Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses

Asuncion Martinez, Gene W. Tyson, and Edward F. DeLong.

Department of Civil and Environmental Engineering, Division of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA

Corresponding author: Edward F. DeLong, Dept. Civil and Environmental Engineering, MIT48-427, Massachusetts Institute of Technology, 15 Vassar St, Cambridge, MA. Phone: 617-253-5271; Fax: 617-253-2679. E-mail: delong@mit.edu.

Running title: Widespread phosphonate utilization in marine bacteria

SUMMARY

Phosphonates (Pn), compounds with a direct C-P bond instead of the more common C-O-P ester bond, constitute a significant fraction of marine dissolved organic phosphorus and recent evidence suggests that they may be an alternative source of P for marine microorganisms. To further characterize the microorganisms and pathways involved in Pn utilization, we screened bacterioplankton genomic libraries for their ability to complement an *Escherichia coli* strain unable to use Pns as a P source. Using this approach we identified a phosphonatase pathway as well as a novel pair of genes that allowed utilization of 2-aminoethylphosphonate (2-AEPn) as the sole P source. These pathways are present in diverse bacteria common in marine plankton including representatives of Proteobacteria, Planctomycetes and Cyanobacteria. Analysis of metagenomic databases for Pn utilization genes revealed that they are widespread and abundant among marine bacteria, suggesting that Pn metabolism is likely to play an important role in P-depleted surface waters, as well as in the more P-rich deep water column.

INTRODUCTION

Phosphorus (P) is an essential element for all living organisms. In its most oxidized form, valence +5, P is found as phosphate esters in numerous metabolic pathways and as a component of essential biomolecules such as nucleic acids, ATP and phospholipids. Perhaps for that reason, the majority of studies on P acquisition by microorganisms including those in marine environments (Dyhrman et al., 2007) have focused on understanding how microbes acquire inorganic phosphate (Pi). However, dissolved inorganic phosphate (DIP) can be a limiting nutrient in ocean surface waters (Rivkin and Anderson, 1997; Wu et al., 2000; Bjorkman and Karl, 2003), so there is considerable interest in understanding the mechanisms, extent and variability of dissolved organic phosphorus (DOP) utilization by marine microorganisms. Although the exact chemical nature of DOP is poorly understood, some insights have been gained by nuclear magnetic resonance studies of high molecular weight

(HMW) DOP concentrated by tangential flow ultrafiltration (Clark et al., 1999; Kolowith et al., 2001). These studies have revealed that across oceans and throughout the water column, about one third of the marine HMW DOP consists of phosphonates (Pn), reduced P compounds (valence +3) that contain a C-P bond instead of the more common C-O-P bond found in phosphate esters. Interestingly, phosphate ester and Pn content decreased with depth in HMW DOM relative to C, indicating that both forms of P are utilized by marine microorganisms (Clark et al., 1999). Benitez-Nelson and colleagues (Benitez-Nelson et al., 2004) analyzed Pn concentration in sediment traps in the anoxic Cariaco Basin and found that the percentage of Pns decreased from 18 to 3% relative to phosphate esters providing evidence for Pn remineralization in sinking particles. These studies suggest that Pn may be a significant source of P for marine organisms.

Known Pns include biogenic and xenobiotic compounds ((Horiguchi, 1984; Ternan et al., 1998) and references therein). 2-Aminoethylphosphonate (2-AEPn) or ciliatine, was first identified in rumen protozoa, and has since been found in phosphonolipids of many marine invertebrates, where it replaces its structural analog ethanolamine phosphate. 2-AEPn is also found in phosphonoglycolipids and it is believed to be the most abundant Pn in the oceans (Horiguchi, 1984). Other biogenic Pns include less abundant compounds of commercial importance such as the antibiotic fosfomycin or the herbicide bialaphos. In addition, synthetic Pns are currently used for various commercial applications because of the strength of their C-P bond that makes them resistant to chemical and enzymatic hydrolysis. Examples of these include the antiviral compound phosphonoacetate, the herbicide glyphosate, and numerous detergent additives. More than 20 x 10^3 tons of organophosphonates are released into the environment each year (Egli, 1988).

The ability of some microorganisms to utilize Pn as a P source has been recognized for many years, and genetic and biochemical analyses have provided detailed information on their Pn utilization pathways (reviewed in (Ternan et al., 1998; Quinn et al., 2007; White and Metcalf, 2007). Known C-P hydrolases include the C-P lyase, a multienzyme complex that

can release phosphate from a variety of alkyl- and aryl-Pns such as methylphosphonate (Mpn) and 2-AEPn, as well as compound specific pathways, like the phosphonatase pathway that acts exclusively on 2-AEPn. The latter is encoded by two genes, 2-AEPn:pyruvate transaminase (*phnW*) and phosphonoacetaldehyde hydrolase or phosphonatase (*phnX*). Phosphonatase catalyzes the hydrolytic cleavage of the C-P bond and its mechanism of action and crystal structure have been determined (Ternan and Quinn, 1998; Morais et al., 2000; Morais et al., 2004). In contrast, the C-P lyase is a complex membrane-bound system encoded by 14 genes (Metcalf and Wanner, 1991, 1993a, b; Yakovleva et al., 1998), and its activity has not yet been reconstructed *in vitro*. Based on genetic analyses, it has been proposed that the first three genes in the operon, *phnCDE*, encode a Pn-specific ABC transporter, while *phnG-phnM* are required for catalytic activity measured by the release of methane from Mpn. The role of the remaining genes, *phnF, phnN, phnO*, and *phnQ* is unclear although they are required for growth on Mpn. The C-P lyase operon appears in distantly related bacteria and phylogenetic analysis suggests that it has been subject of extensive lateral gene transfer (Huang et al., 2005).

Little is currently known about Pn utilization in marine bacteria. Dhyrman and colleagues showed that a complete C-P lyase operon is present in the marine filamentous cyanobacterium *Trichodesmium erythraeum,* and that it is expressed in phosphorus-depleted cultures, and *in situ* in the Sargasso Sea, strongly suggesting that members of this genus can use Pns as an alternative source of P (Dyhrman et al., 2006). Genes similar to a C-P lyase gene (*phnJ*), the *phnX* gene and the *phnA* gene encoding phosphonoacetate hydrolase have also been found in Sargasso Sea metagenomic libraries (Quinn et al., 2007). More recently, it has been shown that the mixed microbial communities at Station ALOHA in the North Pacific Subtropical Gyre release methane upon Mpn addition, indicating that microbes in these communities might posses a C-P lyase pathway (Karl et al., 2008). A high incidence of C-P lyase genes in the Sargasso Sea samples (Venter et al., 2004) was also reported in that study. Sequences homologous to another Pn utilization gene, *phnA*, have been recently identified in

metagenomic and metatranscriptomic analyses of coastal waters (Gilbert et al., 2008). Finally, peptides corresponding to one of the putative components of the Pn ABC-transporter of SAR11 isolates, PhnD, were among the most abundant peptides in a metaproteomic analysis of the Sargasso Sea (Sowell et al., 2009). Taken together, these results suggest that Pn may be an important P source for marine microbes.

To better characterize Pn utilization pathways in marine microorganisms we screened clones in large insert metagenomic libraries for genes able to complement an *E. coli* Phn$^-$ strain. Using this functional genomics approach, which does not depend on any *a priori* sequence knowledge, we identified several clones from a planktonic fosmid library that allowed the *E. coli* host to grow on 2-AEPn as the P source. The complementing cloned DNAs encoded a phosphonatase pathway, as well as a previously undescribed pathway for 2-AEPn utilization. We analyzed metagenomic data to examine frequency and distribution of these genes and showed that model marine microbes containing these genes can grow on Pn as a P source. Our results indicate that Pn is a common alternative P source for marine bacterioplankton and demonstrate the utility of functional screening approaches for assigning environmentally relevant functions to hypothetical genes detected in metagenomic surveys.

RESULTS

Functional screening for Pn utilization.

Seven fosmid libraries representing an open ocean depth profile at the Hawaii Ocean Time Series Station ALOHA (22º45' N, 158^{0}W) (DeLong et al., 2006) were screened for clones that could complement *E. coli* BW16787 for growth on MPn or 2-AEPn as the sole P source. This strain has a partial deletion in the *phn* operon encoding the C-P lyase (Δ*phnHIJKLMNOP*) which renders it incapable of growing on Pn, while still maintaining a functional Pn transporter encoded by *phnCDE*, the first three genes of the operon (**Figure 1**). BW16787 was previously used to successfully identify Pn genes from *Enterobacter aerogenes* and

Salmonella typhimurium by complementation (Lee et al., 1992; Jiang et al., 1995). The presence of the Pn transporter in the screening strain is important since not all Pn degrading clusters contain linked Pn transporter encoding genes (Lee et al., 1992; Huang et al., 2005; Quinn et al., 2007).

MPn utilization genes

Our functional screens did not yield any clones able to complement the ability of BW16787 to grow on MPn as a sole P source, despite the fact that methane production from MPn has been clearly shown in waters from the same location, and that C-P lyase sequence homologs have been identified in these metagenomic libraries (Karl et al., 2008). A previously sequenced fosmid from the HF70m library, HF70_[96]11A08 (APKI441) (Howard et al., 2008), was predicted to encode a cluster containing the regulatory *phnF*, and *phnGHIJKLNM* genes required for catalysis (**Figure 1**) and thus might be capable of complementing BW16787 for Mpn utilization. We transformed HF70_[96]11A08 into BW16787, but transformants could not use either MPn or 2-AEPn for growth (data not shown). The lack of complementation might be due to poor expression of this alphaproteobacterial fosmid in the heterologous host. Since we have previously observed that increasing the copy number of the fosmid vector can lead to increased gene expression in *E. coli* (Martinez et al., 2007), we constructed a copy up derivative of BW16787, BW16787 *trfA*, and tested HF70_[96]11A08 for complementation. Again, no significant growth on Mpn or 2-AEPn was observed under copy-up conditions in the deletion strain. Complementation was observed however in strains harboring individual in-frame deletions of the *phnH* and *phnN* genes (**Figure 1**) indicating that HF70_[96]11A08 does indeed encode a functional C-P lyase, and that the failure to complement the deletion strain might be due to inefficient expression or function of some of the C-P lyase components in the heterologous host.

2-AEPn utilization genes

We identified two unique clones in the HF130m library that allowed BW16787 to grow on 2-AEPn as P source. These clones were designated HF130_AEPn_1 and HF130_AEPn_2. While both clones clearly allow BW16787 to grow on 2-AEPn as sole P source, they did not allow utilization of 2-AEPn as N and C source as has been shown with several bacterial isolates (McGrath et al., 1997; Ternan and Quinn, 1998) (**Figure 2A**). When the complementation test was performed under copy up conditions in the new BW16787*trfA* strain, HF130_AEPn_2 allowed growth on 2-AEPn simultaneously as N and P source, and as N source in the presence of Pi. When used as a P source under copy up conditions, excess Pi was apparently released into the media by HF130_AEPn_2 and allowed the growth of the negative control strain on the same plate. Significant release of Pi into the growth medium was observed in liquid cultures of HF130_AEPn_2 grown on 2-AEPn under copy up conditions (**Figure S1?**). These results indicate that the enzymes encoded in HF130_AEPn_2 are not inhibited by excess phosphate, but more importantly, that they permit utilization of the amino group in 2-AEPn as the sole N source. HF130_AEPn_1, on the other hand, exhibited poor growth on 2-AEPn as P source under copy up conditions with only a few large colonies appearing over a background of microcolonies (**Figure 2A**). This growth pattern is often the result of toxicity caused by high level expression of one or more genes in the fosmid,

The substrate specificity of the pathways encoded by HF130_AEPn_1 and HF130_AEPn_2 was tested in liquid cultures using a variety of Pn compounds known to be transported into the cell and cleaved by the *E. coli* C-P lyase system (Metcalf and Wanner, 1991, 1993b). Both clones grew on 2-AEPn but not on Mpn, phosphonoacetate (PnAc), phosphonoformate or phosphite (**Figure 2B**). This narrow substrate specificity is uncharacteristic of C-P lyase pathways (Quinn et al., 2007). Both clones were fully sequenced and individual transposon insertions into each of the predicted ORFs were tested in the complementation assay to identify the gene(s) that conferred the 2-AEPn$^+$ phenotype.

HF130_AEPn_2 encodes a typical phosphonatase pathway

HF130_AEPn_2 contained a 32.3 kb insert encoding 4 tRNA genes and 25 predicted ORFs (**Table S1**). All the putative proteins were found to be highly similar to proteins of sequenced *Pseudomonas sp*. strains (most proteins with BLAST expectation values of less than 10^{-100}). Only three genes were found to be required for growth on 2-AEPn (**Figure 3**). All three are highly similar to *P. aeruginosa* PAO1 and *P. putida* genes encoding components of the well characterized phosphonatase pathway for 2-AEPn utilization (Dumora et al., 1983; Ternan and Quinn, 1998; Chen et al., 2002; Kim et al., 2002). Genes 20, 21, and 22 appear to form an operon. Gene 22, *phnW*, encodes a putative 2-AEPn:pyruvate aminotransferase (EC 2.6.1.37), the first enzyme of the phosphonatase pathway, which catalyzes the transfer of the amino group of 2-AEPn to pyruvate to form alanine and phosphonoacetaldehyde (**Figure 3**). Gene 21, *phnX*, encodes a protein of high similarity to phosphonoacetaldehyde hydrolase or phosphonatase (EC 3.11.1.1), which cleaves the C-P bond in phosphonoactetaldehyde to release phosphate. Utilization of 2-AEPn as N source with copy up was only observed when both *phnW* and *phnX* were functional (**Figure S1**), perhaps due to a toxic effect of phosphonoacetaldehyde accumulation in the *phnX* mutant. The third gene in this predicted operon, gene 20, *cybB*, encodes a putative cytochrome b561 family protein. A transposon insertion in this gene did not result in a loss of phenotype suggesting that this gene is not required for 2-AEPn metabolism in *E. coli*.

Gene 23, which is transcribed in the opposite direction and encodes a putative transcriptional regulator of the LysR family, was also required for 2-AEPn metabolism. In *P. aeruginosa*, the LysR protein encoded in the phosphonatase cluster binds 2-AEPn and activates transcription of the *phnWX* operon (Quinn et al., 2007). Consistent with its predicted role as transcriptional activator, the *lysR* gene in HF130_AEPn_2 was not required for 2-AEPn utilization under copy up conditions (**Figure S1**).

The phenotype of this clone, including the specificity for 2-AEPn, the capacity to use 2-AEPn as P and N source, and lack of inhibition by Pi, are entirely consistent with the

functional assignment of its Pn utilization genes. This represents the first example of functional characterization of a phosphonatase pathway from a marine microbe.

As an initial analysis of the prevalence of the phosphonatase pathway in marine plankton, we searched the genomes of fully sequenced marine microbes (http://www.moore.org/microgenome/) for *phnX* and *phnW* genes. These genes were often found linked in the chromosomes of many marine microbes (**Figure 3**). For example, *phnW* and *phnX* genes, separated by a putative aminotransferase gene, were common in Gammaproteobacteria, including many members of the Vibrionales (such as *V. angustum* S14), Alteromodales (*Moritella sp. PE36* and *Psychromonas*) and Oceanospirillales (*Marinomonas sp. MED121*). Linked genes were also found in the alphaproteobacterium BAL-199, *Planctomyces maris* DSM8797, and in *Bacillus sp.* B14905. These results suggest that the phosphonatase pathway genes play a significant role in the marine environment.

HF130_AEPn_1 encodes a novel pathway for 2-AEPn utilization

HF130_AEPn_1 contained a 42.1 kb insert encoding 25 predicted ORFs (**Table S2**). In contrast to HF130_AEPn_2, the phylogenetic affiliation of HF130_AEPn_1 is not entirely clear, although BLAST analyses suggest it might belong to a representative of the Deltaproteobacteria. More importantly, none of the predicted ORFs on this environmental fragment encoded proteins with similarity to any known Pn utilization genes, indicating that this clone likely encodes a novel Pn utilization pathway.

Phenotypic analysis of transposon insertion mutants in the predicted ORFs revealed that two genes, gene 5 and gene 6 are required for 2-AEPn utilization **(Figure 4**). We have named these genes *phnY* and *phnZ*, respectively. The best BLAST hits for these genes in the NCBI database are listed in **Table S3**. *phnY* is most similar to poorly characterized proteins, some of which are annotated as phytanoyl-CoA dioxygenases based on weak sequence similarity. Comparison of the predicted protein sequence against the conserved domain database (CDD (Marchler-Bauer et al., 2007)) revealed that PhnY indeed belongs to the PhyH

superfamily (pfam05721 and COG5285 with expectation values of $4x10^{-22}$ and $3x10^{-12}$, respectively). PhyH is a 2-oxoglutarate dependent dioxygenase that catalyzes the alpha-oxidation of phytanoyl-CoA coupled with the oxidative decarboxylation of 2-oxoglutarate to form succinate and carbon dioxide (Schofield and McDonough, 2007). This family includes eukaryotic phytanoyl-CoA dioxygenases and a number of bacterial dioxygenases mostly of unknown function. Interestingly, a member of this family, HtxA, is a hypophosphite dioxygenase in *Pseudomonas stutzeri* WM88 that oxidizes hypophosphite to phosphite, as well as phosphite to phosphate *in vitro* (White and Metcalf, 2002). Despite this similarity, HF130_AEPn_1 did not allow growth on phosphite or hypophosphite. All residues known to be required for Fe^{2+} and 2-oxoglutarate binding in PhyH and related enzymes (Hogan et al., 2000; Schofield and McDonough, 2007) are conserved in the predicted PhnY protein, supporting its functional assignment as a 2-oxoglutarate dioxygenase.

phnZ encodes a protein of the HD superfamily (pfam01966 and COG4341, with expectation values of $9x10^{-06}$ and $6x10^{-29}$, respectively). This family is characterized by the presence of an HD motif and additional conserved D and H residues believed to be involved in metal binding (Aravind and Koonin, 1998). Although members of this family have been predicted to act as divalent metal-dependent phosphohydrolases, none of the genes in the database that share significant sequence similarity with *phnZ* have been functionally characterized. In several bacterial species, however, *phnZ* homologs are found linked to genes similar to other Pn degradation genes, supporting its involvement in Pn metabolism. For example, it is found adjacent to C-P lyase genes in *Nostoc sp*. PCC 7120, and *Pelagibacter sp*. HTCC7211, next to *phnX* in *Gemmata obscuriglobus* UQM 2246, and adjacent to *phnA* in several *Burkholderia sp*. including *B. cepacia* AMMD and *B. cenocepacia* HI2424 (data not shown).

Interestingly, several other marine bacterial species contain linked homologs of both of these genes suggesting that they may also use 2-AEPn as a P source (**Figure 4).** These include *Plesiocystis pacifica* SIR-1 (a deltaproteobacterium), *Planctomyces maris* DSM8797,

and two strains of the unicellular cyanobacterium, *Prochlorococcus marinus* (MIT9301 and MIT9303). *phnY* and *phnZ* are also found in tandem in several fungal species, such a *Aspergillus niger* (**Figure 4**), which is particularly interesting since *A. niger* and other fungi can use 2-AEPn as P source by yet uncharacterized pathways (Krzysko-Lupicka et al., 1997). The sporadic appearance of these genes in isolated strains within a lineage and the fact that their phylogenies (**Figures S2 and S3**) do not generally correspond to phylogenies of highly conserved genes suggests that they might have undergone horizontal gene transfer in the marine environment. Interestingly, a homolog of the HD gene is found in a mimivirus (Raoult et al., 2004), pointing to a possible mechanism for horizontal gene transfer.

Growth of marine microbes on Pn

The presence of genes similar to those identified in the functional screens in fully sequenced marine microorganisms suggests that they might be able to grow on 2-AEPn as a P source. To test this hypothesis we analyzed the ability of several marine bacteria to grow on Pn as the sole P source (**Figure 5**). *Vibrio angustum* S14 (which appears to have a phosphonatase pathway and a C-P lyase pathway), *Planctomyces maris* DSM8797 (which contains a phosphonatase pathway and the new *phnY/phnZ* pathway), and *Ruegeria pomeyori* DSS-3 (which contains a C-P lyase pathway and a homolog of *phnZ*) were all able to grow on both 2-AEPn and Mpn as the sole P source. None of these strains were able to grow on phosphite, a substrate of the C-P lyase in *E. coli*. The lack of growth of *E. coli* BW16787 (Phn^- strain used in the screen) indicates that no significant Pi was present in the Pn-containing media during the course of the experiments.

Abundance of 2-AEPn genes in marine metagenomic databases

We extended previous analyses looking for the presence of Pn genes in marine metagenomic datasets (Quinn et al., 2007; Gilbert et al., 2008; Karl et al., 2008) by examining the abundance of *phnX, phnW, phnY*, and *phnZ* in the Global Ocean Survey (GOS) database

(Yooseph et al., 2007). In addition to the genes identified in our screen, we included the other Pn genes previously identified in marine environments: *phnA*, encoding phosphonoacetate hydrolase (Gilbert et al., 2008), *phnJ* and *phnI* as representatives of the C-P lyase pathway. *recA* and *gyrB* were included as a reference single copy genes. For each analyzed gene, abundance is expressed as the percentage of all bacteria containing the gene in question, based on the assumption that *recA* is present as a single copy gene in every genome. Although the variability across samples is large in some cases, the average abundance of the Pn genes across samples was 2% for *phnY*, 7% for *phnZ*, 8% for *phnW*, and 1% for *phnX* (**Table S4**). These frequencies are within the same order of magnitude as those observed for *phnI, phnJ*, and *phnA* (6, 8 and 9% respectively). These results indicate that the 2-AEPn utilization pathways described in this study, as well as the C-P lyase and PnAc degradation pathways, are widespread and common in surface water marine bacterioplankton.

The distribution of Pn genes throughout the water column was investigated using metagenomic data from depth profiles obtained from station ALOHA in the North Pacific Gyre (from the same site as the DNA used for the fosmid libraries), and from the BATS station near Bermuda in the Sargasso Sea. Analysis of depth profiles is important because the fosmid screen hits came from the130 m sample, well below the surface waters analyzed in the GOS survey. Additionally, these two ecosystems are characterized by vastly different surface phosphate concentrations, with DIP concentrations in the Sargasso Sea (0.2-1.0 nM) being one to two orders of magnitude lower than at ALOHA station (Wu et al., 2000). The results for the available shotgun libraries are shown in **Table 1**. In the Station ALOHA samples, the *phnZ* is found in 10% or less of the microorganisms throughout the water column. The *phnY* gene however was found in the surface shotgun libraries starting at the 125 m sample suggesting that the complete pathway, as found in HF130_AEPn_1, may be more prevalent below surface waters. The *phnX* and *phnW* genes are found in low frequency in the shotgun libraries from the top 125 m of the water column in Station ALOHA , but their frequency increased substantially with depth. In the 4000 m library 63% and 20% of the microorganisms

were estimated to have a *phnW* and *phnX* gene, respectively. The *phnA* gene was also found at high frequency in the deep water samples (present in up to 26% of microorganims). These results were unexpected since the DIP concentration is significantly lower in the upper 100 m of the water column at this location ((Bjorkman and Karl, 2003). The analysis of the putative phylogenetic affiliation of these deep water *phn* genes (**Figure 6 and Table S5**) revealed that many diverse bacterial groups contribute to the *phnA* and *phnW* signals. The *phnX* signal at 4000 m was assigned to Alphaproteobacteria and Eukaryotes based on the top blast hit against the NCBI non-redundant protein database (NR). Although the putative proteins were most similar to eukaryotic proteins, the encoding genes in our libraries did not contain introns suggesting that they are probably bacterial in origin.

In contrast, the 20 and 50 m shotgun libraries from the Sargasso Sea (**Table 1**) had very high frequency of Pn utilization genes, particularly C-P lyase genes (*phnI* and *phnJ*), *phnZ* and *phnX* genes (between 10 and 25% of the microbes). The vast majority of these hits (**Table S5**) were most similar to genes from *Candidatus Pelagibacter sp*. HTCC7211, a highly abundant member of the SAR11 clade of Alphaproteobacteria from the Sargasso Sea (Stingl et al., 2007). Sequence homologs most similar to those in other Alphaproteobacteria (Rhodobacterales family) were also observed, as well as one sequence of a *phnZ* gene most similar to that of *Prochlorococcus marinus* MIT9301. *phnW* and *phnA* genes were present in approximately 5-10% of the microorganisms in these samples, and in contrast with the other *phn* genes analyzed, were not found in SAR11 members but instead appeared to belong primarily to other proteobacterial groups.

To further support these observations we also analyzed metagenomic data obtained by pyrosequencing from the same stations. For Station ALOHA, the pyrosequencing data was obtained from the HOT186 depth profile collected on October 2006, seven months after the HOT179 profile used to construct the shutgun libraries. Despite this difference, the trends observed for *phn* gene abundance are remarkably similar for both libraries. As observed in the shotgun data from Station ALOHA, the frequencies of *phnY, phnZ, phnX* and *phnW*

increased with depth, reaching levels of 6-22% at 500 m, while the *phnI* and *phnJ* genes of the C-P lyase were found at very low frequencies throughout the profile (not more than 1% of the microbes) (**Table 2**). The frequency of Pn utilization genes in surface waters was again significantly higher in the Sargasso Sea samples. Detailed examination of the hits revealed several environmental trends (**Figure 6, Table S6).** For example, the *phnY* and *phnZ* at 50 and 100m include many instances of genes apparently derived from *Prochlorococcus marinus* MIT9303 and MIT9301. These sequences were significantly lower in numbers at Station ALOHA, suggesting that Pn utilization by these microorganisms is more prevalent in the Sargasso Sea. *Pelagibacter sp*. homologs of *phnX, phnZ, phnI* and *phnJ* were highly represented at 20 and 50m, while other Alphaproteobacteria (Rhodobacterales and Rhizobiales) were highly represented in the *phnA* and C-P lyase genes found near the surface. Two matches to C-P lyase genes from Cyanobacteria (*Synechococcus* sp. PCC 7335 and *Cyanothece* sp. PCC 8802) were also detected in these samples. Interestingly, while the C-P lyase genes were no longer detected at 500 m, there was still a significant fraction of microorganisms harboring *phnZ, phnW* and *phnA*. As was the case in the Station ALOHA samples, although Beta- and Gammaproteobacteria were abundant, many diverse bacterial groups appear to contribute to this deep water signal (**Figure 6**, **Table S6**).

We also analyzed publicly available marine metatranscriptomic data (**Table S7**). All the Pn genes analyzed in this study (*phnY, phnZ, phnX, phnW, phnI*, and *phnJ*) were found in a metatranscriptomic analysis of the Sargasso Sea by Moran and colleagues (http://web.camera.calit2.net/cameraweb/id?CAM_PROJ_SargassoSea). Evidence of expression for these genes was also found in surface waters of other regions by Zehr and collegues (http://web.camera.calit2.net/cameraweb/id?CAM_PROJ_GeneExpression).

DISCUSSION

One major challenge in contemporary environmental microbiology is to interpret the vast amounts of genomic information now accumulating (DeLong, 2009). This is particularly true

for marine ecosystems, where large-scale metagenomic and metatranscriptomic sequencing projects such as the Global Ocean Survey have uncovered more than six million new proteins, many of which lack significant homology to proteins with functional assignments (DeLong et al., 2006; Yooseph et al., 2007; Frias-Lopez et al., 2008). One approach for determining the function of novel genes is to use classical genetics to identify mutants in particular functions. This method was used successfully to identify the *dmdA* gene responsible for DMSP utilization in *Ruegeria pomeyori* (Reisch et al., 2008). Unfortunately, this approach is limited to the small number of marine isolates with well-developed genetic systems. An alternative is to perform functional screens in which DNA libraries are analyzed in a heterologous host for expression of a function of interest. We have previously used this approach to identify marine fosmids that express a complete proteorhodopsin photosytem from a marine Alpharoteobacteria in *E. coli* (Martinez et al., 2007). In this study we screened marine fosmid libraries to identify genes that confer the ability to utilize 2-AEPn and Mpn to an *E. coli* strain defective in Pn utilization (Lee et al., 1992). Our results demonstrate both the advantages and limitations of this approach. A major limitation of functional screens is that they require expression and correct protein assembly and localization in the heterologous host. Here, a fosmid predicted to contain all the C-P lyase genes necessary to restore catalytic activity (Yakovleva et al., 1998) was unable to restore growth on Mpn in the deletion strain. Complementation was observed however for strains containing a single, non-polar mutation in *phnG* and *phnN*. These results are perhaps not surprising given the known complexity of the multi-component, membrane associated C-P lyase system (Metcalf and Wanner, 1993a; Yakovleva et al., 1998). Despite these limitations, the major advantage of the functional screening approach is that it is not dependent on prior sequence knowledge and thus, it can identify genes not previously associated with the function of interest.

To explore Pn utilization pathways in marine microbial communities, we screened seven fosmid libraries prepared from microbial DNA from different depths at station ALOHA in the North Pacific Gyre (DeLong et al., 2006). Although the surface waters at station

ALOHA are not as phosphate limited as other marine environments (Karl, 1999; Wu et al., 2000), previous studies have shown that these libraries did contain C-P lyase genes and that methane was produced from Mpn in these waters (Karl et al., 2008), indicating that microbes at this site have the potential to use Pn as P source. Indeed, we identified two pathways for 2-AEPn utilization in these libraries. 2-AEPn is a component of phosphonolipids and phosphonoproteins in numerous marine invertebrates (Horiguchi, 1984; Ternan et al., 1998), thus it should be available for microbial utilization in the marine environment. The first pathway identified is a phophonatase pathway that allows utilization of 2-AEPn as both P, and N source by virtue of the first transamination step that leads to formation of alanine. The gene sequence and genome context in the clone encoding this function is almost identical to that of the well characterized *Pseudomonas* phosphonatase pathways (Dumora et al., 1997; Ternan and Quinn, 1998). In these strains expression of the pathway is not under control of the Pho regulon, but is instead mediated by a LysR-type transcriptional regulator (also present in the fosmid clone) and controlled by substrate availability (Quinn et al., 2007). These findings are consistent with 2-AEPn being a source of not only P, but also of N and even C. Among marine microbes, the complete phophonatase pathway (including both *phnX* and *phnW*) is found in numerous fully sequenced Gammaproteobacteria genomes, in one alphaproteobacterium, one *Planctomyces,* and one *Bacillus* genome.

The second pathway for 2-AEPn utilization identified in this study was a novel pathway, encoded by two genes without prior association with Pn metabolism and whose homologs had not previously been functionally characterized. Although the complete description of the pathway requires further biochemical analysis, we can speculate based on the presence of conserved domains in the encoded proteins. One possible pathway may involve the initial hydrolysis of the C-P bond in 2-AEPn by the predicted HD hydrolase, encoded by *phnZ,* to release phosphite. The phosphite could then be oxidized to phosphate by PhnY. HtxA, hypophosphate dioxygenase, a member of the same protein family as PhnY, has been shown to perform this reaction *in vitro* (White and Metcalf, 2002). However, the strain

containing the HF130_AEPn_1 fosmid was unable to grow on phosphite as a P source as would have been expected if phosphite was an intermediary in this pathway, given that *E. coli* appears to be able to transport Pt into the cell (Metcalf and Wanner, 1991, 1993b). Alternatively, PhnY might act on 2-AEPn first by introducing a OH group in the C1 position, as does *E. coli* TauD, also a 2-oxoglutarate dioxygenase, on taurine (2-aminoethylsulfonate) (Eichhorn et al., 1997). Hydroxylation of the C1 of taurine has been shown to weaken the C-S bond leading to its non-enzymatic breakage. Perhaps PhnY-mediated hydroxylation of 2-AEPn destabilizes the C-P bond making it susceptible to hydrolytic cleavage by PhnZ with the release of phosphate. Indeed, HD domain proteins are predicted to function as phosphohydrolases and contain residues thought to be involved in divalent metal coordination (Aravind and Koonin, 1998) which could stabilize a transition state for the C-P bond cleavage. The fact that a mutant fosmid lacking PhnY cannot grow on 2-AEPn supports the idea that PhnZ cannot hydrolyze 2-AEPn directly, but rather that it needs to be modified by PhnY before C-P bond cleavage. Although initial attempts to characterize these activities in crude extracts were not successful, the availability of the genes encoding these enzymes can now enable their overexpression and purification for biochemical analysis of this novel 2-AEPn pathway.

The *phnY/phnZ* pathway is sporadically found in the genomes of diverse bacterial taxa, including Cyanobacteria, Planctomycetes, and Proteobacteria, and in several fungi. The presence of the pathway in discrete members of distantly related bacteria instead of being lineage specific, suggest that it might have been subject to horizontal gene transfer (HGT) and that there is a strong selection for the ability to utilize Pn in the environment. Similar conclusions had been reached in analyses of the C-P lyase and phosphonatase pathway (Huang et al., 2005). The presence of a *phnZ* gene in the genome of a mimivirus (Raoult et al., 2004) suggests that viruses might be a vehicle for HGT, as has been seen for other host genes (Lindell et al., 2004; Sobecky and Hazen, 2009). Although HGT between bacteria and fungi

seems quite rare, it has also been suggested for other genes, including the *dddP* gene involved in dimethylsulfoniopropionate metabolism in marine bacteria (Todd et al., 2009).

The novel *phnY/phnZ* pathway is present in only 2 of the 11 available genomes of *Prochlorococcus marinus* strains (MIT9301 and MIT9303). Preliminary attempts to grow MIT9301 and MIT9303 on 2-AEPn as a P source have not been successful (Osburne and Chisholm, personal communication), but recent experiments have shown that expression of the *phnZ* and *phnY* genes in MIT9301 is greatly increased under phosphate limiting culture conditions (Coleman and Chisholm, unpublished). In both strains the *phnY/phnZ* pathway is located next to a putative Pn ABC transporter within a genomic island (Coleman et al., 2006; Martiny et al., 2006), again suggesting that these strains may have been acquired this pathway by HGT. Interestingly, both strains were isolated from the Sargasso Sea, which is characterized by extremely low DIP concentrations in the surface layer during the stratified summer months (Wu et al., 2000). The metagenomic analysis reported here indicates that these *Prochlorococcus phnY/phnZ* homologs are significantly enriched in the Sargasso Sea surface waters compared with the higher DIP Station ALOHA samples. An independent analysis of *Prochlorococcus* gene frequencies in the same datasets supports these observations (Colleman and Chisholm, unpublished). These results suggest that the ability to utilize Pn conferred by the *phnY/phnZ* pathway provides a selective advantage for *Prochlorococcus* in low DIP ecosystems. Utilization of Pn by some strains of *Prochlorococcus* would have broad implications for primary productivity in P-limited ecosystems.

The increased representation of Pn utilization genes in surface waters at the Sargasso Sea site compared to Station ALOHA is not exclusive to *Prochlorococcus*. Indeed, all Pn genes analyzed follow that trend. For example, the *phnI* and *phnJ* genes of the C-P lyase pathway are present in more than 16% of the bacteria in the 20 and 50 m samples at BATS, and less than 1% at all depths at Station ALOHA. This signal is almost exclusively from Alphaproteobacteria in both systems, with *Candidatus Pelagibacter sp*. HCCT7211 providing the dominant signal in the Sargasso Sea. We have shown here that *Ruegeria pomeyori* DSS-3,

an alphaproteobacterium represented in these data, is capable of using 2-AEPn and MPn as a P source providing direct evidence for Pn utilization in this group. We also found C-P lyase genes of two cyanobacterial species, *Cyanothece* sp. PCC 8802 and *Synechococcus* sp. PCC 7335, suggesting that they might be able to use Pn as P source, similar to the filamentous cyanobacterium *Trichodesmium* (Dyhrman et al., 2006). Finally, there is high representation of *phnA*, which is involved in phosphonoacetate degradation, and *phnX* and *phnW* genes (phosphonatase pathway) in the Sargasso Sea surface waters. This signal appears to originate in large part from alphaproteobacterial members of the community, but also includes *Pseudomonas*-like sequences similar to those identified in this study. These results provide strong evidence that Pn use is widespread among the microbial community that thrives in P limited waters of the Western North Atlantic. Although no expression data for the Sargasso Sea depth profile analyzed here is yet available, we found hits to *phnY, phnZ, phnA, phnI* and *phnJ* in metatrancriptomic data from the Sargasso Sea (M. A. Moran, unpublished) and other ocean regions (J. Zehr, unpublished). In addition, expression of C-P lyase genes from *Trichodesmium* (Dyhrman et al., 2006) and the periplasmic component of the Pn ABC transporter, PhnD from *Pelagibacter ubique* HTCC7211 (Sowell et al., 2009) have been observed in the Sargasso Sea supporting the conclusion that Pn utilization genes are expressed in this environment.

At both stations, the frequency of some Pn utilization genes was high in deep water samples, even though the measured DIP concentrations are significantly higher than in surface waters. In the HOT samples the increase in *phn* gene frequency was most evident below the 100 m mark (the depth where the measured DIP concentration starts to increase in station ALOHA (Bjorkman and Karl, 2003)) and continued all the way to the 4000 m sample. The frequency of *phnX, phnW,* and *phnA* genes was particularly high and appeared to be derived from representatives of numerous diverse bacterial groups. One possible explanation for Pn utilization in the presence of high concentrations of DIP comes from the fact that the most abundant Pn genes in the deep belong to Pn utilization pathways that allow their use not only

as P but also as a source of N and even C. In other microorganisms where that has been shown to be the case, these genes are expressed in the presence of their substrate but independently of P availability (McMullan and Quinn, 1992; McGrath et al., 1997; Ternan and Quinn, 1998; O'Loughlin et al., 2006; Gilbert et al., 2008). Consistent with this hypothesis, we have shown that the fosmid clone containing the phosphonatase pathway isolated from a 130 m sample allows the use of 2-AEPn as both P and N source. Also, it has been recently shown that marine *Vibrio* isolates that contain a *phnA* gene can use phosphonoacetate as a source of P and C (Gilbert et al., 2008), and that 2-AEPn and PnAc pathways appear to be under catabolite repression in *Agromyces fucosus* Vs2 (O'Loughlin et al., 2006) underscoring the relevance of these compounds as a C source. It seems possible that the increase in *phnX*, *phnW* and *phnA* representation in deep waters reflects an adaptation to the use these compounds as a complete P, N and C source by heterotrophic bacteria and thus are more prevalent below the photic zone. This observation is consistent with chemical analysis data of HMW DOM and sinking particles that have shown that Pns are remineralized throughout the water column (Clark et al., 1999; Benitez-Nelson et al., 2004) and with the observation that the moderately N- and P- rich fraction of the DOM is labile and it is decomposed preferentially (Hopkinson and Vallino, 2005).

In summary, our results suggest that functional screens are a powerful approach for characterizing the function of hypothetical genes in genomic and metagenomic data sets, especially when combined with metagenomic analysis and laboratory experiments in marine model microorganisms. More importantly, our results indicate that the utilization of Pn metabolism in the oceans is widespread among diverse and abundant bacterial groups, and it is likely to play an important role not only in the P, also the N and C biogeochemical cycles.

EXPERIMENTAL PROCEDURES

Chemicals, media, and bacterial strains.

Methylphosphonate (MPn), 2-aminoethylphosphonate (2-AEPn), phosphonoformate (PnF), phosphonoacetate (PnAc), and phosphite (Pt) were from Sigma Aldrich. MOPS minimal medium and MOPS minimal medium without NH_4 were purchased from Tecknova (Hollister, CA). Phosphate Colorimetric Assay kit was from BioVision, Inc. (Mountain View, CA). *E. coli* BW16787 and BW18812 were obtained from B. Wanner through the *E. Coli* Genetic Stock Center. *Ruegeria pomeyori* DSS-3, *Vibrio angustum* S14, and *Planctomyces maris* DSM8797 were obtained from M.A. Moran, S. Kjelleberg, and ATCC, respectively.

Screening of metagenomic libraries for Pn utilization.

Marine picoplankton fosmid libraries were previously constructed from samples collected along a depth profile from the Hawaii Ocean Time series (HOT) station ALOHA (22º45' N, 158^{0}W) at cloned into the copy-control pCC1FOS fosmid vector (Epicentre) (DeLong et al., 2006). These libraries were pooled and fosmid DNA was isolated by alkaline lysis followed by cesium chloride ultracentrifugation (Sambrook et al., 1989). Aliquots of the pooled fosmid DNA were used to transform BW16787 by electroporation (1.2 Kv/cm, 200 Ohms, 25 μF). After electroporation, cells were incubated at 37ºC in 1 ml SOC to allow for antibiotic resistance expression, and rinsed twice in 5 ml of MOPS buffer to remove phosphate prior to plating in selective medium. Plating media consisted of MOPS minimal medium with 0.4% glycerol, 12 μg/ml chloramphenicol, and 0.1 mM MPn or 2-AEPn. A small fraction of each transformation was also plated in minimal medium containing Pi to evaluate the transformation efficiency and estimate library coverage. Screening plates were incubated at 30ºC for up to 20 days. The empty fosmid vector, CCFOS1, was used as a negative control. Negative control and screening plates always contained a large number of microcolonies that arose from the utilization of the residual Pi in this medium (Metcalf and Wanner, 1993b). Only large colonies clearly distinguishable from that background were chosen for further characterization. Phosphonate positive clones were restreaked on selective

medium and fosmid DNA was isolated and retransformed into BW16787 to confirm the phenotype. Restriction analysis and fosmid end sequencing were used to identify replicates. Unique fosmids were sequenced using transposon mutagenesis as described (Martinez et al., 2007). The complete DNA sequence was assembled using Sequencher v. 4.5 (Gene Codes Corporation) and annotated with FGENESB (Softberry) and Artemis v. 6 (The Wellcome Trust Sanger Institute). At least one clone containing a transposon insertion in each predicted ORF was transformed back into BW16787 and screened for growth on Pn as above.

Growth on 2-AEPn as N source.

To evaluate the ability of library clones to grow on 2-AEPn as N source, clones were streaked on NH_4-free MOPS minimal medium containing 5mM 2-AEPn with or without addition of 0.1 mM Pi. Clones were tested in the original BW16787 strain as well as in a new copy-up derivative, BW16787 *trfA*, created by P1 transduction of the copy-up P_{BAD}-*trfA* marker from the Translator strain (Lucigen Corp., Middleton, WI). For copy-up conditions, 0.2% arabinose was added to the growth medium.

Complementation analysis of individual C-P lyase genes.

Strains harboring individual mutations in the catalytic genes of *E. coli* C-P lyase were obtained from the Keio collection and in frame deletions were constructed as described (Datsenko and Wanner, 2000; Baba et al., 2006). Empty CCFOS1 vector and library clones were transformed into the resulting strains and growth on Pn was tested on plates as above.

Pn specificity tests for fosmid library clones.

Pn specificity was evaluated in microtiter plate liquid cultures as follows. BW16787 strains containing the identified Pn-positve clones or the empty fosmid vector CCFOS1, and BW18812 (CCFOS1) as a Pn^+ control were grown for 48 hrs in MOPS glycerol medium with

0.1 mM Pi. Cells were pelleted by centrifugation, rinsed twice in P-free MOPS buffer, and used to inoculate in triplicate 150 ml of medium containing 0.2mM P source. Plates were incubated at 30ºC for 2 days. Growth on each P source was measured using the WST-1 cell proliferation assay (Roche Molecular Diagnostics) which measures respiratory activity by the reduction of the soluble tetrazolium salt WST-1 to colored formazan (Ishiyama et al., 1996)

Growth of marine natural isolates on Pn as a P source.

Bacterial isolates to be tested for growth on Pn as P source were grown for 48h in a modified VNSS medium (0.1xVNSS, 0.1 g peptone, 0.05 g yeast extract, 0.05 g glucose, 0.05 g starch, 1mg $FeSO_4.7H_2O$, 1 mg NaH_2PO_4 per liter of nine salts solution (NSS) (Marden et al., 1985)) at room temperature, rinsed twice in P-free NSS, and resuspended in 1 volume equivalent of NSS. 10 µl of inoculum were added in triplicate to microtiter plate wells containing 160 µl of growth medium (NSS with 4g Glucose, 2.2 g $(NH_4)_2SO_4$, and 1 ml vitamin solution (Gonzalez et al., 2003) per liter) with 0.1 mM P source as indicated. Growth was monitored by measuring optical density (440 nm) in a Biotek Synergy2 plate reader. The stability of the Pn compounds during the course of the experiment was monitored using a colorimetric phosphate assay (BioVision Inc., Mountain View, CA) in an uninoculated plate incubated under the same conditions. In addition, *E. coli* BW16787 (Δ*phn*), which cannot utilize reduced P compounds, was used as negative control. Viability of the negative control strain in the Pn media during the entire course of the experiment was tested by showing that growth could take place upon Pi addition after 5 days of incubation. For strains that only reached low optical density, cell proliferation was also measured using the WST-1 viability assay (Roche Applied Sciences, Indianapolis, IN).

Sample collection, DNA extraction and pyrosequencing.

Bacterioplankton samples were obtained form Hawaii Ocean Time-series Station ALOHA in the North Pacific Subtropical Gyre (22º44'N, 158º2'W) and BATS Station in the

Sargasso Sea (31°40'N, 64°10'W). At each site, bacterioplankton samples were collected from the photic zone at the mixed layer, just below the mixed layer, and at the deep chlorophyll maximum (25, 75, and 125 m for HOT179, 25, 75 and 110m for HOT186, and 25, 50, and 100m for BATS216), and the mesopelagic zone (500m). The H4000m shotgun library has been previously described (Kostantinidis et al., 2009). Sample collection and DNA extraction were performed as previously described (Frias-Lopez et al., 2008). 3-5 μg of community DNA were sequenced using Genome Sequencer FLX (Roche).

Bioinformatics.

Conserved domains were identified using CDD (Marchler-Bauer et al., 2007). Abundance and distribution of Pn genes in the databases was performed as follows. Deduced peptide sequences of the Pn genes identified in the screens, PhnI (4080494), PhnJ (40804945) from *Pseudomonas stutzeri,* PhnA (1196755) from *P. fluorescens*, and *E. coli* RecA and GyrB as single copy reference were used as query to interrogate available databases using NCBI Blast. An expectation cutoff value of $1x10^{-20}$ was used for NCBI NR, GOS, HOTS and BATS shotgun data. For shorter pyrosequencing reads, a bit cutoff value of 40 was used initially (DeLong et al., 2006; Yooseph et al., 2007; Frias-Lopez et al., 2008) but was verified for each gene by comparing the pyosequencing hits against NR and selecting cutoff values that returned as best blast hit only proteins identified as members of the same family as the query gene by annotation or phylogenetic tree analyses (bellow). Based on these analyses, bit score cutoff of 40 was used for all genes except for *phnZ* and *phnW* (b>45 and b>50, respectively). Gene counts were size normalized using the query protein length. The percentage of microbes containing each gene was calculated assuming that *recA* is present in single copy in every microbial genome (Howard et al., 2008; Reisch et al., 2008).

Phylogenetic Analyses

Homologues of HF130_AEPn_1 *phnY* and *phnZ* were identified in public databases

by comparing the amino acids sequences for each gene against the non-redundant (NR) NCBI database and the peptide database available for the Global Ocean Survey (GOS) using BLAST (Altschul et al., 1997). Peptide sequences matching our query sequences with an expectation value of less than $1e^{-20}$ and aligning over >85% of the query protein length was considered significant and used in subsequent analyses. Datasets for each gene aligned using ClustalW version 1.7 (Thompson et al., 1994) and manually refined using the ARB software package (Ludwig et al., 2004). Masks were created in ARB using the base frequency filter tool (20% minimal similarity) to remove hypervariable regions. This produced masked alignments of 233 and 173 amino acids for *phnY* and *phnZ* respectively. Parsimonydendrograms were constructed using the masked amino acid alignments in ARB. Support for interior nodes of the dendrograms was determined using 500 bootstrap resamplings of maximum parsimony (MP) trees calculated in PAUP* version 4.0b2a (D. L. Swofford, Sinauer Associates, Sunderland, MA).

ACKNOWLEDGEMENTS

We thank M.A. Moran, S. Kjelleberg, B. Wanner and the CGSC for bacterial strains; the captain and crew of the R/V Kilo Moana and R/V Atlantic Explorer, the HOTS and BATS teams, and Tracy Mincer, Jay McCarren, Yanmei Shi, Matt Sullivan, Maureen Colleman, Suzanne Kern, and Sarah Bagby for collecting and processing metagenomic samples; Rachel Barry and Stephan Schuster and colleagues for pyrosequencing HOT186 and BATS216 samples, respectively; Tsultrim Palden for assistance with fosmid sequencing; and the Joint Genomics Institute for constructing and sequencing shot-gun libraries. We also thank Marcia Osburne for comments on the manuscript. This work was supported by a grant from the Gordon and Betty Moore Foundation (EFD), the Department of Energy Genomics GTL Program, and NSF Microbial Observatory award (MCB-0348001).

REFERENCES

Altschul, S.F., Madden, T.L., Schaeffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucl. Acids Res.* **25**: 3389-3402.

Aravind, L., and Koonin, E.V. (1998) The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends Biochem Sci* **23**: 469-472.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M. et al. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**: 2006 0008.

Benitez-Nelson, C.R., O'Neill, L., Kolowith, L.C., Pellecia, P., and Thunell, R. (2004) Phosphonates and particulate organic phosphorus cycling in an anoxic marine basin. *Limnol. Oceanogr.* **49**: 1593-1604.

Bjorkman, K.M., and Karl, D.M. (2003) Bioavailability of dissolved organic phosphorus in the euphotic zone at Station ALOHA, North Pacific Subtropical Gyre. *Limnol. Oceanogr.* **48**: 1049-1057.

Chen, C.C., Zhang, H., Kim, A.D., Howard, A., Sheldrick, G.M., Mariano-Dunaway, D., and Herzberg, O. (2002) Degradation pathway of the phosphonate ciliatine: crystal structure of 2-aminoethylphosphonate transaminase. *Biochemistry* **41**: 13162-13169.

Clark, L.L., Ingall, E.D., and Benner, R. (1999) Marine organic phosphorus cycling: novel insights from nuclear magnetic resonance. *Am J Science* **2999**: 724-737.

Coleman, M.L., Sullivan, M.B., Martiny, A.C., Steglich, C., Barry, K., Delong, E.F., and Chisholm, S.W. (2006) Genomic islands and the ecology and evolution of Prochlorococcus. *Science* **311**: 1768-1770.

Datsenko, K.A., and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**: 6640-6645.

DeLong, E.F. (2009) Microbial oceanography from genomes to biomes. *Nature* **459**.

DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.-U. et al. (2006) Community Genomics Among Stratified Microbial Assemblages in the Ocean's Interior. *Science* **311**: 496-503.

Dumora, C., Lacoste, A.M., and Cassaigne, A. (1983) Purification and properties of 2-aminoethylphosphonate:pyruvate aminotransferase from Pseudomonas aeruginosa. *Eur J Biochem* **133**: 119-125.

Dumora, C., Marche, M., Doignon, F., Aigle, M., Cassaigne, A., and Crouzet, M. (1997) First characterization of the phosphonoacetaldehyde hydrolase gene of Pseudomonas aeruginosa. *Gene* **197**: 405-412.

Dyhrman, S.T., Ammerman, J.W., and Van Moy, B.A.S. (2007) Microbes and the marine phosphorus cycle. *Oceanography* **20**: 110-116.

Dyhrman, S.T., Chappell, P.D., Haley, S.T., Moffett, J.W., Orchard, E.D., Waterbury, J.B., and Webb, E.A. (2006) Phosphonate utilization by the globally important marine diazotroph Trichodesmium. **439**: 68-71.

Egli, T. (1988) (An)aerobic breakdown of chelating agents used in household detergents. *Microbiol Sci* **5**: 36-41.

Eichhorn, E., van der Ploeg, J.R., Kertesz, M.A., and Leisinger, T. (1997) Characterization of alpha-ketoglutarate-dependent taurine dioxygenase from Escherichia coli. *J Biol Chem* **272**: 23031-23036.

Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., and Delong, E.F. (2008) Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci U S A* **105**: 3805-3810.

Gilbert, J.A., Thomas, S., Cooley, N.A., Kulakova, A., Field, D., Booth, T. et al. (2008) Potential for phosphonoacetate utilization by marine bacteria in temperate coastal waters. *Environ Microbiol.*

Gonzalez, J.M., Covert, J.S., Whitman, W.B., Henriksen, J.R., Mayer, F., Scharf, B. et al. (2003) Silicibacter pomeroyi sp. nov. and Roseovarius nubinhibens sp. nov.,

dimethylsulfoniopropionate-demethylating bacteria from marine environments. *Int J Syst Evol Microbiol* **53**: 1261-1269.
Hogan, D.A., Smith, S.R., Saari, E.A., McCracken, J., and Hausinger, R.P. (2000) Site-directed mutagenesis of 2,4-dichlorophenoxyacetic acid/alpha-ketoglutarate dioxygenase. Identification of residues involved in metallocenter formation and substrate binding. *J Biol Chem* **275**: 12400-12409.
Hopkinson, C.S., Jr., and Vallino, J.J. (2005) Efficient export of carbon to the deep ocean through dissolved organic matter. *Nature* **433**: 142-145.
Horiguchi, M. (1984) Occurence, identification and properties of phosphonic and phosphinic acids. In *The biochemistry of natural C-P compounds*. Hoti, T., Horiguchi, M., and Hayshi, A. (eds). Kyoto, Japan: Maruzen, pp. 24-52.
Howard, E.C., Sun, S., Biers, E.J., and Moran, M.A. (2008) Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. *Environ Microbiol* **10**: 2397-2410.
Huang, J., Su, Z., and Xu, Y. (2005) The evolution of microbial phosphonate degradative pathways. *J Mol Evol* **61**: 682-690.
Ishiyama, M., Tominaga, H., Shiga, M., Sasamoto, K., Ohkura, Y., and Ueno, K. (1996) A combined assay of cell viability and in vitro cytotoxicity with a highly water-soluble tetrazolium salt, neutral red and crystal violet. *Biol Pharm Bull* **19**: 1518-1520.
Jiang, W., Metcalf, W.W., Lee, K.S., and Wanner, B.L. (1995) Molecular cloning, mapping, and regulation of Pho regulon genes for phosphonate breakdown by the phosphonatase pathway of Salmonella typhimurium LT2. *J Bacteriol* **177**: 6411-6421.
Karl, D.M. (1999) A Sea of Change: Biogeochemical variability in the North Pacific Subtropical Gyre. *Ecosystems* **2**: 181-214.
Karl, D.M., Beversdorf, L., Bjorman, K.M., Church, M.J., Martinez, A., and DeLong, E.F. (2008) Aerobic production of methane in the sea. *Nature Geosciences* **1**: 473-478.
Kim, A.D., Baker, A.S., Dunaway-Mariano, D., Metcalf, W.W., Wanner, B.L., and Martin, B.M. (2002) The 2-aminoethylphosphonate-specific transaminase of the 2-aminoethylphosphonate degradation pathway. *J Bacteriol* **184**: 4134-4140.
Kolowith, L.C., Ingall, E.D., and Benner, R. (2001) Composition and cycling of marine organic phosphorus. *Limnol. Oceanogr.* **46**: 309-320.
Kostantinidis, K., Braff, J., Karl, D.M., and DeLong, E.F. (2009) Comparative metagenomic analysis of an abyssal microbial community from 4000 m deep at Station ALOHA in the North Pacific Ocean. *Appl Environ Microbiol* **in press**.
Krzysko-Lupicka, T., Strof, W., Kubs, K., Skorupa, M., Wieczorek, P., Lejczak, B., and Kafarski, P. (1997) The ability of soil-borne fungi to degrade organophosphonate carbon-to-phosphorus bonds. *Appl Microbiol Biotechnol* **48**: 549-552.
Lee, K.S., Metcalf, W.W., and Wanner, B.L. (1992) Evidence for two phosphonate degradative pathways in Enterobacter aerogenes. *J Bacteriol* **174**: 2501-2510.
Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F., and Chisholm, S.W. (2004) Transfer of photosynthesis genes to and from Prochlorococcus viruses. *PNAS* **101**: 11013-11018.
Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar et al. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363-1371.
Marchler-Bauer, A., Anderson, J.B., Derbyshire, M.K., DeWeese-Scott, C., Gonzales, N.R., Gwadz, M. et al. (2007) CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res* **35**: D237-240.
Marden, P., Tunlid, A., Malmcrona-Friberg, K., Odham, G., and Kjelleberg, S. (1985) Physioogical and morpholgical changes during short term starvation of marine bacterial isolates. *Arch Microbiol* **142**: 326-332.
Martinez, A., Bradley, A.S., Waldbauer, J.R., Summons, R.E., and DeLong, E.F. (2007) Proteorhodopsin photosystem gene expression enables photophosphorylation in a heterologous host. *Proc Natl Acad Sci U S A* **104**: 5590-5595.

Martiny, A.C., Coleman, M.L., and Chisholm, S.W. (2006) Phosphate acquisition genes in Prochlorococcus ecotypes: evidence for genome-wide adaptation. *Proc Natl Acad Sci U S A* **103**: 12552-12557.

McGrath, J.W., Ternan, N.G., and Quinn, J.P. (1997) Utilization of organophosphonates by environmental microorganisms. *Letters in Applied Microbiology* **24**: 69-73.

McMullan, G., and Quinn, J.P. (1992) Detection of a novel carbon-phosphorus bond cleavage activity in cell-free extracts of an environmental Pseudomonas fluorescens isolate. *Biochem Biophys Res Commun* **184**: 1022-1027.

Metcalf, W.W., and Wanner, B.L. (1991) Involvement of the Escherichia coli phn (psiD) gene cluster in assimilation of phosphorus in the form of phosphonates, phosphite, Pi esters, and Pi. *J Bacteriol* **173**: 587-600.

Metcalf, W.W., and Wanner, B.L. (1993a) Evidence for a fourteen-gene, phnC to phnP locus for phosphonate metabolism in Escherichia coli. *Gene* **129**: 27-32.

Metcalf, W.W., and Wanner, B.L. (1993b) Mutational analysis of an Escherichia coli fourteen-gene operon for phosphonate degradation, using TnphoA' elements. *J Bacteriol* **175**: 3430-3442.

Morais, M.C., Zhang, W., Baker, A.S., Zhang, G., Dunaway-Mariano, D., and Allen, K.N. (2000) The crystal structure of bacillus cereus phosphonoacetaldehyde hydrolase: insight into catalysis of phosphorus bond cleavage and catalytic diversification within the HAD enzyme superfamily. *Biochemistry* **39**: 10385-10396.

Morais, M.C., Zhang, G., Zhang, W., Olsen, D.B., Dunaway-Mariano, D., and Allen, K.N. (2004) X-ray crystallographic and site-directed mutagenesis analysis of the mechanism of Schiff-base formation in phosphonoacetaldehyde hydrolase catalysis. *J Biol Chem* **279**: 9353-9361.

O'Loughlin, S.N., Graham, R.L., McMullan, G., and Ternan, N.G. (2006) A role for carbon catabolite repression in the metabolism of phosphonoacetate by Agromyces fucosus Vs2. *FEMS Microbiol Lett* **261**: 133-140.

Quinn, J.P., Kulakova, A.N., Cooley, N.A., and McGrath, J.W. (2007) New ways to break an old bond: the bacterial carbon-phosphorus hydrolases and their role in biogeochemical phosphorus cycling. *Environ Microbiol* **9**: 2392-2400.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H. et al. (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* **306**: 1344-1350.

Reisch, C.R., Moran, M.A., and Whitman, W.B. (2008) Dimethylsulfoniopropionate-dependent demethylase (DmdA) from Pelagibacter ubique and Silicibacter pomeroyi. *J Bacteriol* **190**: 8018-8024.

Rivkin, R.B., and Anderson, M.R. (1997) Inorganic nutrient limitation of oceanic bacterioplancton. *Limnol. Oceanogr.* **42**: 730-740.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: a laboratory manual*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Schofield, C.J., and McDonough, M.A. (2007) Structural and mechanistic studies on the peroxisomal oxygenase phytanoyl-CoA 2-hydroxylase (PhyH). *Biochem Soc Trans* **35**: 870-875.

Sobecky, P.A., and Hazen, T.H. (2009) Horizontal gene transfer and mobile genetic elements in marine systems. *Methods Mol Biol* **532**: 435-453.

Sowell, S.M., Wilhelm, L.J., Norbeck, A.D., Lipton, M.S., Nicora, C.D., Barofsky, D.F. et al. (2009) Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *Isme J* **3**: 93-105.

Stingl, U., Tripp, H.J., and Giovannoni, S.J. (2007) Improvements of high-throughput culturing yielded novel SAR11 strains and other abundant marine bacteria from the Oregon coast and the Bermuda Atlantic Time Series study site. *Isme J* **1**: 361-371.

Ternan, N.G., and Quinn, J.P. (1998) Phosphate starvation-independent 2-aminoethylphosphonic acid biodegradation in a newly isolated strain of Pseudomonas putida, NG2. *Syst Appl Microbiol* **21**: 346-352.

Ternan, N.G., Mc Grath, J.W., Mc Mullan, G., and Quinn, J.P. (1998) Review: organophosphonates: occurence, synthesis and biodegradation. *World Journal of Microbiology & Biotechnology* **14**: 635-647.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**: 4673-4680.

Todd, J.D., Curson, A.R., Dupont, C.L., Nicholson, P., and Johnston, A.W. (2009) The dddP gene, encoding a novel enzyme that converts dimethylsulfoniopropionate into dimethyl sulfide, is widespread in ocean metagenomes and marine bacteria and also occurs in some Ascomycete fungi. *Environ Microbiol*.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A. et al. (2004) Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science* **304**: 66-74.

White, A.K., and Metcalf, W.W. (2002) Isolation and biochemical characterization of hypophosphite/2-oxoglutarate dioxygenase. A novel phosphorus-oxidizing enzyme from Psuedomonas stutzeri WM88. *J Biol Chem* **277**: 38262-38271.

White, A.K., and Metcalf, W.W. (2007) Microbial metabolism of reduced phosphorus compounds. *Annu Rev Microbiol* **61**: 379-400.

Wu, J., Sunda, W., Boyle, E.A., and Karl, D.M. (2000) Phosphate depletion in the western North Atlantic Ocean. *Science* **289**: 759-762.

Yakovleva, G.M., Kim, S.K., and Wanner, B.L. (1998) Phosphate-independent expression of the carbon-phosphorus lyase activity of Escherichia coli. *Appl Microbiol Biotechnol* **49**: 573-578.

Yooseph, S., Sutton, G., Rusch, D.B., Halpern, A.L., Williamson, S.J., Remington, K. et al. (2007) The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol* **5**: e16.

Table 1. Abundance of *phn* genes in shotgun metagenomic libraries from Station ALOHA and the Sargasso Sea. Results expressed as % of microorganisms estimated to contain gene of interest (assuming *recA* is present in single copy in every bacterium). Expectation cutoff value $1x10^{-20}$.

	Station ALOHA					Sargasso Sea	
Gene	**HOT179_25m**	**HOT179_75m**	**HOT179_125m**	**HOT179_500m**	**HOT4000m**	**BATS216_20m**	**BATS216_50m**
HF130_AEPn_1 *phnY*	0	0	3	3	5	0	0
HF130_AEPn_1 *phnZ*	10	9	9	4	3	24	15
HF130_AEPn_2 *phnW*	0	3	3	13	63	5	9
HF130_AEPn_2 *phnX*	3	0	0	12	20	14	5
phnI	0	0	2	0	6	25	18
phnJ	0	0	0	0	4	35	17
phnA	13	3	14	20	26	7	10
recA	100	100	100	100	100	100	100
gyrB	66	51	39	45	142	114	108

Table 2. Abundance of *phn* genes in pyrosequencing metagenomic databases from Station ALOHA and the Sargasso Sea. Results expressed as % of microorganisms estimated to contain gene of interest (assuming *recA* is present in single copy in every bacterium). Bitscore (b) cutoff value b>40 except for *phnZ* (b>45) and *phnW* (b>50).

	Station ALOHA				Sargasso Sea			
Query	**HOT186_25m**	**HOT186_75m**	**HOT186_110m**	**HOT186_500m**	**BATS216_20m**	**BATS216_50m**	**BATS216_110m**	**BATS216_500m**
HF130_AEPn_1 *phnY*	0	1	3	6	1	4	7	4
HF130_AEPn_1 *phnZ*	2	3	11	22	8	10	19	10
HF130_AEPn_2 *phnW*	5	4	8	12	6	5	9	18
HF130_AEPn_2 *phnX*	0	0	1	8	9	7	5	0
phnI	1	0	0	1	20	16	5	0
phnJ	0	1	0	1	28	16	12	0
phnA	6	4	5	16	10	4	10	26
recA	100	100	100	100	100	100	100	100
gyrB	102	102	118	111	105	102	118	92

FIGURE LEGENDS

Figure 1. A. Schematic representation of the C-P lyase operon in wild type *E. coli* and the deletion strain BW16787. Genes are marked according to their assigned function as follows: putative ABC transporter components (light blue), catalytic components (dark blue), regulatory (red), and accessory proteins (orange). BW16787 has a deletion encompassing *phnG-phnP* which renders it Phn^-. **B.** Predicted C-P lyase operon in HF70_[96]11A08 (APKI441) and MPn complementation assay for individual in-frame deletions in the C-P lyase operon. Genes are marked according to their assigned functions as above. A hypothetical gene of unknown function preceding *phnM* is marked in white. For the Mpn complementation assay, strains with in-frame deletions in each of the genes encoding catalytic subunits of the C-P lyase were transformed with either HF70_[96]11A08 (APKI441) or the empty fosmid, CCFOS1, And spotted in MOPS glycerol medium containing 0.2 mM MPn as P source. HF70_[96]11A08 (APKI441) was able to complement *phnH* and *phnN* mutations for growth on Mpn.

Figure 2.A. 2-AEPn phenotypes of HF130_AEPn_1 and HF130_AEPn_2. **A.** The complementation phenotype for growth on 2-AEPn as P, as N, and as P and N source was tested in the original screening strain, BW16787 (top) and under copy up conditions in BW16787 *trfA* (bottom). CCFOS1 is shown as negative control. HF130_AEPn_1 and HF130_AEPn_2 allow growth on 2-AEPn as P source in single copy, and only HF130_AEPn_2 allows utilization of 2-AEPn as N source (marked with arrow). HF130_AEPn_130_1 appears to be toxic under copy up conditions. **B**. Pn specificity assay. BW16787 harboring CCFOS1, HF130_AEPn_1, and HF130_AEPn_2, were grown in liquid cultures with 0.2 mM of the following P sources: phosphate (Pi), 2-AEPn, MPn, phosphonoacetate (PnAc), phosphonoformate (PnF), and phosphite (Pt). Growth was measured using a WST-1 proliferation assay after 2 days at 30ºC. BW18812 (Phn^+) containing CCFOS1 was used as the positive control.

Figure 3. Genes required for 2-AEPn utilization in HF130_AEpn_2. Three genes were required for growth on 2-AEPn (marked with an asterisk): *lysR*, encoding a LysR-type transcriptional activator (blue), *phnW* encoding a 2-AEPn:pyruvate aminotransferase (red), and *phnX* encoding a 2-phosphonoacetaldehyde hydrolase or phosphonatase (green). A putative *cybB* gene (orange) in the same predicted operon was not required in *E. coli*. Gene arrangement of the phosphonatase operon in *P. fluorescens* and selected marine bacteria in shown. A *phnZ* homolog found adjacent to *phnX* and *phnW* in *Planctomyces mari*s DSM8797 is marked (dark green). The phosphonatase pathway is shown for reference.

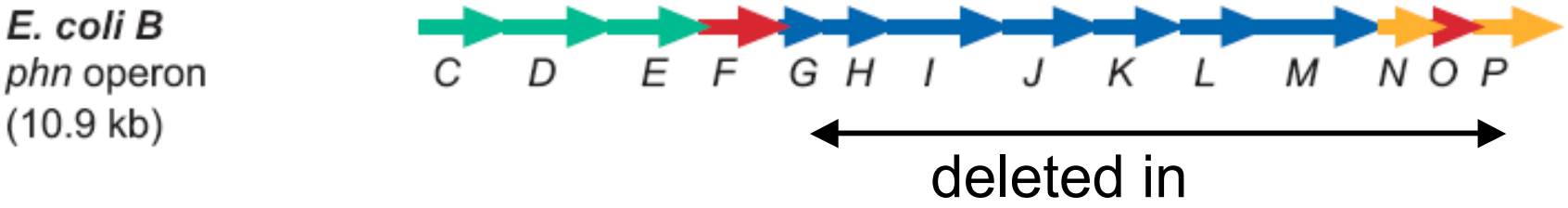
Figure 4. Genes required for 2-AEPn in HF130_AEpn_1. Two genes were required for growth on 2-AEPn (marked with an asterisk): *phnY* (pink), encoding a putative 2-oxoglutarate dioxygensase, and *phnZ* (green), encoding a protein of the HD phosphohydrolase family. Gene arrangement in other microorganisms containing similar genes is shown. The location of genes encoding a putative Pn ABC transporter in *P. marinus* MIT9303 and MIT9301 is shown in orange.

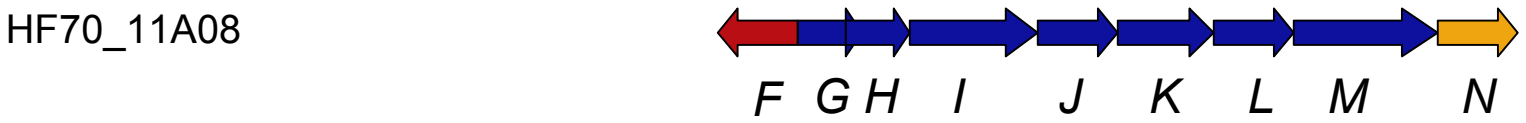
Figure 5. Growth of marine bacteria on Pns as P source. *Vibrio angustum* S14, *Ruegeria pomeyori* DSS-3, and *Planctomyces maris* DSM8797 were grown in marine synthetic medium containing 0.1 mM P source (2-AEPn, MPn, Pt or Pi as indicated). Growth was monitored measuring optical density (440 nm). *E. coli* BW16787 (Phn^-) was used as a negative control to monitor Pn stability during the course of the experiments. Addition of Pi to this strain after 5 days incubation results in detectable growth in all media.

Figure 6. Putative taxonomic affiliation of *phn* genes in metagenomic databases. Putative taxonomic affiliation was derived from the top BLAST hit against NR database for each *phn* sequence identified in the libraries. **A.** Deep water *phn* sequences in Station ALOHA for

1 the4000 m shotgun library (H4000_SG) (top) and 500 m pyrosequenced library
2 (H186_500m) (bottom). **B.** *phn* sequences in the Sargasso Sea surface waters. Data from
3 BATS216 50m pyrosequenced library. See Supplementary Tables S5 and S6 for more
4 information.
5
6

Figure 1


E. coli B
phn operon
(10.9 kb)
C D E F G H I J K L M N O P
deleted in


HF70_11A08
F G H I J K L M N

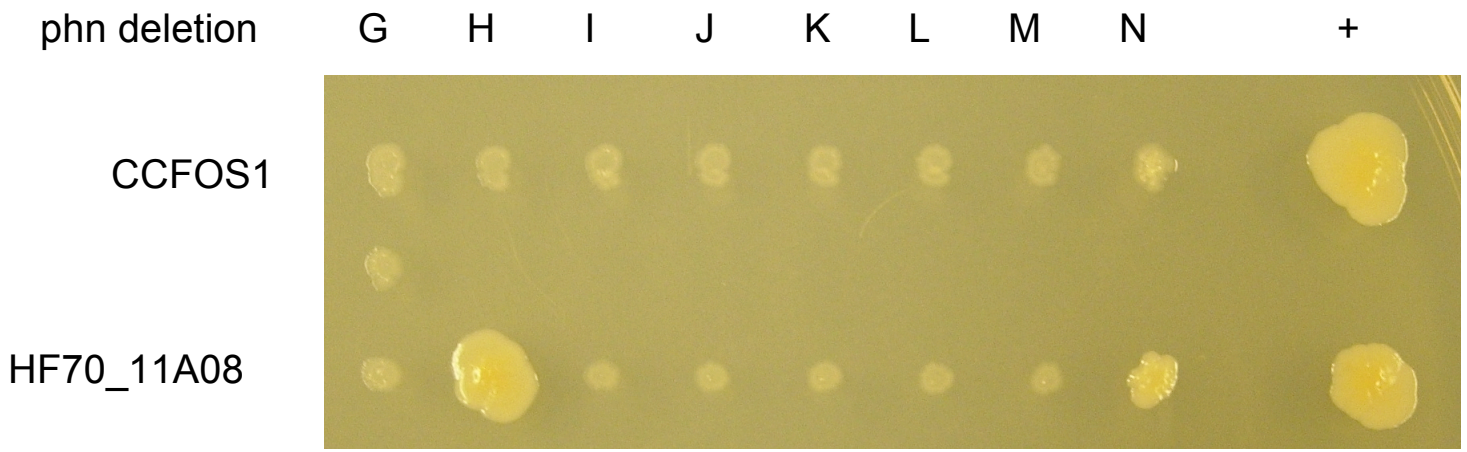
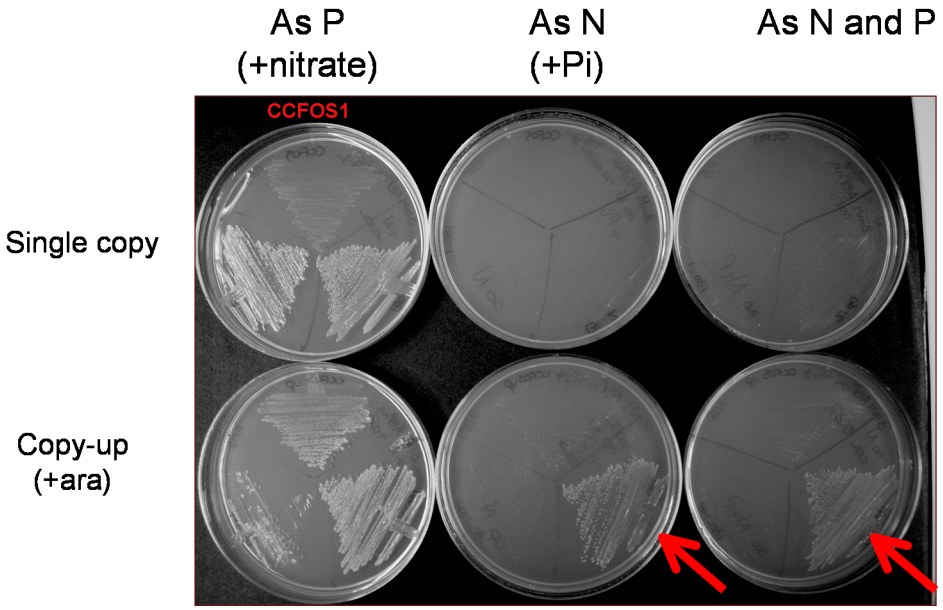

phn deletion
G H I J K L M N +
CCFOS1
HF70_11A08

Figure 2. Phenotype of complementing clones



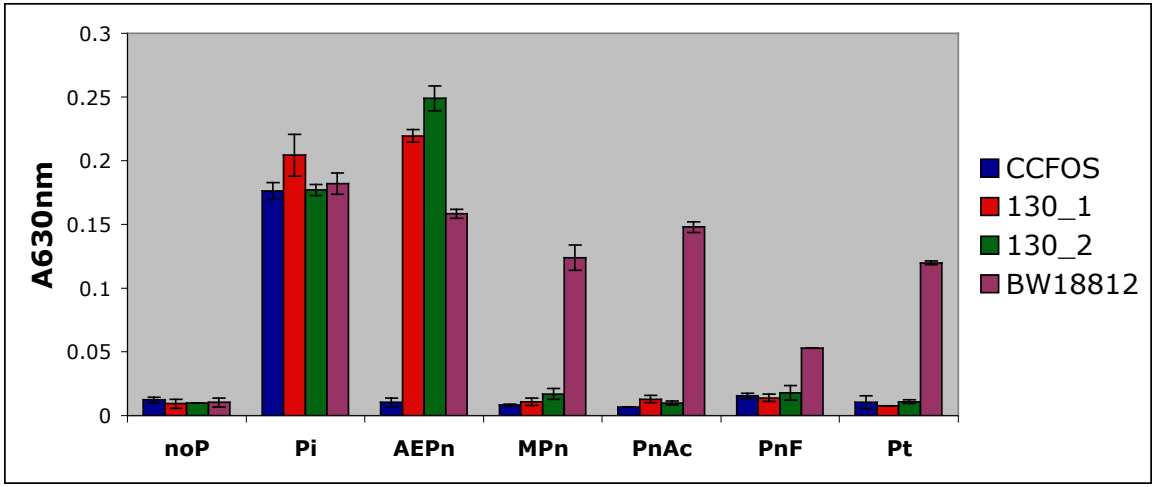
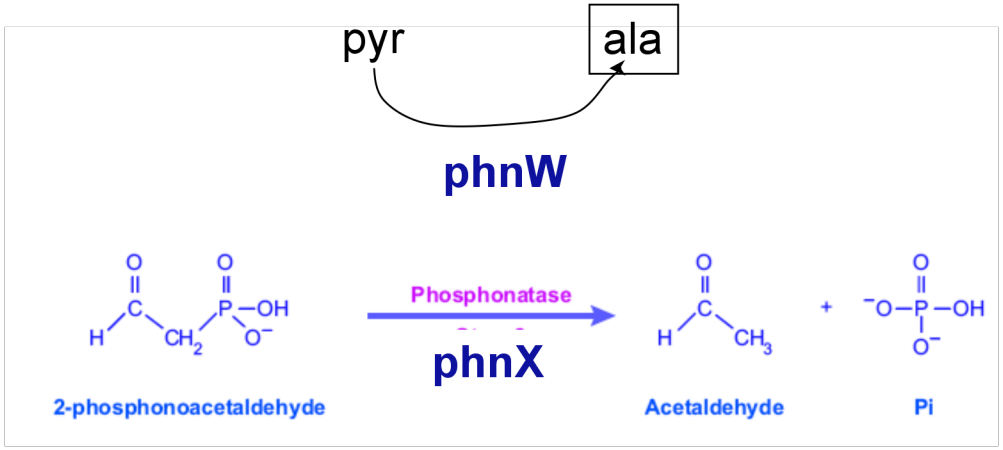
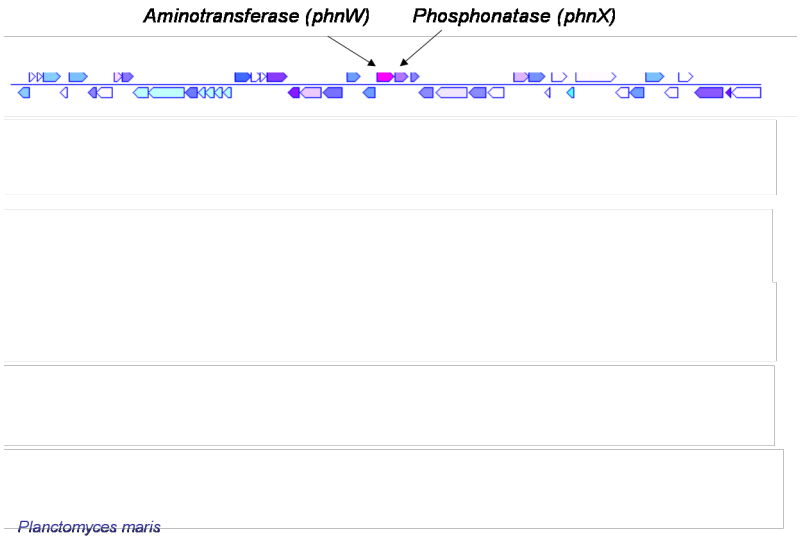


Figure 3. Note: I want to include some simpler version of the lower panel (just the operon in a few marine strains)





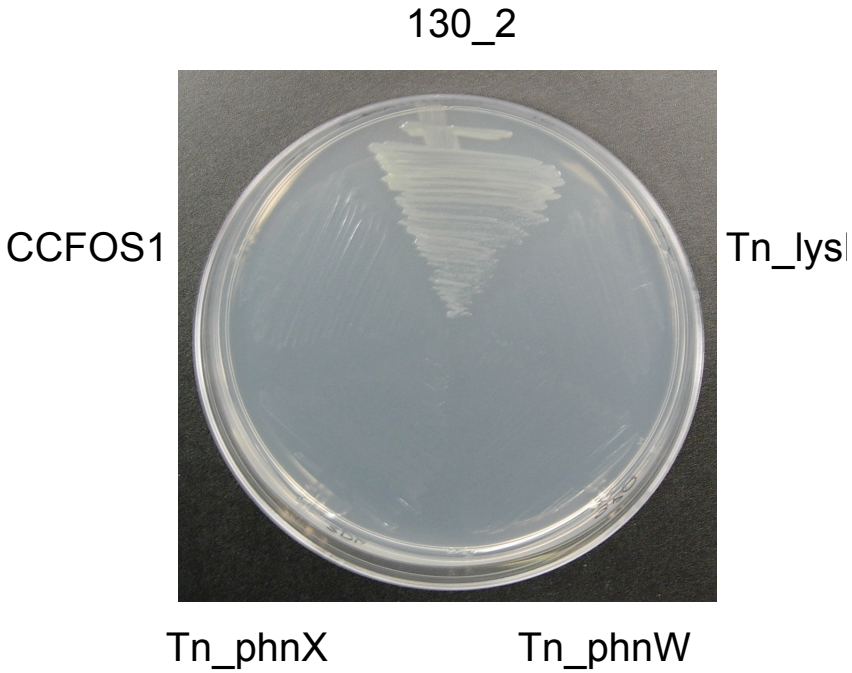
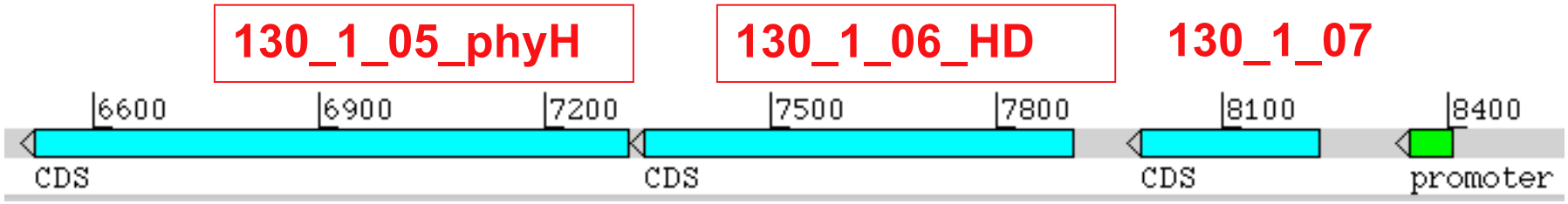
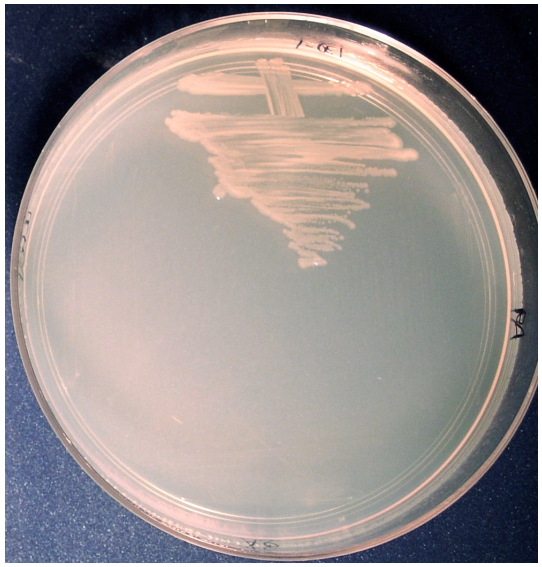


Figure 4. Required genes are boxed



130_1

CCFOS1



Tn_5_phy

Tn_6_HD

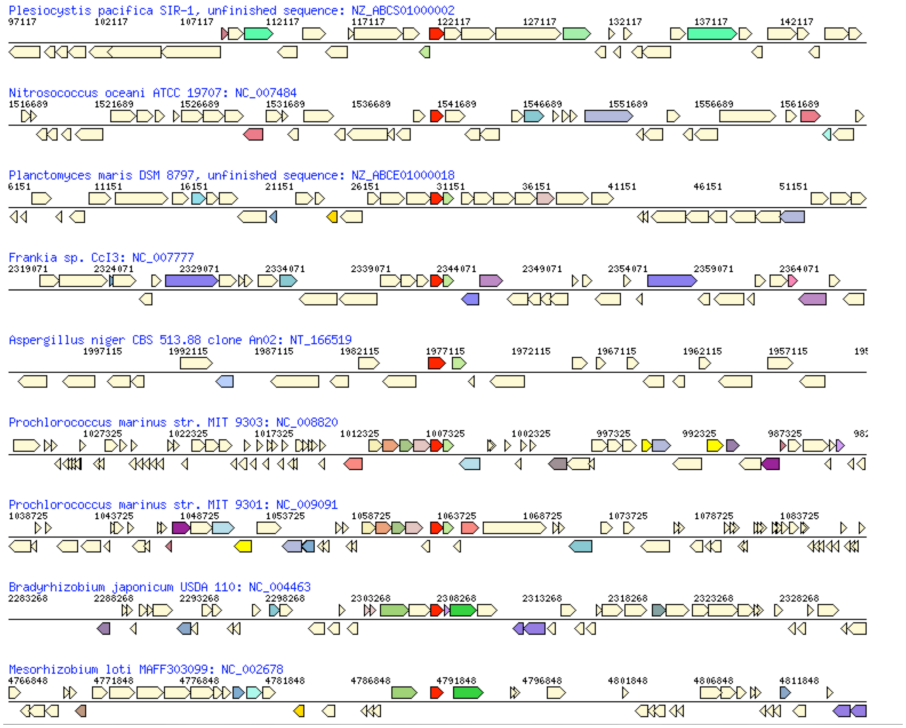


Figure 5. Growth curves on P

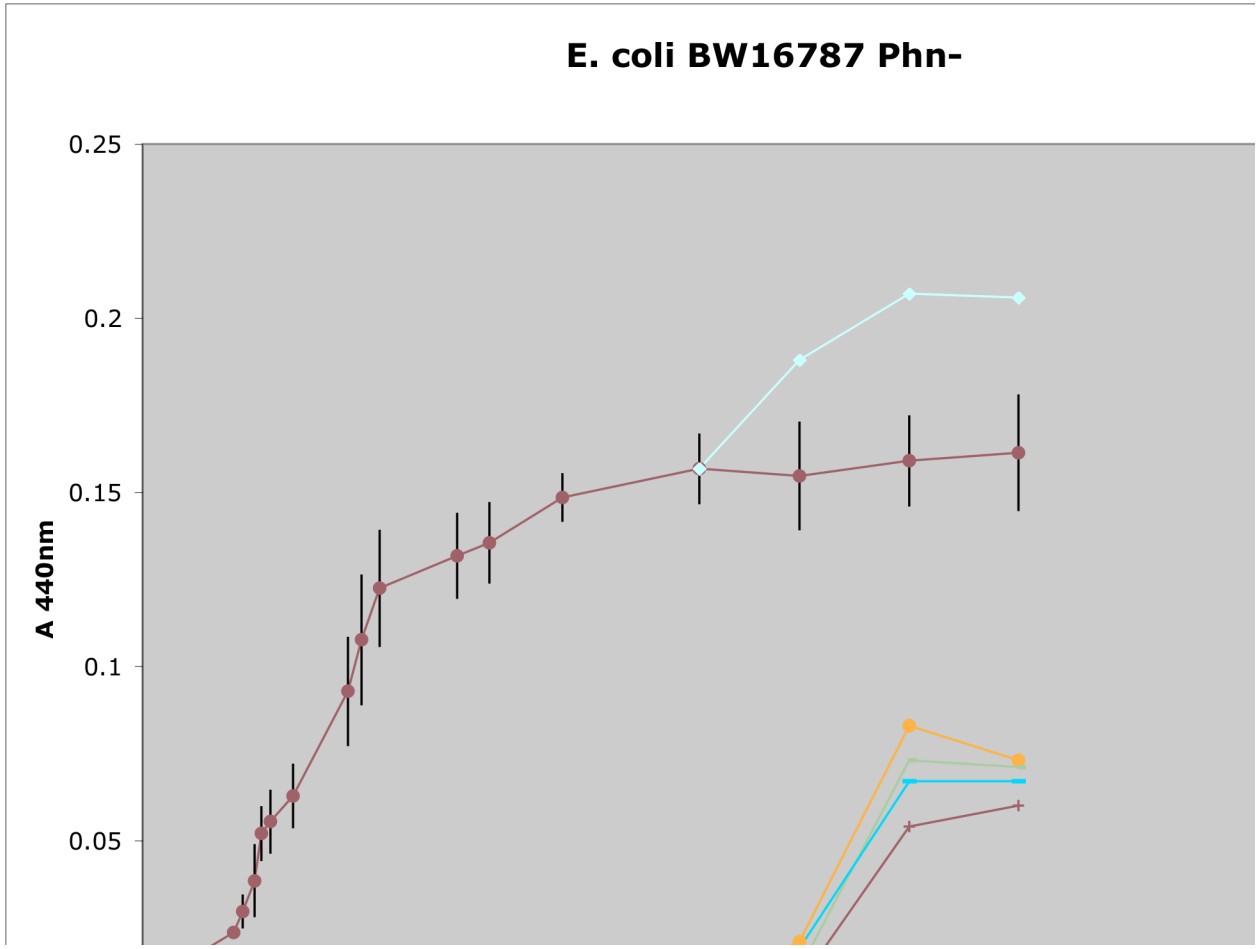


Table 1. Pn gene counts in shotgun libraries

size normalyzed % bugs recA

cutoff e-20 (only 1hsp/hit)

Gene	HOT179_25m	HOT179_75m	HOT179_125m	HOT179_500m	HF4000	HOT186_25m	BATS216_20m	BATS216_50m
130_1_5_phyH	0	0	3	3	5	4	0	0
130_1_6_HD	10	9	9	4	3	0	24	15
130_2_22_phnW	0	3	3	13	63	8	5	9
130_2_21_phnX	3	0	0	12	20	0	14	5
phnI	0	0	2	0	6	0	25	18
phnJ	0	0	0	0	4	0	35	17
phnA	13	3	14	20	26	9	7	10
recA	100	100	100	100	100	100	100	100
gyrB	66	51	39	45	142	106	114	108

Table 2. Pn gene counts in 454 libraries

size normalyzed % bugs recA

Query	HOT186_25m	HOT186_75M	HOT186_110M	HOT186_500M	BATS_20m	BATS_50m	BATS_110m	BATS_500m
130_1_5_phyH	0	1	3	6	1	4	7	4
130_1_6_HD	2	3	11	22	8	10	19	10
130_2_22_phnW	5	4	8	12	6	5	9	18
130_2_21_phnX	0	0	1	8	9	7	5	0
phnI	1	0	0	1	20	16	5	0
phnJ	0	1	0	1	28	16	12	0
PhnA	6	4	5	16	10	4	10	26
RecA	100	100	100	100	100	100	100	100
GyrB	102	102	118	111	105	102	118	92

cutoff b=40 except for HD (b=45)and phnW (b=50)