DNA phosphorothioation is widespread and quantized in bacterial genomes

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
DNA phosphorothioation is widespread and quantized in bacterial genomes

Lianrong Wang\textsuperscript{a,b,c}, Shi Chen\textsuperscript{b,c,1}, Kevin L. Vergin\textsuperscript{d}, Stephen J. Giovannoni\textsuperscript{d}, Simon W. Chan\textsuperscript{a}, Michael S. DeMott\textsuperscript{a}, Koli Taghizadeh\textsuperscript{h}, Otto X. Cordero\textsuperscript{b}, Michael Cutler\textsuperscript{b}, Sonia Timberlake\textsuperscript{a}, Eric J. Alm\textsuperscript{a,b}, Martin F. Polz\textsuperscript{f}, Jarone Pinhasi\textsuperscript{i}, Zixin Deng\textsuperscript{c,d} and Peter C. Dedon\textsuperscript{a,c,1}

\textsuperscript{a}Department of Biological Engineering, \textsuperscript{b}Center for Environmental Health Sciences, and \textsuperscript{c}Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; \textsuperscript{d}Laboratory of Microbial Metabolism and School of Life Science and Biotechnology, Shanghai Jiao tong University, Shanghai 200030, China; \textsuperscript{e}College of Pharmacy, Wuhan University, Wuhan 430071, China; \textsuperscript{f}Department of Microbiology, Oregon State University, Corvallis, OR 97331; and \textsuperscript{g}Marine Microbiology, School of Natural Sciences, Linnaeus University, SE-39182 Kalmar, Sweden

Edited\textsuperscript{a} by Gerald N. Wogan, Massachusetts Institute of Technology, Cambridge, MA, and approved January 7, 2011 (received for review November 17, 2010)

Phosphorothioate (PT) modification of DNA, with sulfur replacing a nonbridging phosphate oxygen, was recently discovered as a product of the \textit{dnd} genes found in bacteria and archaea. Given our limited understanding of the biological function of PT modifications, including sequence context, genomic frequencies, and relationships to the diversity of \textit{dnd} gene clusters, we undertook a quantitative study of PT modifications in prokaryotic genomes using a liquid chromatography-coupled tandem quadrupole mass spectrometry approach. The results revealed a diversity of unique PT sequence contexts and three discrete genomic frequencies in a wide range of bacteria. Metagenomic analyses of PT modifications revealed unique ecological distributions, and a phylogenetic comparison of \textit{dnd} genes and PT sequence contexts strongly supports the horizontal transfer of \textit{dnd} genes. These results are consistent with the involvement of PT modifications in a type of restriction-modification system with wide distribution in prokaryotes.

\textbf{DNA modification \mid bioanalytical chemistry \mid sulfur}

**Results and Discussion**

**Development of a Sensitive Method to Quantify PT Modifications in Bacterial Genomes.** We approached the problem of defining the biological function of PT modifications by quantifying them and defining their sequence context. Specifically, we developed a highly quantitative electrospray ionization LC-MS/MS technique that identifies the two-nucleotide sequence context of PT modification. By using synthetic dinucleotides containing PT in the \textit{R}\textsubscript{P} configuration (the only stereochemistry observed to date), the HPLC retention times, collision-induced dissociation molecular transitions, and limits of detection were optimized for all 16 possible dinucleotide sequence contexts for PT (Fig. 1 and Table S1). We also optimized conditions for nucleic PT, which is inhibited by PT in the \textit{R}\textsubscript{P} configuration, and alkaline phosphatase hydrolysis of genomic DNA to PT-containing dinucleotides and canonical nucleosides (Fig. S1), with canonical nucleosides eluting well before the PT-modified dinucleotides. Quantification was achieved by using the \textit{S}\textsubscript{p} stereoisomer of d(CGpT) as an internal standard with multiple reaction monitoring mode of the mass spectrometer (Fig. 1). We were thus able to detect PT modifications at levels as low as 1 per 10\textsuperscript{6} \textit{n} for d(CTpT) to 2 per 10\textsuperscript{8} \textit{n} for d(CpT) in 20 \textmu g of genomic DNA (Table S1). This quantitative bioanalytical approach is a rigorous means to screen prokaryotic genomes for the sequence context and quantity of PT modifications, with immediate implications for the understanding of biological function.

**Widescap Distribution and Sequence Diversity Contexts for PT Modifications in Bacteria and Archaea.** The LC-MS/MS method was first applied to define the PT sequence contexts and quantify PT modifications in bacteria known to harbor \textit{dnd} gene clusters. These taxonomically unrelated bacteria, including \textit{Salmonella enterica} serovar Cerro 87, \textit{Escherichia coli} B7A, \textit{P. fluorescens} PS-1, \textit{S. lividans} 1326, \textit{Geobacter uraniumreducens} Rf4, \textit{Halobacterium salinarum} NC6a, \textit{B. subtilis} C600, \textit{B. megaterium} subsp. \textit{B. thermosphactum} and \textit{S. mutans} ATCC700354, represent genera of variable \textit{dnd} gene cluster sequences, but they are all able to resist nuclease degradation (1). However, we recently discovered that the \textit{dnd} gene product incorporates sulfur into the DNA backbone as a PT in a sequence- and stereo-specific manner (2). Beginning with the original observation in \textit{Streptomyces lividans} 1326 that the five-gene \textit{dnd} cluster (\textit{dndA–E}) caused DNA degradation during electrophoresis (3), the presence of \textit{dnd} genes has been established in dozens of different bacteria and archaea (4). An emerging picture of \textit{Dnd} protein function reveals that \textit{DndA} acts as a cysteine desulfurase and assembles \textit{DndC} as a 4Fe-4S cluster protein (5). \textit{DndC} possesses ATP pyrophosphatase activity and is predicted to have PAPS reductase activity, whereas \textit{DndB} has homology to a group of transcriptional regulators (4, 6). A \textit{DndD} homologue in \textit{Pseudomonas fluorescens} PS-1, \textit{S. lividans} 1326, \textit{Geobacter uraniumreducens} Rf4, \textit{Halobacterium salinarum} NC6a, \textit{B. subtilis} C600, \textit{B. megaterium} subsp. \textit{B. thermosphactum} and \textit{S. mutans} ATCC700354, represent genera of variable origins and diverse habitats, such as soil-dwelling and marine microbes, aerobic and anaerobic microbes, nonpathogenic saprophytes, and human pathogens. One immediate discovery was the presence of a PT sequence context, d(GpT), in bacteria


The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

To whom correspondence may be addressed. E-mail: shichen@whu.edu.cn or pcdedon@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1017261108/-/DCSupplemental.
originally observed to possess d(GPSA) (Table 1) (2). In the cases of S. enterica serovar Cerro 87 and E. coli B7A, d(GPSA) and d(GPSA) occurred at the same level of approximately 4 PT per 10^6 nt (Table 1). However, this pairing of PT sequence contexts was not universal, with marine bacteria B. marisrubri RED65 and H. chejuensis KCTC2396 possessing d(GPSA) accompanied by only barely detectable d(GPSG) and S. pealeana ATCC700345 possessing a ratio of d(GPSA) to d(GPSG) of approximately 2:1 (Table 1). Apart from the combination of d(GPSG) and d(GPSG), both d(GPSG) and d(GPSG) were simultaneously present in G. uraniumreducens Rf4 and S. lividans 1326, but at levels that differed by two orders of magnitude (Table 1). Distinct from the others, P. fluorescens P0-1 possesses only a single PT context, d(GPSG), at a level close to that in S. lividans 1326 and G. uraniumreducens Rf4. These results in bacteria known to possess dnd genes revealed the potential for a wide range of sequence contexts for PT modifications, as might be expected for a restriction-modification system.

Further insights into PT function were gained from surveys of bacteria not previously known to possess dnd genes. To this end, we interrogated 63 Vibrio strains derived from an approximate 1,000-strain library of ecologically differentiated coastal Vibrio-naceae with defined seasonal and habitat preferences (8). Seven isolates were found to possess PT modifications, with levels of four to six per 10^6 nt in d(GPSA)/d(GPSG) and d(GPSG) in three strains (1F267, ZS139, 1F230). Strikingly, four other isolates possessed 10-fold higher levels of PT (2 – 3 per 10^6 nt) in a new sequence context, d(CPSG) (Table 2). Partial genome sequence

### Table 1. PT modifications of DNA in bacteria

<table>
<thead>
<tr>
<th>Bacterium</th>
<th>d(GPSA)</th>
<th>d(GPSG)</th>
<th>d(GPSA)</th>
<th>d(GPSA)</th>
<th>d(GPSG)</th>
<th>d(GPSG)</th>
<th>Total PT</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli B7A</td>
<td>370 ± 11</td>
<td>398 ± 17</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>768 ± 27</td>
</tr>
<tr>
<td>S. enterica 87</td>
<td>362 ± 9</td>
<td>370 ± 11</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>732 ± 20</td>
</tr>
<tr>
<td>DH108 (pJTU1980)</td>
<td>529 ± 48</td>
<td>543 ± 62</td>
<td>2 ± 0</td>
<td>3 ± 0</td>
<td>2 ± 0</td>
<td>1,078 ± 109</td>
<td></td>
</tr>
<tr>
<td>DH108 (pJTU1238)</td>
<td>717 ± 52</td>
<td>774 ± 53</td>
<td>4 ± 0</td>
<td>6 ± 1</td>
<td>5 ± 0</td>
<td>1,505 ± 103</td>
<td></td>
</tr>
<tr>
<td>P. fluorescens P0-1</td>
<td>—</td>
<td>—</td>
<td>451 ± 9</td>
<td>—</td>
<td>—</td>
<td>451 ± 9</td>
<td></td>
</tr>
<tr>
<td>S. lividans 1326</td>
<td>2 ± 0</td>
<td>471 ± 39</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>474 ± 39</td>
<td></td>
</tr>
<tr>
<td>G. uraniumreducens Rf4</td>
<td>—</td>
<td>3 ± 0</td>
<td>517 ± 14</td>
<td>—</td>
<td>—</td>
<td>520 ± 13</td>
<td></td>
</tr>
<tr>
<td>B. marisrubri RED65</td>
<td>438 ± 23</td>
<td>3 ± 0</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>440 ± 23</td>
<td></td>
</tr>
<tr>
<td>S. pealeana ATCC700345</td>
<td>316 ± 9</td>
<td>172 ± 2</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>489 ± 11</td>
<td></td>
</tr>
<tr>
<td>H. chejuensis KCTC2396</td>
<td>286 ± 9</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>286 ± 9</td>
<td></td>
</tr>
</tbody>
</table>

Values represent mean ± SD for three analyses of 20 μg of bacterial DNA; dash indicates that the dinucleotide was not detected. *Signal for the dinucleotide was detected but below the limit of quantification.

Fig. 1. Analysis of PT-linked dinucleotides by LC-MS/MS in multiple reaction monitoring mode. All of the 16 possible PT-linked dinucleotides in Rp configuration were resolved by reversed-phase HPLC followed by MS/MS detection by using the ion transitions labeled under each dinucleotide. Bold arrow indicates d(GPSA) S internal standard.
information for these isolates revealed dnd gene homologues in three of the strains (1F267, ZS139, and 1F230; Materials and Methods). These results both expand the repertoire of sequence contexts and broaden the range of PT levels in bacterial genomes to cover three orders of magnitude.

**PT Modifications Are Quantized in Three Discrete Frequencies.** Analysis of the quantitative data revealed that the levels of PT modifications were quantized into three distinct levels: two to three per 10^6 nt, three to eight per 10^6 nt, and one to six per 10^6 nt (Tables 1 and 2). Along with defined sequence contexts, the first two frequency ranges are consistent with a restriction-modification system (9). The highest frequency of two to three PT modifications per 10^6 nt (one PT modification in 333–500 nt) was observed in Vibrio species as d(CpGpC) (Table 2), which is consistent with a 4-nt consensus sequence, such as GGCpGpC or CpGpCGG, with a statistical frequency of once every 256 nt. Analysis of available partial genome sequence data for these strains (Table S2) reveals that the CCpG motif occurs every approximately 905 ± 130 nt and the GGCpG motif every approximately 518 ± 54 nt (averages for four species), whereas the CC dinucleotide motif occurs every 24 nt. With allowance for the A/T richness of the Vibrio genomes (Table S2), the GGCpG and CCpG motifs are thus reasonable candidates for palindromic PT consensus sequences.

The other PT frequency consistent with a restriction-modification function is three to eight PT per 10^6 nt (Tables 1 and 2), which is equivalent to one PT in 1,250 to 3,333 nt or a 5–6-nt consensus sequence (~1 modification every 1,024–4,096 nt). This frequency was observed for the d(GpSpG) motif in S. lividans (Table 1) (2), in which a loose consensus sequence of 5′-cCGC-3′ (GGCC strictly conserved) was determined based on cloning of dnd phenotype break sites (10, 11). Analysis of the limited genome sequence data available for S. lividans 1326 reveals that the strictly conserved 4 nt GpSpG GCC core sequence occurs at too high a frequency (one site per ~110 nt; Table S3) to serve as the consensus sequence for a PT-based restriction-modification system. Such is also the case for P. fluorescens P01-0 and G. uramunareducens RF4, in which GGC occurs, on average, every 114 nt and 177 nt, respectively (Table S3). Therefore, a 6-nt consensus, such as CGGpCpGGCC (GGCC strictly conserved) occurs at a frequency of 6 to 11 sites per 10^6 nt in S. lividans 1326, P. fluorescens P01-0, and G. uramunareducens RF4 (Table S3), is more consistent with the observed levels of PT modifications (three to eight per 10^6 nt).

A similar situation holds for E. coli B7A and S. enterica 87, in which we observed a 1:1 ratio of d(GpSpA) and d(GpSpG) with each PT-containing site occurring every 2,500 to 2,800 nt on average (Table 1). This frequency is too low for a 4-nt consensus sequence such as GpSpAAGpSpTTCC, which occurs in the related E. coli DH10B once in every 258 nt (Table S4), but it is consistent with a 1- to 2-nt extension of this core sequence. The situation with B. marisrubri RED65 and H. chejuensis KCTC2396, in which d(GpSpA) was observed as the predominant dinucleotide context at frequencies consistent with a 5- to 6-nt consensus (Table 1), suggests a palindromic core sequence such as GpSpATC.

One consequence of the high sensitivity of the LC-MS/MS method is that we were able to detect PT modifications at levels well below those expected for a restriction-modification system. As shown in Tables 1 and 2, we observed two to three PT modifications per 10^6 nt in the d(GpSpG) motif in S. lividans 1326 and G. uramunareducens RF4 for which d(GpSpG) was the high frequency modification site, whereas d(CpGpA), d(TpSpA), and d(ApSpC) occurred at low levels (two to six per 10^6 nt) in E. coli expressing the S. enterica dnd genes (Table 1) and d(ApSpC) and d(TpSpC) were minor sites observed in Vibrio isolates in Table 2.

There are two explanations for the low PT levels: a function for PT other than restriction-modification, as is the case for DNA methylation in many prokaryotes (9, 12); and biochemical non-specificity for the Dnd protein responsible for target selection. The latter would be similar to a restriction enzyme cleaving at a thermodynamically or kinetically disfavored site (i.e., secondary cleavage sites), or cleaving with altered sequence specificity as a result of salt- or pH-induced alterations in protein–DNA interactions (i.e., “star activity”) (9). As one test of the latter hypothesis, we compared the levels of d(GpSpT) and d(GpSpA) in E. coli DH10B harboring low and high copy number plasmids containing the S. enterica 87 dnd gene cluster, pJTU1980 and pJTU1238, respectively. The high copy number vector produced a 16-fold increase in dndC transcription (Table S5). As shown in Table 1, the increased expression of dnd genes caused PT modifications in d(GpSpT) and d(GpSpA) to increase 1.5- and 2-fold, respectively, in comparison with S. enterica 87 (Table 1). This is consistent with relaxation of a strict modification consensus sequence or an increase in the proportion of a consensus sequence population that becomes modified with a PT. Support for the former model comes from the appearance of three low-frequency PT modifications in d(GpSpA), d(TpSpA), and d(ApSpC) in direct proportion with dnd gene expression (Table 1). The results support the hypothesis that low-frequency PT modifications result from a degree of relaxed DNA target recognition by Dnd proteins.

**Phylogenetic Analysis of dnd Genes and PT Sequence Contexts Is Consistent with Horizontal Gene Transfer.** In addition to quantifying PT levels, support for the involvement of PT modifications in a restriction-modification system comes from an apparent association of PT sequence contexts with phylogenetic relationships drawn from Dnd protein sequences but not species phylogeny. There is strong evidence for the distribution of classical methylation-based restriction-modification systems (13) and for the distribution of dnd gene clusters (14) by horizontal gene transfer and mobile genetic elements, as opposed to the vertical gene transfer associated with species phylogeny. As shown in Fig. 2, there is a strong correlation between the Dnd protein sequence phylogeny and the distribution of PT sequence modifications, with the exception of outlier Candidatus Pelagibacter ubique. The fact that the Dnd phylogenies do not follow their corresponding species tree suggests that PT sequence context is dependent on Dnd protein sequence and not on the phylogenetic descent of the strains. This is most clearly seen in the three Vibrio isolates (ZS139, 1F230, and 1F267), which are phylogenetically incoherent in all four Dnd proteins. The phylogenetic differentiation of the Vibrio isolates also suggests horizontal gene transfer of the whole dnd cluster. The fact that this split is matched by a corresponding switch in PT sequence context is a strong indication of the dependency between dnd genes and PT contexts.

**Insights into Ecological Distributions of PT Modifications and dnd Genes by Metagenomic Analysis of Ocean Genomes.** The widespread distribution of dnd genes and PT modifications is further illustrated by a metagenomic analysis of ocean bacteria (15). To
initiate these studies, we identified a dnd gene cluster homologue in *C. Pelagibacter ubique* strain HTCC1002, one of the smallest known free-living bacteria and the first cultured member of the alphaproteobacterial SAR11 clade. SAR11 is a ubiquitous group of marine bacteria that can account for as much as 35% of bacterioplankton populations in the ocean surface (16). The SAR11 dnd gene cluster was located in a hypervariable region of the genome and was not found in closely related strain HTCC1062, which is consistent with our observation of PT only in strain HTCC1002 as d(GPSA) and d(GPSG) R modifications.

Having established a PT modification in strain HTCC1002, we used tBLASTn with the HTCC1002 dndBCDE genes as queries to detect the distribution of dnd genes in oceanic metagenomes. We observed significant hits (E<10^{-20}) in several oceanic metagenomes (Table S6): 136 reads contained dndB, 123 reads contained dndC, 80 reads contained dndD, and nine reads contained dndE. In particular, one of the most significant hits was detected in the metagenome of Sargasso Sea, a low-nutrient, low-productivity, subtropical ocean gyre. DNA samples collected from depths to 200 m in the Sargasso Sea revealed natural DNA PT contexts of d(CPSG) and d(ACGT) in all samples, with d(GPSG) and d(GPSG) occurring mainly at lower depths (Table 3). A similar analysis of DNA samples collected from 5 m, 20 m, and 40 m in the highly productive, temperate Oregon coastal waters revealed the presence of four major PT contexts: d(CPSG), d(GPSA), d(GPSG), and d(GPSG) occurring mainly at lower depths (Table 3). A dramatic transition in the composition of microbial communities as a function of depth have been documented in previous studies (17). These results suggest that d(GPSG) is associated with microbial taxa that occupy the dark mesopelagic ocean region beneath the euphotic zone, which may have implications for horizontal gene transfer among microbes occupying the various ocean communities.

In summary, these results provide insights into the function of PT modifications in bacterial genomes. Our LC-MS/MS approach to studying the only defined chemical modification of the DNA backbone, with potential application to the study of putative arsenic-modified microbial nucleic acids (18), provides a rich source of information that complements genetic and molecular approaches to defining biological function. The data reveal quantized levels of PT, which suggest involvement of PT in a new restriction-modification system. This is consistent with our recent observation that a dptF-H cluster adjacent to dnd genes in *S. enterica* serovar Cerro 87 restricts the uptake of plasmids without PT (19) and with studies showing inhibition of methylation-based restriction enzymes by PT modifications (20, 21). The observations of widespread ecological distribution and strong phylogenetic relationships of dnd genes and

### Table 3. PT detection in seawater at varying depths

<table>
<thead>
<tr>
<th>DNA sample location/depth</th>
<th>d(CPSG)</th>
<th>d(GPSA)</th>
<th>d(GPSG)</th>
<th>d(GPSG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oregon coast</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 m</td>
<td>+</td>
<td>+</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>20 m</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>—</td>
</tr>
<tr>
<td>40 m</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Sargasso Sea</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 m</td>
<td>+</td>
<td>+</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>80 m</td>
<td>+</td>
<td>+</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>200 m</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Plus sign denotes detectable levels of PT in 2 μg of DNA; dash indicates that the PT-containing dinucleotide was not detectable. Oregon samples were collected from station SH50 (44° 15′ N, 124° 10′ W) on April 2, 2007; Sargasso Sea samples were collected at the BATS station on August 7, 2001 (15).
PT modifications demonstrate the importance of PT modifications in prokaryotic physiology.

**Materials and Methods**

**Materials, Bacterial Strains, and Plasmids.** EnantiomERICally pure, PT-containing dinucleotides in E. coli or S. enterica were isolated from a commercial egg-cluster in DH10B(pJTU1238) and DH10B(pJTU1980), primers were used to amplify SAR11 strains HTCC1002 and HTCC1062 were grown at 16 °C in 30 mM sodium acetate, pH 5.3, 0.5 mM ZnCl₂. To this end, we provide complete release of PT-containing dinucleotides.

**Controlled Enzymatic Digestion of PT Modified DNA.** The first step in the development of the LC-MS/MS technique to quantify PT modifications involved nuclease P1-mediated hydrolysis of PT-containing DNA to a limit digest of nucleosides and PT-bridged dinucleotides, as observed in earlier studies (2). Optimal parameters and MS parameters for nucleases for resolving and quantifying the PT-containing dinucleotides. To this end, we defined the HPLC retention times for the set of all possible 16 PT-linked dinucleotides by using a Thermo HyperSil GOLD aQ column (150 × 2.1 mm, 3 μm) with elution conducted at 35 °C and a flow rate of 0.3 mL/min, with a gradient of 97% buffer A (0.1% acetic acid in water) and 3% buffer B (0.1% acetic acid in acetonitrile) for 5 min, followed by 3% to 6% buffer B over a period of 30 min and 6% to 98% buffer B over a period of 1 min. Retention times are noted with other parameters later, with canonical nucleosides eluting as follows: dC, 2.2 min; dA, 3.8 min; dG, 3.9 min; and dT, 5.0 min; the d(PP)₅ assay was optimized to yield the most contiguous (i.e., highest N50) assembly, independently for each strain. N50 values ranged from 3 to 200 kb, with an average strain reaching an N50 of 40 kb and N80 of 20 kb.

**Analysis of Bacterial Genomes.** Bacterial genomes were analyzed by searching for PT motifs by using the PTdroG genome browser with sequences obtained from the National Center for Biotechnology Information (NCBI) Genome database. Complete or partial genome sequences were uploaded into the browser and the number of sequence motifs determined for both genomic DNA strands (Table S2). Dnd protein sequences for 12 strains analyzed here were retrieved from the NCBI Protein database and aligned by using Muscle (35) and phylogenetic trees were computed with the program phyML (36), with phylogenies displayed using iTol (37).

**Acknowledgments.** The authors thank Dr. Evgenya Shelobolina for providing G. uranireducens Rf4, Dr. Jian-Shen Zhao for S. pealeana ATCC700343, Dr. Hong Kum Lee for H. chejuensis KCTC2396, Dr. Eduardo Robleti for P. fluorescens PMö-1, Dr. Toshiyuki Murase for S. enterica serovar Cerro 87, Joshua Kitner for culturing the SAR11 cells, Anthony Bertagnolli and the crew of the R.V. Elakha for collecting and processing Oregon Coast DNA samples, Rachel Parsons and the crew of the R.V. Weatherbird II for collecting and processing BATS samples, and Dr. H. James Tripp for assistance with the hyperspectral region analysis. Chromatography and mass spectrometry were performed in the Bioanalytical Facilities Core of the Massachusetts Institute of Technology Center for Environmental Health Sciences, supported by National Institute of Environmental Health Sciences Grant 3S002109. This work was supported by National Science Foundation Grant CHE-1019990.


