

## Analysis of Microbial Gene Transcripts in Environmental Samples†

Rachel S. Poretsky,<sup>1</sup> Nasreen Bano,<sup>1</sup> Alison Buchan,<sup>2</sup> Gary LeClerc,<sup>1</sup> Jutta Kleikemper,<sup>3</sup>  
 Maria Pickering,<sup>1</sup> Whitney M. Pate,<sup>1</sup> Mary Ann Moran,<sup>1\*</sup> and James T. Hollibaugh<sup>1</sup>

*Department of Marine Sciences, University of Georgia, Athens, Georgia 30602<sup>1</sup>; Department of Molecular, Cellular, and Developmental Biology, P.O. Box 208103, Yale University, New Haven, Connecticut 06520<sup>2</sup>; and Institute of Terrestrial Ecology, Swiss Federal Institute of Technology Zurich, CH-8952 Schlieren, Switzerland<sup>3</sup>*

Received 12 August 2004/Accepted 21 January 2005

**We analyzed gene expression in marine and freshwater bacterioplankton communities by the direct retrieval and analysis of microbial transcripts. Environmental mRNA, obtained from total RNA by subtractive hybridization of rRNA, was reverse transcribed, amplified with random primers, and cloned. Approximately 400 clones were analyzed, of which ~80% were unambiguously mRNA derived. mRNAs appeared to be from diverse taxonomic groups, including both *Bacteria* (mainly  $\alpha$ - and  $\gamma$ -*Proteobacteria*) and *Archaea* (mainly *Euryarchaeota*). Many transcripts could be linked to environmentally important processes such as sulfur oxidation (*soxA*), assimilation of C1 compounds (*fdh1B*), and acquisition of nitrogen via polyamine degradation (*aphA*). Environmental transcriptomics is a means of exploring functional gene expression within natural microbial communities without bias toward known sequences, and provides a new approach for obtaining community-specific variants of key functional genes.**

The technology of environmental genomics is based on sequence analysis of fragments of environmental DNA and retrieves genes without any previous sequence information and with relatively little apparent bias (1, 18, 21). An analogous method for environmental mRNA (i.e., environmental transcriptomics) could similarly retrieve the transcriptome of a microbial assemblage without any prior information on what genes the community might be expressing. The prospect for using environmental transcriptomics to link genetic potential with biogeochemical activity of microbes has been hindered, however, by the difficulties of working with mRNAs. Prokaryotic transcripts generally lack the poly(A) tails that make isolation of most eukaryotic messages straightforward (12). Some mRNAs degrade quickly, with half-lives as short as 30 s based on studies of cultured bacteria (2). Finally, mRNA molecules are much less abundant than rRNA molecules in total RNA extracts, so the mRNA signal is often overwhelmed by background.

We have developed a protocol to analyze partial environmental transcriptomes by collecting total RNA from the environment, enriching for mRNA by subtractive hybridization of rRNA, and using randomly primed reverse transcription (RT) to produce a cDNA template population. The templates are amplified by PCR and used to generate cDNA clone libraries. Here we report results from the analysis of approximately 400 environmental gene transcripts retrieved directly from bacterioplankton communities of Sapelo Island, GA, and Mono Lake, CA.

**Protocol for library generation.** Water samples were collected from the Sapelo Island Microbial Observatory (SIMO; tidal salt marsh creek in the southeastern United States; <http://simo.marsci.uga.edu/>) and the Mono Lake Microbial Observatory (MLMO; closed-basin, hypersaline soda lake near Lake Tahoe, CA; <http://www.monolake.uga.edu/>). SIMO water samples (10 liters) were collected in October 2002 and August 2003 and screened immediately after collection to remove particles of >3.0  $\mu$ m, including most eukaryotic cells. Cells for RNA extraction were collected on a 0.2- $\mu$ m-pore-size polycarbonate membrane filter. MLMO samples (8 liters) were collected in May 2003 at depths of 5 m (surface) and 23 m (chemocline). Because the dominant phytoplankter (*Picocystis salinarum*) is of a size similar to the bacterioplankton, MLMO samples were not screened. MLMO samples were stored on ice during transport to the laboratory and then filtered onto a 0.2- $\mu$ m-pore-size membrane filter.

The process from sample collection to RNA extraction was done as rapidly as possible to limit degradation of mRNA. RNA was extracted using a RNAqueous-Midi kit (Ambion, Austin, TX) with several modifications (see the supplemental material for detailed protocol). For SIMO samples, the elapsed time between water collection and RNA extraction was less than 30 min. For MLMO samples, the elapsed time was ~2 h. Subtractive hybridization was used to selectively remove rRNA (MICROBExpress Bacterial mRNA enrichment kit; Ambion). DNase-treated mRNA preparations were amplified by RT-PCR using two of six possible random primers (see Table S1 in the supplemental material): 10-mer primers OPA04, OPA13, and OPA17 from a commercial primer stock (Operon Technologies, Inc., Alameda, CA), primer SD14 designed to target the Shine-Dalgarno region of bacterial mRNAs (5), and primers SES3-1 and SESRT-3 designed with low G+C content (MLMO only). Clone libraries of some PCR products were screened to eliminate sequences derived from contaminating rRNA using probes constructed by amplifying rRNA genes from DNA harvested from the same sample (see the supplemental material for detailed protocol). Sequences of 347 SIMO clones (40 from October 2002 and 307 from August

2003) were analyzed. The results of this analysis are presented in the supplemental material.

\* Corresponding author. Mailing address: Department of Marine Sciences, University of Georgia, Athens, GA 30602. Phone: (706) 542-6481. Fax: (706) 542-5888. E-mail: mmoran@uga.edu.

† Supplemental material for this article may be found at <http://aem.asm.org/>.

2003) and 60 MLMO clones were analyzed using the BLASTX and BLASTN tools (<http://www.ncbi.nlm.nih.gov/BLAST/>). Additionally, 282 of the August 2003 SIMO clones were automatically annotated using the Annotation Engine service provided by The Institute for Genomic Research (Rockville, MD).

**Environmental transcript libraries.** We calculate that  $2.4 \times 10^{13}$  bacterial mRNAs were present in the 10-liter SIMO water samples collected in August 2003, of which 80,000 were unique (calculated assuming a late summer population of  $1.7 \times 10^6$  bacterial cells  $\text{ml}^{-1}$  [<http://gce-lter.marsci.uga.edu>] each with 1,380 total mRNA molecules per cell [10] and 200 bacterial species represented in the community [<http://simo.marsci.uga.edu/MainWeb/pages/database.htm/>] each with 400 unique mRNAs per cell [10]). Thus the 342 SIMO clones and 60 MLMO clones analyzed here were a small fraction of the total transcript pool in each environment. Yet while these small libraries do not provide a quantitative inventory of bacterioplankton transcripts, they offer a novel glimpse of microbial activity that is unconstrained by existing sequence data and not restricted to previously characterized processes. Further, the standard cloning and sequencing methods used for these manually assembled libraries can be readily adapted to high-throughput approaches, potentially allowing the sequencing of thousands of amplicons from a single community.

Sublibraries were generated from a single sample using different primer combinations, with one primer chosen at random for the RT step and that primer used in combination with a second primer in the PCR step (see Table S2 in the supplemental material). When we compared transcript retrieval with different permutations of the random primers, the SD14 primer appeared to outcompete the others. Often, both ends of the amplicons were primed by SD14. No amplicons were generated for MLMO samples without SD14 in either the RT or PCR step, although several primer combinations without SD14 were used with success in the SIMO samples (see Table S2 in the supplemental material). The higher amplification efficiency with the SD14 primer is not surprising, as primers designed to bind to the Shine-Dalgarno region (the ribosomal binding site) have been used in differential display analyses of mRNA transcripts in both pure cultures and in soils (5). When used as a PCR primer, it ostensibly biases amplification to the 5' end of mRNA transcripts for bacteria that possess a typical *Escherichia coli*-like Shine-Dalgarno region (e.g., AGGAGG) (10). When used as an RT primer, we expected to see the SD14 primer sequence only for polycistronic operons because it would target the Shine-Dalgarno site at the beginning of the gene downstream from the one that was reverse transcribed. Because the SD14 primer sequence was often identified at both the beginning and end of sequences following RT-PCR, we concluded that SD14 does not necessarily target the Shine-Dalgarno site exclusively when used under low-specificity PCR conditions.

Although we still observed a few rRNA-generated cDNA sequences after repeated subtractive hybridizations, analysis of the clone libraries indicated that the protocol for removing rRNA worked effectively, as typically fewer than 20% of the clones were derived from 16S, 23S, and 28S rRNA combined (see Table S2 in the supplemental material). Results from colony hybridizations of the SIMO clone libraries indicated that perhaps a higher percentage of clones were rRNA gener-

ated, but this screening step reduced the number of rRNA clones sequenced. Even though they were not screened by hybridization, MLMO cDNA libraries contained few rRNA genes, indicating that the subtractive hybridization alone worked efficiently for these samples.

**Apparent taxonomic representation.** The putative taxonomic origin of the transcripts was used to assess diversity in relation to the known microbial composition of the two systems. Putative taxonomic origin was assigned based on the taxon of the most similar sequence by BLAST analysis (see Table S3 in the supplemental material). The accuracy of this assignment is negatively affected by lateral gene transfer and positively correlated with the taxonomic coverage of the database for any given gene. Given these caveats, the SIMO libraries appeared to be almost entirely bacterial derived, although similarities to known genes were sometimes low (see Table S3 in the supplemental material). Using only those sequences with E values of  $\leq e^{-10}$ , gene expression at the SIMO site was inferred for  $\alpha$ -,  $\beta$ -,  $\gamma$ -,  $\delta$ -, and  $\epsilon$ -*Proteobacteria*, *Bacteroidetes*, *Chlorobi*, *Cyanobacteria*, *Firmicutes*, *Actinobacteria*, *Spirochaetes*, *Planctomycetes*, *Euryarchaeota*, and *Crenarchaeota*. Apparent archaeal sequences represented a significant portion of the August 2003 SIMO library (Fig. 1A). Almost all of the putative archaeal transcripts were most similar to genes from *Sulfolobus tokodaii* or *Methanococcus voltae*, but identification may be skewed toward organisms for which a genome sequence is available.

A small-subunit rRNA database of SIMO bacterioplankton that was generated during a different year, but for the same season (summer) and the same size fraction (0.2 to 3.0  $\mu\text{m}$ ) (<http://simo.marsci.uga.edu/>), provided a comparison with the putative taxonomic assignment of the transcripts in the August 2003 mRNA library. The SIMO 16S rRNA libraries were dominated by sequences from  $\alpha$ - and  $\gamma$ -*Proteobacteria* (18 and 16%) (Fig. 1B). These two taxonomic groups were similarly represented in the mRNA library (16 and 19%, respectively). *Cyanobacteria* played a larger role in the 16S rRNA library (Fig. 1B) than in the mRNA library while *Chlorobi*,  $\epsilon$ -*Proteobacteria*, and *Spirochaetes* appeared to contribute to the mRNA pool but were not well represented in the 16S rRNA library. Overall, relatively similar distributions among apparent taxonomic groups existed between the two libraries. The transcripts in the August 2003 SIMO library were also compared to the genome of *Silicibacter pomeroyi*, a marine  $\alpha$ -*Proteobacteria* isolated from coastal water near the SIMO site ([http://www.marsci.uga.edu/s\\_pomeroyi/](http://www.marsci.uga.edu/s_pomeroyi/)) (6, 15). Using BLASTX, almost 10% of the clones in the SIMO transcript library matched predicted proteins in the *S. pomeroyi* genome with identities higher than other entries in GenBank, with E values between  $e^{-70}$  and  $e^{-97}$  in most cases.

At MLMO, 33% of the 60 transcripts appeared to be eukaryotic in origin; not surprising given that the spring phytoplankton bloom was under way during sample collection and eukaryotes were too small to be removed by size-selective screening. Prokaryotic MLMO transcripts matched genes from *Firmicutes*, *Cyanobacteria*, *Bacteroidetes*, *Spirochaetes*, *Actinobacteria*, *Planctomycetes*, and  $\alpha$ -,  $\beta$ -,  $\delta$ -, and  $\epsilon$ -*Proteobacteria*. Putative taxonomic affiliations of MLMO transcripts were consistent with a 16S rRNA gene library constructed in July 2000, as evidenced by the presence of  $\gamma$ -*Proteobacteria*-like se-

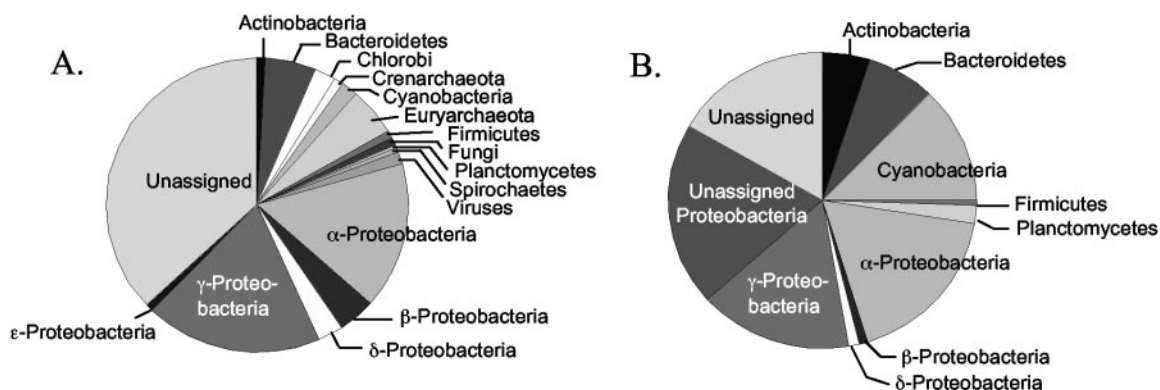


FIG. 1. Taxonomic assignment of SIMO mRNA (A) and bacterial 16S rRNA (B) sequences from the 0.2- to 3.0- $\mu\text{m}$ -size fraction of SIMO bacterioplankton. *Archaea*, viruses, and fungi were not captured by the 16S rRNA primers used in the rRNA library. No putative taxonomic assignment could be made for 37% of the mRNA clones because they had highest similarity to genes from unclassified organisms. Twenty-three percent of the mRNA clones did not match any database sequence well using a BLASTX E-value cutoff of  $e^{-10}$  and were not included in this figure. Approximately 17% of the rRNA clones could not be readily assigned to a phylum using a cutoff value of 80% similarity to described organisms, while 20% could be classified as *Proteobacteria* but not assigned a class within this phylum.

quences in surface and chemocline samples in both mRNA and 16S rRNA libraries, as well as a cyanobacterial-like mRNA and 16S rRNA sequences in the chemocline (9). Although there is uncertainty in the taxonomic assignment of mRNA sequences as discussed above, environmental transcripts appeared to be retrieved from a diversity of microorganisms at both the SIMO and MLMO sites.

**Transcript annotation.** Most of the sequences obtained were not full-length transcripts ( $\sim 200$  to 500 bp), although some amplicons were large ( $>1,000$  bp). In all cases, there was no amplification in controls that lacked the RT step. The mRNA sequences appeared to be transcribed from a range of house-keeping genes, components of transport systems, and genes for energy metabolism (Table 1). Like the taxonomic assignments, the identities of transcripts were inferred from the closest matches by BLASTX (see Table S3 in the supplemental material). These assignments are only as good as the existing database, and genes that are rare in genomes because they code for unusual or specialized traits are particularly susceptible to poor database coverage. For example, a MLMO transcript with a strong BLAST hit to an arsenite transporter

(*arsA*) from *Arabidopsis thaliana* (Table 1) predicts a function expected in Mono Lake given the high concentration of arsenite (200  $\mu\text{M}$ ) (14) but predicts an organism quite distant from any lake plankton.

In all libraries, several instances of multiple mRNAs transcribed from homologous genes were seen. Some of the repeated mRNAs, such as those having sequence similarity to *soxA* (sulfur oxidation; eight sequences) and *surE* (stationary-phase survival protein; four sequences) were found in different sublibraries (i.e., libraries constructed from the same RNA sample but using different primer pairs). In all but one case, the homologous sequences were found in eight or fewer clones, with the exception being transcripts putatively encoding acetyl-polyamine amidohydrolase (*aphA*) that accounted for 35 of 307 clones in the August 2003 sample.

Annotation of the clones from the SIMO libraries revealed that the majority ( $\sim 80\%$ ) were found only once in the library. In contrast, nearly half (42%) of the 60 MLMO sequences were homologous to another sequence in the library. Four clones from the SIMO August 2003 sample and three clones from MLMO had no significant matches using an EXPECT

TABLE 1. Selected mRNA sequences with inferred functions of ecological or geochemical relevance in cDNA libraries constructed from SIMO (top) and MLMO (bottom) samples<sup>a</sup>

Gene	Putative function	Closest match source	Accession no.	% Identity <sup>b</sup>
<i>soxA</i>	Inorganic sulfur oxidation	<i>Chlorobium tepidum</i>	NP_661911	51
<i>fdh1B</i>	Formate dehydrogenase beta-subunit (C1 metabolism)	<i>Methanococcus voltae</i>	AAK57554	58
<i>trkA</i>	Potassium uptake	<i>Bacteroides thetaiotaomicron</i>	NP_813009	63
<i>kefA</i>	Potassium efflux	<i>Pseudomonas syringae</i>	AAO58452	48
<i>psbA2</i>	Photosystem II protein	<i>Synechococcus</i> sp. strain WH 8102	NP_897076	78
<i>surE</i>	Stationary-phase survival protein	<i>Coxiella burnetii</i>	NP_820653	58
<i>proV</i>	Proline/glycine betaine/DMSP <sup>c</sup> transport system	<i>Streptomyces coelicolor</i>	AAD29279	63
<i>mexE</i>	Multidrug efflux membrane protein	<i>Pseudomonas syringae</i>	NP_792891	85
<i>chi</i>	Chitinase	<i>Bacillus thuringiensis</i>	AAM88400	27
<i>arsA</i>	Arsenite-transporting ATPase	<i>Arabidopsis thaliana</i>	NP_563640	68

<sup>a</sup> A complete list of all transcript annotations is provided in Table S3 in the supplemental material.

<sup>b</sup> Based on deduced amino acid sequences

<sup>c</sup> DMSP, dimethylsulfoniopropionate.



TABLE 2. SIMO transcript identities and assigned role categories (August 2003 sample) as determined by the TIGR Annotation Engine

Main role and subrole(s)	No.	% <sup>a</sup>
Amino acid biosynthesis	6	2.5
Aspartate	2	
Pyruvate	4	
Biosynthesis of cofactors	1	0.4
Cell envelope	1	0.4
Cellular processes	13	5.5
Chemotaxis and motility	1	
Detoxification	8	
DNA transformation	2	
Toxin production and resistance	2	
Central intermediary metabolism	44	18.5
DNA metabolism	5	2.1
Replication, recombination, repair	5	
Energy metabolism	10	4.2
Aerobic	1	
Amino acids and amines	3	
ATP proton motive force	1	
Electron transport	1	
Glycolysis/gluconeogenesis	1	
Pentose phosphate pathway	1	
Photosynthesis	1	
Tricarboxylic acid cycle	1	
Protein fates	3	1.3
Degradation of proteins, peptides, and glycopeptides	1	
Protein and peptide secretion	2	
Protein synthesis	12	5.0
Ribosomal proteins	10	
Translation factors	1	
tRNA aminoacylation	1	
Regulatory functions	10	4.2
Transport and binding proteins	9	3.8
Amino acids, peptides, amines	2	
Carbohydrates, alcohols, and organic acids	2	
Unknowns	5	
Unknown functions	2	0.8
Unclassified <sup>b</sup>	28	11.8
Conserved hypothetical proteins <sup>c</sup>	10	4.2
Hypothetical proteins	84	35.3
rRNA	44	
Total no. of clones	282	

<sup>a</sup> Calculation of percent representation in the library does not include the 44 rRNA sequences.

<sup>b</sup> Unclassified proteins have a known function but have not been assigned to a role category.

<sup>c</sup> Conserved hypothetical proteins have homologs in other organisms, but none of the homologs have known functions.

threshold of 10 in homology searches (BLASTN and BLASTX) and thus either are transcripts of novel genes or are not transcripts.

The Institute for Genomic Research Annotation Engine organized the August 2003 SIMO sequences into role categories based on assigned functions of the highest matching gene sequences, including central intermediary metabolism (18.5% of the clones), cellular processes (5.5%), and protein synthesis (5.0%) (Table 2). Transcripts that appeared to code for transport/binding proteins (3.8%) were potentially involved in amino acid, carbohydrate, and organic acid and alcohol transport and metabolism (Table 2). The largest fraction of transcripts was categorized as hypothetical (35.3%), and 12% were "unclassified" (typically of known function but not readily

placed in a role category during autoannotation). As discussed above, inferred functional assignments of transcripts are subject to effects of database coverage and lateral gene transfer.

There are significant methodological obstacles in retrieving an environmental transcriptome that may result in the unequal capture of transcripts, including choice of primer, preferential targeting of transcripts, and bias toward the longest-lived mRNAs. In assessing the issue of targeting bias, we found evidence for selective amplification of some targets by a given primer pair, such as *soxA* transcript amplification only when both OPA13 and OPA17 primers were used and *aphA* transcript amplification only if primer SD14 was used. In assessing the issue of mRNA lifetime, we examined three gene categories predicted to have longer half-lives based on studies of *E. coli* transcripts: cell envelope genes, energy metabolism genes, and transport/binding genes (3). The 307-member August 2003 SIMO environmental transcript library was not dominated by any of these functional categories, although evidence from organisms such as *Bacillus subtilis* indicates that there are both long and short half-life transcripts in almost all gene classes (7). For the MLMO transcript library, for which the time from collection to processing was ~2 h, potential biases related to mRNA half-life could not be evaluated.

**Applications of environmental transcriptomics.** A promising application of environmental transcriptomics is the retrieval of community-specific functional gene sequences with relevance for quantitative ecological studies. Functional gene discovery in natural environments is typically based on primer sets designed from a limited database that is heavily biased toward cultured organisms (17). Environmental transcript libraries can alleviate this problem by supplying site-specific functional gene sequences from active cells without the constraints of prior sequence information. For example, the eight putative *soxA* sequences in the SIMO library were similar to one another but distinct from those found in cultured bacteria (Fig. 2). Quantitative PCR analysis of DNA from an August 2004 SIMO bacterioplankton community, using a primer set designed to target only the SIMO clade *soxA* genes, indicated that they were present at concentrations of  $\sim 4.6 \times 10^6$  liter<sup>-1</sup>, or in 1 of every 370 cells (assuming  $1.7 \times 10^6$  cells ml<sup>-1</sup> and one gene copy per cell). Further, *soxA* transcripts were retrieved from four samples collected within an 11-h period in August 2004 using RT-quantitative PCR (averaging  $2.6 \times 10^3$  transcripts liter<sup>-1</sup>), suggesting that SIMO clade *soxA* genes are consistently transcribed within the bacterioplankton community. The putative chitinase transcript in the MLMO library (Table 1), which has low identity to known chitinase sequences (<27%), is also of significance because chitinase genes cannot be amplified from the Mono Lake ecosystem using existing *chi* primer sets (11). Nevertheless, the abundance of arthropod exoskeletons in the lake along with previous demonstrations of chitinase activity (11) suggest that chitin degradation is a major microbial process in this system. Environmental transcriptomics thus provides gene sequences of biogeochemical interest (Table 1) without constraints imposed by existing sequence data and with preference for those genes being actively expressed.

Environmental transcriptomics also has considerable potential for generating novel hypotheses about microbial processes. In the SIMO library, putative acetylpolyamine amidohydrolase

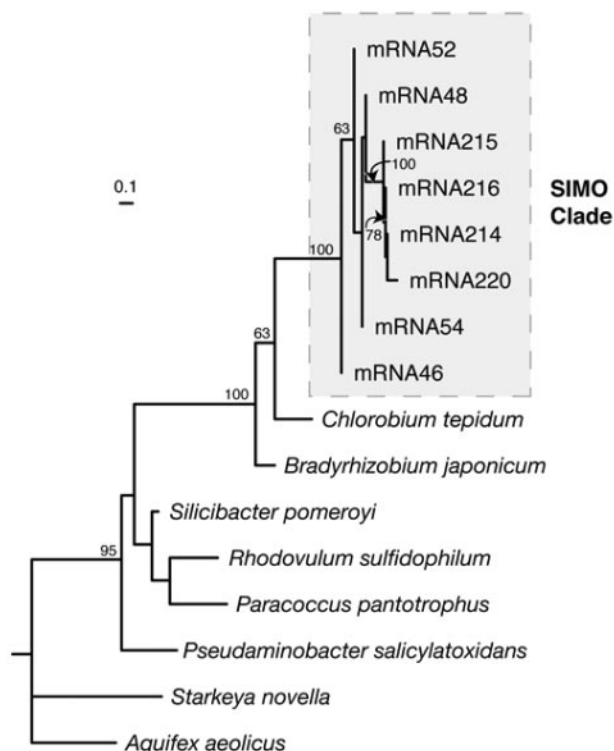


FIG. 2. Phylogenetic tree of SIMO-specific *soxA* sequences and those from representative cultured organisms constructed using the neighbor-joining method of the PHYLIP package (4). The tree is based on the deduced amino acids encoded by the *soxA* transcripts or genes (positions 8 to 247; *Chlorobium tepidum* numbering system) and is unrooted, with *soxA* from *Aquifex aeolicus* (AE000757) as the outgroup. Bootstrap values of  $\geq 50\%$  are indicated at branch nodes. The scale bar indicates Dayhoff percent accepted mutation (PAM) distance.

(*aphA*) transcripts accounted for 11% of the sequences in the August 2003 SIMO library, represented by at least seven distinct sequences in four sublibraries. The possible ecological relevance of these sequences is not immediately apparent because prokaryotic *aphA* genes are poorly characterized. However, they are suspected to encode proteins involved in the degradation of polyamines (19), a class of nitrogen-rich compounds including putrescine and spermidine that form complexes with DNA and RNA and act as important signaling compounds for cell growth (20). Evidence in support of a hypothesis that the *aphA* transcripts reflect a role for polyamines as a nitrogen source for coastal bacterioplankton includes the facts that polyamines are produced by marine algae, plants, invertebrates, and microorganisms (8, 13, 16, 20), they reach concentrations of 30 nM in coastal seawater during algal blooms (16), and they are readily assimilated by coastal and open ocean bacterioplankton communities (8). Further, the genome sequence of marine bacterium *S. pomeroyi* contains an *aphA* homolog located in an apparent operon with a polyamine transporter (*potABCD*) (15) and candidate genes for a putrescine degradation pathway (putrescine transaminase and aminobutyraldehyde dehydrogenase). The *aphA* transcripts may be a response to unusual conditions caused by sample processing (e.g., a spike in polyamine concentrations due to

eukaryotic cell breakage during filtration), but nonetheless indicate an ability of bacteria to respond rapidly to the availability of these nitrogen-rich compounds in seawater. While polyamine assimilation by marine bacteria has been considered in the past (8), the SIMO transcript library forms the foundation of a hypothesis that these compounds are a more important source of dissolved organic nitrogen for coastal bacterioplankton than is currently suspected.

Our environmental transcriptomics protocol was used successfully to survey two very different types of aquatic communities for microbial gene expression, without the constraints of targeting specific organisms, phylogenetic groups, or metabolic pathways. While the libraries analyzed here were small, this approach can be readily adapted for high-throughput processing and automated annotation and can be coupled to environmental genomics methods to assess genetic potential and patterns of activity in natural microbial assemblages.

**Nucleotide sequence accession numbers.** Newly determined sequences were deposited in GenBank under the accession numbers AY793704 to AY794012.

We thank J. R. Henriksen for help with sequence analysis, E. C. Hierling and J. Shalack for assistance with sample collection and analysis, and three anonymous reviewers for constructive comments.

This work was supported by National Science Foundation (NSF) grants MCB-0084164 to the Sapelo Island Microbial Observatory and MCB-9977886 to the Mono Lake Microbial Observatory, as well as by the Gordon and Betty Moore Foundation. R.S.P. was funded by an NSF graduate research fellowship, and M.P. and W.M.P. were funded by NSF REU fellowships. The Annotation Engine service was provided by The Institute for Genomic Research as a result of funding from the Department of Energy and the NSF.

#### REFERENCES

- Béjà, O., M. T. Suzuki, E. V. Koonin, L. Aravind, A. Hadd, L. P. Nguyen, R. Villacorta, M. Amjadi, C. Garrigues, S. B. Jovanovich, R. A. Feldman, and E. F. DeLong. 2000. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ. Microbiol.* 2:516–529.
- Belasco, J. G., and G. Brawerman. 1993. Control of messenger RNA stability. Academic Press, San Diego, Calif.
- Bernstein, J. A., A. B. Khodursky, P. H. Lin, S. Lin-Chao, and S. N. Cohen. 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. USA* 99:9697–9702.
- Felsenstein, J. 1989. PHYLIP, phylogenetic inference package (version 3.2). *Climatics* 5:164–166.
- Fleming, J. T., W. H. Yao, and G. S. Saylor. 1998. Optimization of differential display of prokaryotic mRNA: application to pure culture and soil microcosms. *Appl. Environ. Microbiol.* 64:3698–3706.
- González, J. M., J. S. Covert, W. B. Whitman, J. R. Henriksen, F. Mayer, B. Scharf, R. Schmitt, A. Buchan, J. A. Fuhrman, R. P. Kiene, and M. A. Moran. 2003. *Silicibacter pomeroyi* sp. nov. and *Roseovarius nubinhibens* sp. nov., dimethylsulfoniopropionate-demethylating bacteria from marine environments. *Int. J. Syst. Evol. Microbiol.* 53:1261–1269.
- Hambraeus, G., C. von Wachenfeldt, and L. Hederstedt. 2003. Genome-wide survey of mRNA half-lives in *Bacillus subtilis* identifies extremely stable mRNAs. *Mol. Genet. Genomics* 269:706–714.
- Höfle, M. G. 1984. Degradation of putrescine and cadaverine in seawater cultures by marine bacteria. *Appl. Environ. Microbiol.* 47:843–849.
- Humayoun, S. B., N. Bano, and J. T. Hollibaugh. 2003. Depth distribution of microbial diversity in Mono Lake, a meromictic soda lake in California. *Appl. Environ. Microbiol.* 69:1030–1042.
- Ingraham, J. L., O. Maaløe, and F. C. Neidhardt. 1983. Growth of the bacterial cell. Sinauer Associates, Sunderland, Mass.
- LeCleir, G. R., A. Buchan, and J. T. Hollibaugh. 2004. Chitinase gene sequences from diverse aquatic habitats reveal complex patterns of diversity. *Appl. Environ. Microbiol.* 70:6977–6983.
- Liang, P., and A. B. Pardee. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257:967–971.
- Lu, Y. H., and D.-F. Hwang. 2002. Polyamine profile in the paralytic shellfish poison-producing alga *Alexandrium minutum*. *J. Plankton Res.* 24:275–279.

14. Maest, A. S., S. P. Pasilis, L. G. Miller, and D. K. Nordstrom. 1992. Redox geochemistry of arsenic and iron in Mono Lake, California, p. 507–511. *In* Y. K. Kharaka and A. S. Maest (ed.), *Water-rock interaction*. A. A. Balkema, Leiden, The Netherlands.
15. Moran, M. A., A. Buchan, J. M. González, J. F. Heidelberg, W. B. Whitman, R. P. Kiene, J. R. Henriksen, G. M. King, R. Belas, C. Fuqua, L. Brinkac, M. Lewis, S. Johri, B. Weaver, G. Pai, J. A. Eisen, E. Rahe, W. M. Sheldon, W. Ye, T. R. Miller, J. Carlton, D. A. Rasko, I. T. Paulsen, Q. Ren, S. C. Daugherty, R. T. Deboy, R. J. Dodson, A. S. Durkin, R. Madupu, W. C. Nelson, S. A. Sullivan, M. J. Rosovitz, D. H. Haft, J. Selengut, and N. Ward. 2004. Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* **432**:910–913.
16. Nishibori, N., A. Yuasa, M. Sakai, S. Fujihara, and S. Nishio. 2001. Free polyamine concentrations in coastal seawater during phytoplankton bloom. *Fish. Sci.* **67**:79–83.
17. Rodríguez-Valera, F. 2004. Environmental genomics, the big picture? *FEMS Microbiol. Lett.* **231**:153–158.
18. Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L. Tiong, M. Gilman, M. S. Osburne, J. Clardy, J. Handelsman, and R. M. Goodman. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* **66**:2541–2547.
19. Sakurada, K., T. Ohta, K. Fujishiro, M. Hasegawa, and K. Aisaka. 1996. Acetylpolyamine amidohydrolase from *Mycoplana ramosa*: gene cloning and characterization of the metal-substituted enzyme. *J. Bacteriol.* **178**:5781–5786.
20. Tabor, C. W., and H. Tabor. 1984. Polyamines. *Annu. Rev. Biochem.* **53**:749–790.
21. Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar, and J. F. Banfield. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**:37–43.