa are now recognized as important components of marine microbial communities (Caron, 1991). While piezotolerant protists (flagellates, ciliates, and amoebae) from the bathypelagic have been observed using microscopic analysis of cultures grown under high pressure or preserved samples (Patterson, et al., 1993, Atkins, et al., 2000, Edgcomb, et al., 2011), less is known about the effects of pressure changes on protozoa than for Bacteria and Archaea. Numbers of species of free-living and particle-associated protists declines from the photic zone into the bathypelagic, and this decline is thought to be attributed to reductions in diversity and quantities of available food, and potentially also due to pressure effects (Patterson, et al., 1993). Flagellates are known to have variable sensitivities to changes in pressure, and even species within genera can vary significantly in their responses (Turley & Carstens, 1991). This implies that significant depressurization may cause some protist taxa to lyse during recovery if cells are not preserved in situ. Supporting this notion, in studies of anoxic waters from 900m depth in the Cariaco Basin,
Venezuela, it was necessary to preserve water samples in situ for fluorescence and scanning electron microscopy in order to more accurately assess protist abundances (Edgcomb, et al., 2011). Deep ocean microorganisms are known to alter the phospholipid composition of their membranes to maintain adequate membrane fluidity under higher pressures (Yano, et al., 1998). It is conceivable that for an unknown fraction of the microbial community from mesopelagic and bathypelagic realms the higher membrane fluidity upon depressurization may make them more susceptible to lysing or leaking. Depressurization during sample recovery not only has the potential to alter gene expression from what was occurring in the sampled habitat, but to bias against the recovery of an unknown fraction of the community from any sample, due to loss of cell integrity for some taxa.

CONCLUSIONS

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

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

ACKNOWLEDGMENTS

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

TABLE AND FIGURE LEGENDS
Table 1. Summary of transcriptome sequencing, quality control, and number of annotated reads (based on the COG database) for the Niskin water sample and the three replicate MS-SID samples.

Figure 1. Microbial Sampler-Submersible Incubation Device (MS-SID). Panel A, schematic of the instrument illustrating major modular components. Each Fixation Filter Unit (lower left inset) possesses a reservoir for containing an appropriate (denser than water) chemical preservative (in our study RNAlater). During filtration, preservative loss is prevented by a poppet that seals the access hole shown. Upon cessation of filtration, the poppet settles by gravity, opening the reservoir, allowing preservative to flow onto the filter surface by convection. Panel B, deployment of SID-ISMS for collection of microbial samples from Mediterranean Sea for grazing studies.

Figure 2. Percentage of total annotated reads assigned to Gammaproteobacteria in each COG category for the Niskin sample (black) and the three replicate MS-SID samples (shades of grey).

Figure 3. Percentage of total annotated reads assigned to protist taxa in each COG category for the Niskin sample (black) and the combined MS-SID samples (grey).

Supplementary Figure 1. Percentage of total annotated reads assigned to protist taxa in each COG category for the Niskin sample (black) and the pooled MS-SID samples (grey).
Supplementary Figure 2. DEGseq analysis comparing differential expression between the three replicate MS-SID transcript libraries vs. the Niskin library. A) MA-plot B) Boxplot of read counts (log2) for each gene in the MS-SID vs. Niskin libraries. Y axis depicts the log2 fold change in gene expression (M) and X axis depicts the mean of normalized counts (A). Red data points are gene transcripts with significant differences (p<0.001) in expression levels.

REFERENCES


Danovaro R, Dell'Anno A, Pusceddu A & Fabiano M (1999) Nucleic acid concentrations (DNA, RNA) in the continental and deep-sea sediments of the eastern Mediterranean:


barophilic bacteria in response to changes in growth pressure. *Appl Environ Microbiol* 
**64**: 479-485.

Zobell CE (1941) Apparatus for collecting water samples from different depths for 
% 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
- Posttranslational 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 metabolism 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

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)
- Posttranslational 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 metabolism (F)
- Energy Production and Conversion (C)
- Replication, recombination and repair (L)
- Inorganic Ion Transport and Metabolism (P)
- Amino acid transport and metabolism (E)

Percentage of Total Annotated Reads for Each Dataset

- Niskin
- MS-SID
Table 1. Overall sequencing results for each sample collected from 2222m depth.

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

Percentage of Total Annotated Reads for Each Dataset

- Niskin
- MS-SID