

MIT Open Access Articles

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

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Martinez, Asuncion, Gene W. Tyson, and Edward F. DeLong. "Widespread Known and Novel Phosphonate Utilization Pathways in Marine Bacteria Revealed by Functional Screening and Metagenomic Analyses." *Environmental Microbiology* 12.1 (2010): 222–238.

As Published: <http://dx.doi.org/10.1111/j.1462-2920.2009.02062.x>

Publisher: Society for Applied Microbiology and Blackwell Publishing Ltd

Persistent URL: <http://hdl.handle.net/1721.1/69553>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike 3.0



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

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

1 **SUMMARY**

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

14

15 **INTRODUCTION**

16

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

1 (HMW) DOP concentrated by tangential flow ultrafiltration (Clark et al., 1999; Kolowitz et
2 al., 2001). These studies have revealed that across oceans and throughout the water column,
3 about one third of the marine HMW DOP consists of phosphonates (Pn), reduced P
4 compounds (valence +3) that contain a C-P bond instead of the more common C-O-P bond
5 found in phosphate esters. Interestingly, phosphate ester and Pn content decreased with depth
6 in HMW DOM relative to C, indicating that both forms of P are utilized by marine
7 microorganisms (Clark et al., 1999). Benitez-Nelson and colleagues (Benitez-Nelson et al.,
8 2004) analyzed Pn concentration in sediment traps in the anoxic Cariaco Basin and found that
9 the percentage of Pns decreased from 18 to 3% relative to phosphate esters providing
10 evidence for Pn remineralization in sinking particles. These studies suggest that Pn may be a
11 significant source of P for marine organisms.

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

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

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

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

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

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

17

18 **RESULTS**

19

20 **Functional screening for Pn utilization.**

21 Seven fosmid libraries representing an open ocean depth profile at the Hawaii Ocean Time
22 Series Station ALOHA (22°45' N, 158°W) (DeLong et al., 2006) were screened for clones that
23 could complement *E. coli* BW16787 for growth on MPn or 2-AEPn as the sole P source. This
24 strain has a partial deletion in the *phn* operon encoding the C-P lyase ($\Delta phnHIJKLMNOP$)
25 which renders it incapable of growing on Pn, while still maintaining a functional Pn
26 transporter encoded by *phnCDE*, the first three genes of the operon (**Figure 1**). BW16787 was
27 previously used to successfully identify Pn genes from *Enterobacter aerogenes* and

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

5

6 **MPn utilization genes**

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

26

27 **2-AEPn utilization genes**

1 We identified two unique clones in the HF130m library that allowed BW16787 to grow on 2-
2 AEPn as P source. These clones were designated HF130_AEPn_1 and HF130_AEPn_2.
3 While both clones clearly allow BW16787 to grow on 2-AEPn as sole P source, they did not
4 allow utilization of 2-AEPn as N and C source as has been shown with several bacterial
5 isolates (McGrath et al., 1997; Ternan and Quinn, 1998) (**Figure 2A**). When the
6 complementation test was performed under copy up conditions in the new BW167877 strain, HF130_AEPn_2 allowed growth on 2-AEPn simultaneously as N and P source, and as
8 N source in the presence of Pi. When used as a P source under copy up conditions, excess Pi
9 was apparently released into the media by HF130_AEPn_2 and allowed the growth of the
10 negative control strain on the same plate. Significant release of Pi into the growth medium
11 was observed in liquid cultures of HF130_AEPn_2 grown on 2-AEPn under copy up
12 conditions (**Figure S1?**). These results indicate that the enzymes encoded in HF130_AEPn_2
13 are not inhibited by excess phosphate, but more importantly, that they permit utilization of the
14 amino group in 2-AEPn as the sole N source. HF130_AEPn_1, on the other hand, exhibited
15 poor growth on 2-AEPn as P source under copy up conditions with only a few large colonies
16 appearing over a background of microcolonies (**Figure 2A**). This growth pattern is often the
17 result of toxicity caused by high level expression of one or more genes in the fosmid,

18 The substrate specificity of the pathways encoded by HF130_AEPn_1 and
19 HF130_AEPn_2 was tested in liquid cultures using a variety of Pn compounds known to be
20 transported into the cell and cleaved by the *E. coli* C-P lyase system (Metcalf and Wanner,
21 1991, 1993b). Both clones grew on 2-AEPn but not on Mpn, phosphonoacetate (PnAc),
22 phosphonoformate or phosphite (**Figure 2B**). This narrow substrate specificity is
23 uncharacteristic of C-P lyase pathways (Quinn et al., 2007). Both clones were fully
24 sequenced and individual transposon insertions into each of the predicted ORFs were tested in
25 the complementation assay to identify the gene(s) that conferred the 2-AEPn⁺ phenotype.

26

1 **HF130_AEPn_2 encodes a typical phosphonatase pathway**

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

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

26 The phenotype of this clone, including the specificity for 2-AEPn, the capacity to use
27 2-AEPn as P and N source, and lack of inhibition by P_i , are entirely consistent with the

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

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

13

14 **HF130_AEPn_1 encodes a novel pathway for 2-AEPn utilization**

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

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

1 superfamily (pfam05721 and COG5285 with expectation values of 4×10^{-22} and 3×10^{-12} ,
2 respectively). PhyH is a 2-oxoglutarate dependent dioxygenase that catalyzes the alpha-
3 oxidation of phytanoyl-CoA coupled with the oxidative decarboxylation of 2-oxoglutarate to
4 form succinate and carbon dioxide (Schofield and McDonough, 2007). This family includes
5 eukaryotic phytanoyl-CoA dioxygenases and a number of bacterial dioxygenases mostly of
6 unknown function. Interestingly, a member of this family, HtxA, is a hypophosphite
7 dioxygenase in *Pseudomonas stutzeri* WM88 that oxidizes hypophosphite to phosphite, as
8 well as phosphite to phosphate *in vitro* (White and Metcalf, 2002). Despite this similarity,
9 HF130_AEPn_1 did not allow growth on phosphite or hypophosphite. All residues known to
10 be required for Fe²⁺ and 2-oxoglutarate binding in PhyH and related enzymes (Hogan et al.,
11 2000; Schofield and McDonough, 2007) are conserved in the predicted PhnY protein,
12 supporting its functional assignment as a 2-oxoglutarate dioxygenase.

13 *phnZ* encodes a protein of the HD superfamily (pfam01966 and COG4341, with
14 expectation values of 9×10^{-06} and 6×10^{-29} , respectively). This family is characterized by the
15 presence of an HD motif and additional conserved D and H residues believed to be involved
16 in metal binding (Aravind and Koonin, 1998). Although members of this family have been
17 predicted to act as divalent metal-dependent phosphohydrolases, none of the genes in the
18 database that share significant sequence similarity with *phnZ* have been functionally
19 characterized. In several bacterial species, however, *phnZ* homologs are found linked to
20 genes similar to other Pn degradation genes, supporting its involvement in Pn metabolism.
21 For example, it is found adjacent to C-P lyase genes in *Nostoc sp.* PCC 7120, and
22 *Pelagibacter sp.* HTCC7211, next to *phnX* in *Gemmata obscuriglobus* UQM 2246, and
23 adjacent to *phnA* in several *Burkholderia sp.* including *B. cepacia* AMMD and *B. cenocepacia*
24 HI2424 (data not shown).

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

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

10

11 **Growth of marine microbes on Pn**

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

23

24 **Abundance of 2-AEPn genes in marine metagenomic databases**

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

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

13 The distribution of Pn genes throughout the water column was investigated using
14 metagenomic data from depth profiles obtained from station ALOHA in the North Pacific
15 Gyre (from the same site as the DNA used for the fosmid libraries), and from the BATS
16 station near Bermuda in the Sargasso Sea. Analysis of depth profiles is important because the
17 fosmid screen hits came from the 130 m sample, well below the surface waters analyzed in the
18 GOS survey. Additionally, these two ecosystems are characterized by vastly different surface
19 phosphate concentrations, with DIP concentrations in the Sargasso Sea (0.2-1.0 nM) being
20 one to two orders of magnitude lower than at ALOHA station (Wu et al., 2000). The results
21 for the available shotgun libraries are shown in **Table 1**. In the Station ALOHA samples, the
22 *phnZ* is found in 10% or less of the microorganisms throughout the water column. The *phnY*
23 gene however was found in the surface shotgun libraries starting at the 125 m sample
24 suggesting that the complete pathway, as found in HF130_AEPn_1, may be more prevalent
25 below surface waters. The *phnX* and *phnW* genes are found in low frequency in the shotgun
26 libraries from the top 125 m of the water column in Station ALOHA, but their frequency
27 increased substantially with depth. In the 4000 m library 63% and 20% of the microorganisms

1 were estimated to have a *phnW* and *phnX* gene, respectively. The *phnA* gene was also found
2 at high frequency in the deep water samples (present in up to 26% of microorganisms). These
3 results were unexpected since the DIP concentration is significantly lower in the upper 100 m
4 of the water column at this location ((Bjorkman and Karl, 2003). The analysis of the putative
5 phylogenetic affiliation of these deep water *phn* genes (**Figure 6 and Table S5**) revealed that
6 many diverse bacterial groups contribute to the *phnA* and *phnW* signals. The *phnX* signal at
7 4000 m was assigned to Alphaproteobacteria and Eukaryotes based on the top blast hit against
8 the NCBI non-redundant protein database (NR). Although the putative proteins were most
9 similar to eukaryotic proteins, the encoding genes in our libraries did not contain introns
10 suggesting that they are probably bacterial in origin.

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

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

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

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

24

25 **DISCUSSION**

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

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

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

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

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

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

18 The *phnY/phnZ* pathway is sporadically found in the genomes of diverse bacterial taxa,
19 including Cyanobacteria, Planctomycetes, and Proteobacteria, and in several fungi. The
20 presence of the pathway in discrete members of distantly related bacteria instead of being
21 lineage specific, suggest that it might have been subject to horizontal gene transfer (HGT) and
22 that there is a strong selection for the ability to utilize Pn in the environment. Similar
23 conclusions had been reached in analyses of the C-P lyase and phosphonate pathway
24 (Huang et al., 2005). The presence of a *phnZ* gene in the genome of a mimivirus (Raoult et al.,
25 2004) suggests that viruses might be a vehicle for HGT, as has been seen for other host genes
26 (Lindell et al., 2004; Sobecky and Hazen, 2009). Although HGT between bacteria and fungi

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

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

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

1 an alphaproteobacterium represented in these data, is capable of using 2-AEPn and MPn as a
2 P source providing direct evidence for Pn utilization in this group. We also found C-P lyase
3 genes of two cyanobacterial species, *Cyanothece* sp. PCC 8802 and *Synechococcus* sp. PCC
4 7335, suggesting that they might be able to use Pn as P source, similar to the filamentous
5 cyanobacterium *Trichodesmium* (Dyhrman et al., 2006). Finally, there is high representation
6 of *phnA*, which is involved in phosphonoacetate degradation, and *phnX* and *phnW* genes
7 (phosphonatase pathway) in the Sargasso Sea surface waters. This signal appears to originate
8 in large part from alphaproteobacterial members of the community, but also includes
9 *Pseudomonas*-like sequences similar to those identified in this study. These results provide
10 strong evidence that Pn use is widespread among the microbial community that thrives in P
11 limited waters of the Western North Atlantic. Although no expression data for the Sargasso
12 Sea depth profile analyzed here is yet available, we found hits to *phnY*, *phnZ*, *phnA*, *phnI* and
13 *phnJ* in metatranscriptomic data from the Sargasso Sea (M. A. Moran, unpublished) and other
14 ocean regions (J. Zehr, unpublished). In addition, expression of C-P lyase genes from
15 *Trichodesmium* (Dyhrman et al., 2006) and the periplasmic component of the Pn ABC
16 transporter, PhnD from *Pelagibacter ubique* HTCC7211 (Sowell et al., 2009) have been
17 observed in the Sargasso Sea supporting the conclusion that Pn utilization genes are expressed
18 in this environment.

19 At both stations, the frequency of some Pn utilization genes was high in deep water
20 samples, even though the measured DIP concentrations are significantly higher than in surface
21 waters. In the HOT samples the increase in *phn* gene frequency was most evident below the
22 100 m mark (the depth where the measured DIP concentration starts to increase in station
23 ALOHA (Bjorkman and Karl, 2003)) and continued all the way to the 4000 m sample. The
24 frequency of *phnX*, *phnW*, and *phnA* genes was particularly high and appeared to be derived
25 from representatives of numerous diverse bacterial groups. One possible explanation for Pn
26 utilization in the presence of high concentrations of DIP comes from the fact that the most
27 abundant Pn genes in the deep belong to Pn utilization pathways that allow their use not only

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

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

24

25 **EXPERIMENTAL PROCEDURES**

26 **Chemicals, media, and bacterial strains.**

1 Methylphosphonate (MPn), 2-aminoethylphosphonate (2-AEPn), phosphonoformate
2 (PnF), phosphonoacetate (PnAc), and phosphite (Pt) were from Sigma Aldrich. MOPS
3 minimal medium and MOPS minimal medium without NH₄ were purchased from Tecknova
4 (Hollister, CA). Phosphate Colorimetric Assay kit was from BioVision, Inc. (Mountain
5 View, CA). *E. coli* BW16787 and BW18812 were obtained from B. Wanner through the *E.*
6 *Coli* Genetic Stock Center. *Ruegeria pomeyori* DSS-3, *Vibrio angustum* S14, and
7 *Planctomyces maris* DSM8797 were obtained from M.A. Moran, S. Kjelleberg, and ATCC,
8 respectively.

9

10 **Screening of metagenomic libraries for Pn utilization.**

11 Marine picoplankton fosmid libraries were previously constructed from samples
12 collected along a depth profile from the Hawaii Ocean Time series (HOT) station ALOHA
13 (22°45' N, 158°W) at cloned into the copy-control pCC1FOS fosmid vector (Epicentre)
14 (DeLong et al., 2006). These libraries were pooled and fosmid DNA was isolated by alkaline
15 lysis followed by cesium chloride ultracentrifugation (Sambrook et al., 1989). Aliquots of
16 the pooled fosmid DNA were used to transform BW16787 by electroporation (1.2 Kv/cm,
17 200 Ohms, 25 μF). After electroporation, cells were incubated at 37°C in 1 ml SOC to allow
18 for antibiotic resistance expression, and rinsed twice in 5 ml of MOPS buffer to remove
19 phosphate prior to plating in selective medium. Plating media consisted of MOPS minimal
20 medium with 0.4% glycerol, 12 μg/ml chloramphenicol, and 0.1 mM MPn or 2-AEPn. A
21 small fraction of each transformation was also plated in minimal medium containing Pi to
22 evaluate the transformation efficiency and estimate library coverage. Screening plates were
23 incubated at 30°C for up to 20 days. The empty fosmid vector, CCFOS1, was used as a
24 negative control. Negative control and screening plates always contained a large number of
25 microcolonies that arose from the utilization of the residual Pi in this medium (Metcalf and
26 Wanner, 1993b). Only large colonies clearly distinguishable from that background were
27 chosen for further characterization. Phosphonate positive clones were restreaked on selective

1 medium and fosmid DNA was isolated and retransformed into BW16787 to confirm the
2 phenotype. Restriction analysis and fosmid end sequencing were used to identify replicates.
3 Unique fosmids were sequenced using transposon mutagenesis as described (Martinez et al.,
4 2007). The complete DNA sequence was assembled using Sequencher v. 4.5 (Gene Codes
5 Corporation) and annotated with FGENESB (Softberry) and Artemis v. 6 (The Wellcome
6 Trust Sanger Institute). At least one clone containing a transposon insertion in each
7 predicted ORF was transformed back into BW16787 and screened for growth on Pn as
8 above.

9

10 **Growth on 2-AEPn as N source.**

11 To evaluate the ability of library clones to grow on 2-AEPn as N source, clones were
12 streaked on NH₄-free MOPS minimal medium containing 5mM 2-AEPn with or without
13 addition of 0.1 mM Pi. Clones were tested in the original BW16787 strain as well as in a
14 new copy-up derivative, BW16787 *trfA*, created by P1 transduction of the copy-up P_{BAD-*trfA*}
15 marker from the Translator strain (Lucigen Corp., Middleton, WI). For copy-up conditions,
16 0.2% arabinose was added to the growth medium.

17

18 **Complementation analysis of individual C-P lyase genes.**

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

23

24 **Pn specificity tests for fosmid library clones.**

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

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

6

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

8 Bacterial isolates to be tested for growth on Pn as P source were grown for 48h in a
9 modified VNSS medium (0.1xVNSS, 0.1 g peptone, 0.05 g yeast extract, 0.05 g glucose,
10 0.05 g starch, 1mg FeSO₄·7H₂O, 1 mg NaH₂PO₄ per liter of nine salts solution (NSS)
11 (Marden et al., 1985)) at room temperature, rinsed twice in P-free NSS, and resuspended in 1
12 volume equivalent of NSS. 10 µl of inoculum were added in triplicate to microtiter plate
13 wells containing 160 µl of growth medium (NSS with 4g Glucose, 2.2 g (NH₄)₂SO₄, and 1 ml
14 vitamin solution (Gonzalez et al., 2003) per liter) with 0.1 mM P source as indicated.
15 Growth was monitored by measuring optical density (440 nm) in a Biotek Synergy2 plate
16 reader. The stability of the Pn compounds during the course of the experiment was
17 monitored using a colorimetric phosphate assay (BioVision Inc., Mountain View, CA) in an
18 uninoculated plate incubated under the same conditions. In addition, *E. coli* BW16787
19 (*Δphn*), which cannot utilize reduced P compounds, was used as negative control. Viability
20 of the negative control strain in the Pn media during the entire course of the experiment was
21 tested by showing that growth could take place upon Pi addition after 5 days of incubation.
22 For strains that only reached low optical density, cell proliferation was also measured using
23 the WST-1 viability assay (Roche Applied Sciences, Indianapolis, IN).

24

25 **Sample collection, DNA extraction and pyrosequencing.**

26 Bacterioplankton samples were obtained from Hawaii Ocean Time-series Station
27 ALOHA in the North Pacific Subtropical Gyre (22°44'N, 158°2'W) and BATS Station in the

1 Sargasso Sea (31°40'N, 64°10'W). At each site, bacterioplankton samples were collected
2 from the photic zone at the mixed layer, just below the mixed layer, and at the deep
3 chlorophyll maximum (25, 75, and 125 m for HOT179, 25, 75 and 110m for HOT186, and
4 25, 50, and 100m for BATS216), and the mesopelagic zone (500m). The H4000m shotgun
5 library has been previously described (Kostantinidis et al., 2009). Sample collection and
6 DNA extraction were performed as previously described (Frias-Lopez et al., 2008). 3-5 µg
7 of community DNA were sequenced using Genome Sequencer FLX (Roche).

8

9 **Bioinformatics.**

10 Conserved domains were identified using CDD (Marchler-Bauer et al., 2007).
11 Abundance and distribution of Pn genes in the databases was performed as follows. Deduced
12 peptide sequences of the Pn genes identified in the screens, PhnI (4080494), PhnJ
13 (40804945) from *Pseudomonas stutzeri*, PhnA (1196755) from *P. fluorescens*, and *E. coli*
14 RecA and GyrB as single copy reference were used as query to interrogate available
15 databases using NCBI Blast. An expectation cutoff value of 1×10^{-20} was used for NCBI NR,
16 GOS, HOTS and BATS shotgun data. For shorter pyrosequencing reads, a bit cutoff value of
17 40 was used initially (DeLong et al., 2006; Yooseph et al., 2007; Frias-Lopez et al., 2008) but
18 was verified for each gene by comparing the pyrosequencing hits against NR and selecting
19 cutoff values that returned as best blast hit only proteins identified as members of the same
20 family as the query gene by annotation or phylogenetic tree analyses (bellow). Based on
21 these analyses, bit score cutoff of 40 was used for all genes except for *phnZ* and *phnW* (b>45
22 and b>50, respectively). Gene counts were size normalized using the query protein length.
23 The percentage of microbes containing each gene was calculated assuming that *recA* is
24 present in single copy in every microbial genome (Howard et al., 2008; Reisch et al., 2008).

25

26 **Phylogenetic Analyses**

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

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

14

15 **ACKNOWLEDGEMENTS**

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

26

1 REFERENCES

- 2 Altschul, S.F., Madden, T.L., Schaeffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman,
3 D.J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database
4 search programs. *Nucl. Acids Res.* **25**: 3389-3402.
- 5 Aravind, L., and Koonin, E.V. (1998) The HD domain defines a new superfamily of metal-
6 dependent phosphohydrolases. *Trends Biochem Sci* **23**: 469-472.
- 7 Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M. et al. (2006)
8 Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio
9 collection. *Mol Syst Biol* **2**: 2006 0008.
- 10 Benitez-Nelson, C.R., O'Neill, L., Kolowith, L.C., Pellecia, P., and Thunell, R. (2004)
11 Phosphonates and particulate organic phosphorus cycling in an anoxic marine basin.
12 *Limnol. Oceanogr.* **49**: 1593-1604.
- 13 Bjorkman, K.M., and Karl, D.M. (2003) Bioavailability of dissolved organic phosphorus in
14 the euphotic zone at Station ALOHA, North Pacific Subtropical Gyre. *Limnol. Oceanogr.*
15 **48**: 1049-1057.
- 16 Chen, C.C., Zhang, H., Kim, A.D., Howard, A., Sheldrick, G.M., Mariano-Dunaway, D., and
17 Herzberg, O. (2002) Degradation pathway of the phosphonate ciliatine: crystal structure
18 of 2-aminoethylphosphonate transaminase. *Biochemistry* **41**: 13162-13169.
- 19 Clark, L.L., Ingall, E.D., and Benner, R. (1999) Marine organic phosphorus cycling: novel
20 insights from nuclear magnetic resonance. *Am J Science* **2999**: 724-737.
- 21 Coleman, M.L., Sullivan, M.B., Martiny, A.C., Steglich, C., Barry, K., Delong, E.F., and
22 Chisholm, S.W. (2006) Genomic islands and the ecology and evolution of
23 Prochlorococcus. *Science* **311**: 1768-1770.
- 24 Datsenko, K.A., and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in
25 Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**: 6640-6645.
- 26 DeLong, E.F. (2009) Microbial oceanography from genomes to biomes. *Nature* **459**.
- 27 DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.-U. et al. (2006)
28 Community Genomics Among Stratified Microbial Assemblages in the Ocean's Interior.
29 *Science* **311**: 496-503.
- 30 Dumora, C., Lacoste, A.M., and Cassaigne, A. (1983) Purification and properties of 2-
31 aminoethylphosphonate:pyruvate aminotransferase from Pseudomonas aeruginosa. *Eur J*
32 *Biochem* **133**: 119-125.
- 33 Dumora, C., Marche, M., Doignon, F., Aigle, M., Cassaigne, A., and Crouzet, M. (1997)
34 First characterization of the phosphonoacetaldehyde hydrolase gene of Pseudomonas
35 aeruginosa. *Gene* **197**: 405-412.
- 36 Dyhrman, S.T., Ammerman, J.W., and Van Moy, B.A.S. (2007) Microbes and the marine
37 phosphorus cycle. *Oceanography* **20**: 110-116.
- 38 Dyhrman, S.T., Chappell, P.D., Haley, S.T., Moffett, J.W., Orchard, E.D., Waterbury, J.B.,
39 and Webb, E.A. (2006) Phosphonate utilization by the globally important marine
40 diazotroph Trichodesmium. **439**: 68-71.
- 41 Egli, T. (1988) (An)aerobic breakdown of chelating agents used in household detergents.
42 *Microbiol Sci* **5**: 36-41.
- 43 Eichhorn, E., van der Ploeg, J.R., Kertesz, M.A., and Leisinger, T. (1997) Characterization of
44 alpha-ketoglutarate-dependent taurine dioxygenase from Escherichia coli. *J Biol Chem*
45 **272**: 23031-23036.
- 46 Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., and
47 Delong, E.F. (2008) Microbial community gene expression in ocean surface waters. *Proc*
48 *Natl Acad Sci U S A* **105**: 3805-3810.
- 49 Gilbert, J.A., Thomas, S., Cooley, N.A., Kulakova, A., Field, D., Booth, T. et al. (2008)
50 Potential for phosphonoacetate utilization by marine bacteria in temperate coastal waters.
51 *Environ Microbiol.*
- 52 Gonzalez, J.M., Covert, J.S., Whitman, W.B., Henriksen, J.R., Mayer, F., Scharf, B. et al.
53 (2003) *Silicibacter pomeroyi* sp. nov. and *Roseovarius nubinihibens* sp. nov.,

- 1 dimethylsulfoniopropionate-demethylating bacteria from marine environments. *Int J Syst*
2 *Evol Microbiol* **53**: 1261-1269.
- 3 Hogan, D.A., Smith, S.R., Saari, E.A., McCracken, J., and Hausinger, R.P. (2000) Site-
4 directed mutagenesis of 2,4-dichlorophenoxyacetic acid/alpha-ketoglutarate dioxygenase.
5 Identification of residues involved in metallocenter formation and substrate binding. *J*
6 *Biol Chem* **275**: 12400-12409.
- 7 Hopkinson, C.S., Jr., and Vallino, J.J. (2005) Efficient export of carbon to the deep ocean
8 through dissolved organic matter. *Nature* **433**: 142-145.
- 9 Horiguchi, M. (1984) Occurrence, identification and properties of phosphonic and phosphinic
10 acids. In *The biochemistry of natural C-P compounds*. Hoti, T., Horiguchi, M., and
11 Hayashi, A. (eds). Kyoto, Japan: Maruzen, pp. 24-52.
- 12 Howard, E.C., Sun, S., Biers, E.J., and Moran, M.A. (2008) Abundant and diverse bacteria
13 involved in DMSP degradation in marine surface waters. *Environ Microbiol* **10**: 2397-
14 2410.
- 15 Huang, J., Su, Z., and Xu, Y. (2005) The evolution of microbial phosphonate degradative
16 pathways. *J Mol Evol* **61**: 682-690.
- 17 Ishiyama, M., Tominaga, H., Shiga, M., Sasamoto, K., Ohkura, Y., and Ueno, K. (1996) A
18 combined assay of cell viability and in vitro cytotoxicity with a highly water-soluble
19 tetrazolium salt, neutral red and crystal violet. *Biol Pharm Bull* **19**: 1518-1520.
- 20 Jiang, W., Metcalf, W.W., Lee, K.S., and Wanner, B.L. (1995) Molecular cloning, mapping,
21 and regulation of Pho regulon genes for phosphonate breakdown by the phosphonate
22 pathway of *Salmonella typhimurium* LT2. *J Bacteriol* **177**: 6411-6421.
- 23 Karl, D.M. (1999) A Sea of Change: Biogeochemical variability in the North Pacific
24 Subtropical Gyre. *Ecosystems* **2**: 181-214.
- 25 Karl, D.M., Beversdorf, L., Bjorman, K.M., Church, M.J., Martinez, A., and DeLong, E.F.
26 (2008) Aerobic production of methane in the sea. *Nature Geosciences* **1**: 473-478.
- 27 Kim, A.D., Baker, A.S., Dunaway-Mariano, D., Metcalf, W.W., Wanner, B.L., and Martin,
28 B.M. (2002) The 2-aminoethylphosphonate-specific transaminase of the 2-
29 aminoethylphosphonate degradation pathway. *J Bacteriol* **184**: 4134-4140.
- 30 Kolowith, L.C., Ingall, E.D., and Benner, R. (2001) Composition and cycling of marine
31 organic phosphorus. *Limnol. Oceanogr.* **46**: 309-320.
- 32 Kostantinidis, K., Braff, J., Karl, D.M., and DeLong, E.F. (2009) Comparative metagenomic
33 analysis of an abyssal microbial community from 4000 m deep at Station ALOHA in the
34 North Pacific Ocean. *Appl Environ Microbiol* **in press**.
- 35 Krzysko-Lupicka, T., Strof, W., Kubs, K., Skorupa, M., Wieczorek, P., Lejczak, B., and
36 Kafarski, P. (1997) The ability of soil-borne fungi to degrade organophosphonate carbon-
37 to-phosphorus bonds. *Appl Microbiol Biotechnol* **48**: 549-552.
- 38 Lee, K.S., Metcalf, W.W., and Wanner, B.L. (1992) Evidence for two phosphonate
39 degradative pathways in *Enterobacter aerogenes*. *J Bacteriol* **174**: 2501-2510.
- 40 Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F., and Chisholm, S.W.
41 (2004) Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *PNAS* **101**:
42 11013-11018.
- 43 Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar et al. (2004)
44 ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363-1371.
- 45 Marchler-Bauer, A., Anderson, J.B., Derbyshire, M.K., DeWeese-Scott, C., Gonzales, N.R.,
46 Gwadz, M. et al. (2007) CDD: a conserved domain database for interactive domain
47 family analysis. *Nucleic Acids Res* **35**: D237-240.
- 48 Marden, P., Tunlid, A., Malmcrona-Friberg, K., Odham, G., and Kjelleberg, S. (1985)
49 Physiological and morphological changes during short term starvation of marine bacterial
50 isolates. *Arch Microbiol* **142**: 326-332.
- 51 Martinez, A., Bradley, A.S., Waldbauer, J.R., Summons, R.E., and DeLong, E.F. (2007)
52 Proteorhodopsin photosystem gene expression enables photophosphorylation in a
53 heterologous host. *Proc Natl Acad Sci U S A* **104**: 5590-5595.

- 1 Martiny, A.C., Coleman, M.L., and Chisholm, S.W. (2006) Phosphate acquisition genes in
2 Prochlorococcus ecotypes: evidence for genome-wide adaptation. *Proc Natl Acad Sci U S*
3 *A* **103**: 12552-12557.
- 4 McGrath, J.W., Ternan, N.G., and Quinn, J.P. (1997) Utilization of organophosphonates by
5 environmental microorganisms. *Letters in Applied Microbiology* **24**: 69-73.
- 6 McMullan, G., and Quinn, J.P. (1992) Detection of a novel carbon-phosphorus bond
7 cleavage activity in cell-free extracts of an environmental *Pseudomonas fluorescens*
8 isolate. *Biochem Biophys Res Commun* **184**: 1022-1027.
- 9 Metcalf, W.W., and Wanner, B.L. (1991) Involvement of the *Escherichia coli* *phn* (*psiD*)
10 gene cluster in assimilation of phosphorus in the form of phosphonates, phosphite, Pi
11 esters, and Pi. *J Bacteriol* **173**: 587-600.
- 12 Metcalf, W.W., and Wanner, B.L. (1993a) Evidence for a fourteen-gene, *phnC* to *phnP* locus
13 for phosphonate metabolism in *Escherichia coli*. *Gene* **129**: 27-32.
- 14 Metcalf, W.W., and Wanner, B.L. (1993b) Mutational analysis of an *Escherichia coli*
15 fourteen-gene operon for phosphonate degradation, using *TnphoA'* elements. *J Bacteriol*
16 **175**: 3430-3442.
- 17 Morais, M.C., Zhang, W., Baker, A.S., Zhang, G., Dunaway-Mariano, D., and Allen, K.N.
18 (2000) The crystal structure of *Bacillus cereus* phosphonoacetaldehyde hydrolase: insight
19 into catalysis of phosphorus bond cleavage and catalytic diversification within the HAD
20 enzyme superfamily. *Biochemistry* **39**: 10385-10396.
- 21 Morais, M.C., Zhang, G., Zhang, W., Olsen, D.B., Dunaway-Mariano, D., and Allen, K.N.
22 (2004) X-ray crystallographic and site-directed mutagenesis analysis of the mechanism of
23 Schiff-base formation in phosphonoacetaldehyde hydrolase catalysis. *J Biol Chem* **279**:
24 9353-9361.
- 25 O'Loughlin, S.N., Graham, R.L., McMullan, G., and Ternan, N.G. (2006) A role for carbon
26 catabolite repression in the metabolism of phosphonoacetate by *Agromyces fucus* Vs2.
27 *FEMS Microbiol Lett* **261**: 133-140.
- 28 Quinn, J.P., Kulakova, A.N., Cooley, N.A., and McGrath, J.W. (2007) New ways to break an
29 old bond: the bacterial carbon-phosphorus hydrolases and their role in biogeochemical
30 phosphorus cycling. *Environ Microbiol* **9**: 2392-2400.
- 31 Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H. et al. (2004) The 1.2-
32 megabase genome sequence of Mimivirus. *Science* **306**: 1344-1350.
- 33 Reisch, C.R., Moran, M.A., and Whitman, W.B. (2008) Dimethylsulfoniopropionate-
34 dependent demethylase (*DmdA*) from *Pelagibacter ubique* and *Silicibacter pomeroyi*. *J*
35 *Bacteriol* **190**: 8018-8024.
- 36 Rivkin, R.B., and Anderson, M.R. (1997) Inorganic nutrient limitation of oceanic
37 bacterioplankton. *Limnol. Oceanogr.* **42**: 730-740.
- 38 Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: a laboratory*
39 *manual*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.
- 40 Schofield, C.J., and McDonough, M.A. (2007) Structural and mechanistic studies on the
41 peroxisomal oxygenase phytanoyl-CoA 2-hydroxylase (*PhyH*). *Biochem Soc Trans* **35**:
42 870-875.
- 43 Sobecky, P.A., and Hazen, T.H. (2009) Horizontal gene transfer and mobile genetic elements
44 in marine systems. *Methods Mol Biol* **532**: 435-453.
- 45 Sowell, S.M., Wilhelm, L.J., Norbeck, A.D., Lipton, M.S., Nicora, C.D., Barofsky, D.F. et al.
46 (2009) Transport functions dominate the SAR11 metaproteome at low-nutrient extremes
47 in the Sargasso Sea. *Isme J* **3**: 93-105.
- 48 Stingl, U., Tripp, H.J., and Giovannoni, S.J. (2007) Improvements of high-throughput
49 culturing yielded novel SAR11 strains and other abundant marine bacteria from the
50 Oregon coast and the Bermuda Atlantic Time Series study site. *Isme J* **1**: 361-371.
- 51 Ternan, N.G., and Quinn, J.P. (1998) Phosphate starvation-independent 2-
52 aminoethylphosphonic acid biodegradation in a newly isolated strain of *Pseudomonas*
53 *putida*, NG2. *Syst Appl Microbiol* **21**: 346-352.

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

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

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

11 Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A. et al.
12 (2004) Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science* **304**:
13 66-74.

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

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

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

21 Yakovleva, G.M., Kim, S.K., and Wanner, B.L. (1998) Phosphate-independent expression of
22 the carbon-phosphorus lyase activity of *Escherichia coli*. *Appl Microbiol Biotechnol* **49**:
23 573-578.

24 Yooseph, S., Sutton, G., Rusch, D.B., Halpern, A.L., Williamson, S.J., Remington, K. et al.
25 (2007) The Sorcerer II Global Ocean Sampling expedition: expanding the universe of
26 protein families. *PLoS Biol* **5**: e16.

27
28
29
30

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 1×10^{-20} .

Gene	Station ALOHA					Sargasso Sea	
	HOT179_25m	HOT179_75m	HOT179_125m	HOT179_500m	HOT4000m	BATS216_20m	BATS216_50m
HF130_AEPn_1 <i>phnY</i>	0	0	3	3	5	0	0
HF130_AEPn_1 <i>phnZ</i>	10	9	9	4	3	24	15
HF130_AEPn_2 <i>phnW</i>	0	3	3	13	63	5	9
HF130_AEPn_2 <i>phnX</i>	3	0	0	12	20	14	5
<i>phnI</i>	0	0	2	0	6	25	18
<i>phnJ</i>	0	0	0	0	4	35	17
<i>phnA</i>	13	3	14	20	26	7	10
<i>recA</i>	100	100	100	100	100	100	100
<i>gyrB</i>	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).

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

1 **FIGURE LEGENDS**

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

15

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

1

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

10

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

17

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

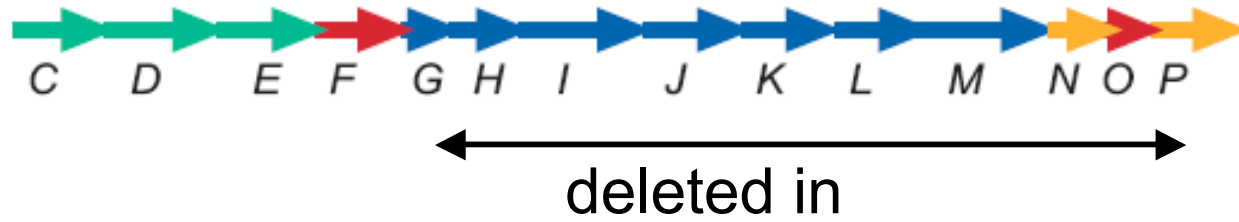
24

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

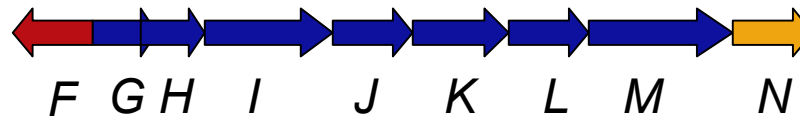
1 the4000 m shotgun library (H4000_SG) (top) and 500 m pyrosequenced library
2 (H186_500m) (bottom). **B.** *phn* sequences in the Sargasso Sea surface waters. Data from
3 BATS216 50m pyrosequenced library. See Supplementary Tables S5 and S6 for more
4 information.
5
6

Figure 1

E. coli B
phn operon
(10.9 kb)



HF70_11A08



phn deletion

G H I J K L M N +

CCFOS1

HF70_11A08

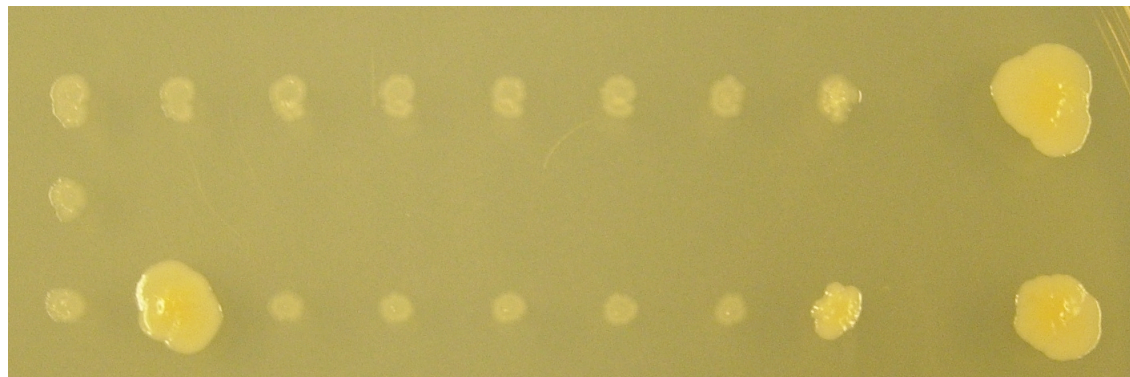


Figure 2. Phenotype of complementing clones

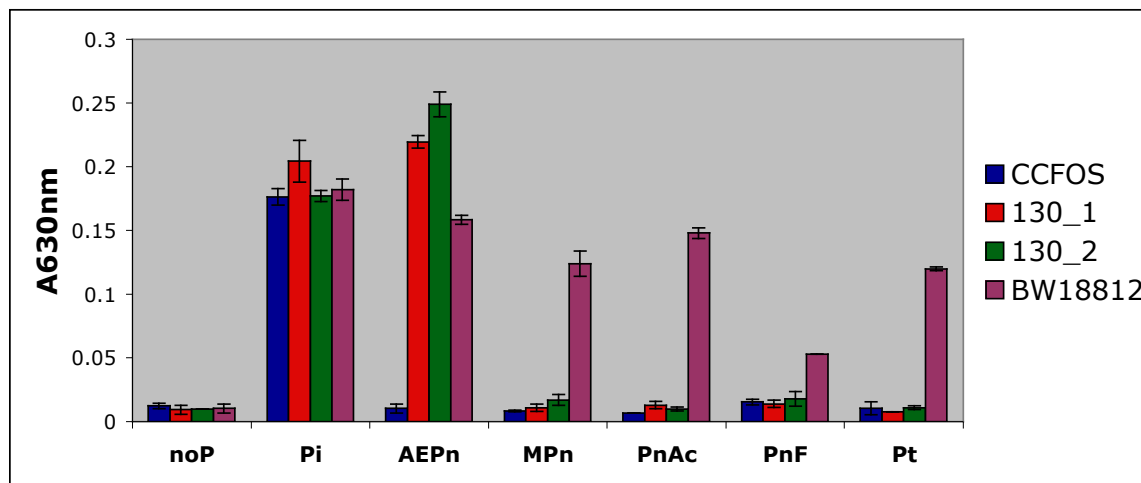
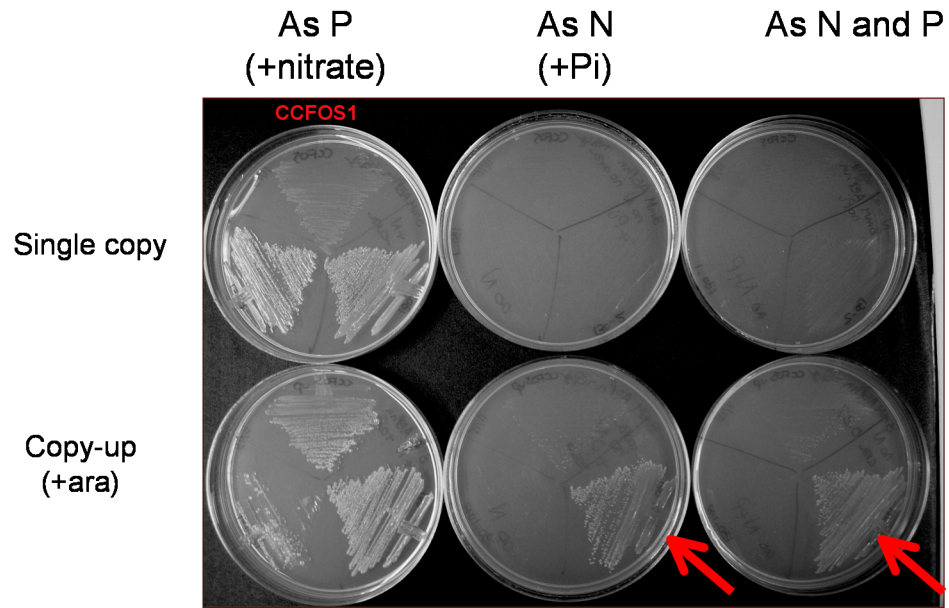


Figure 3. Note: I want to include some simpler version of the lower panel (just the operon in a few marine strains)

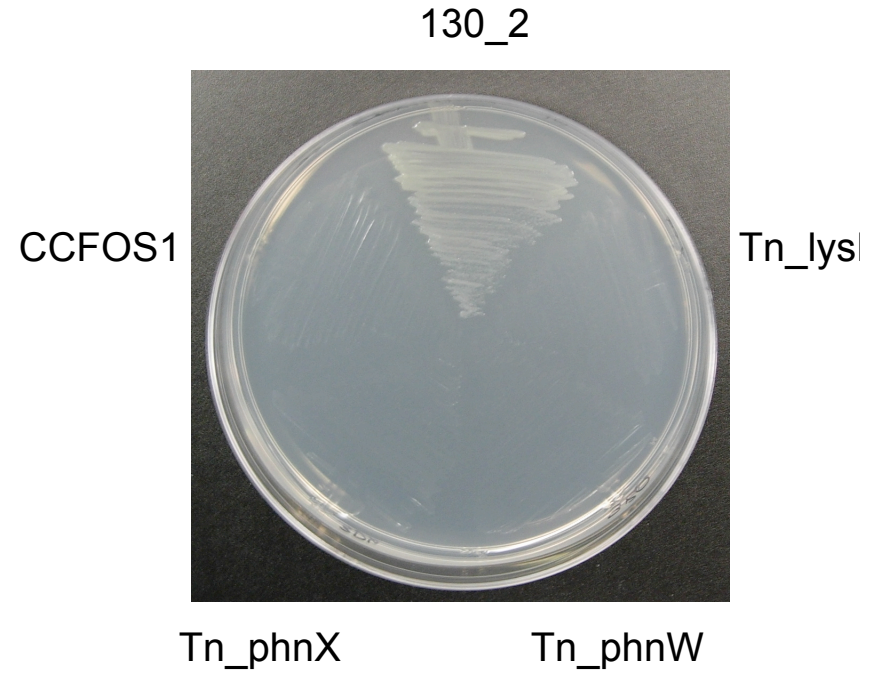
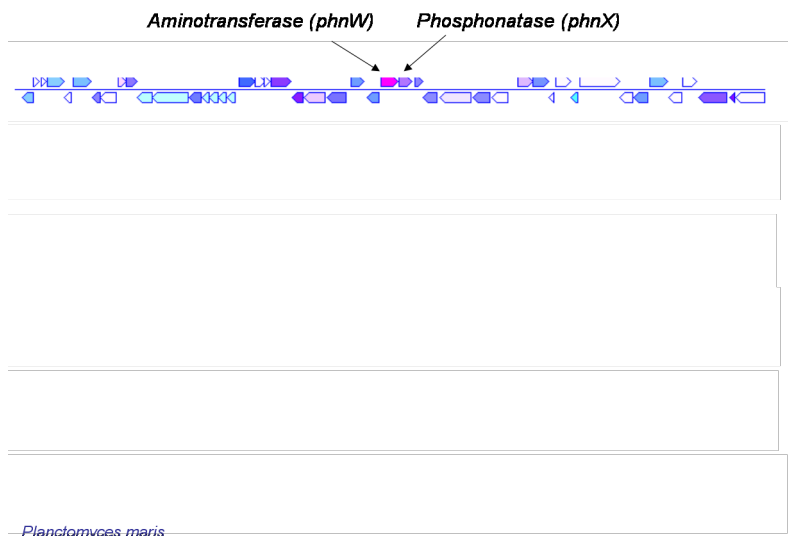
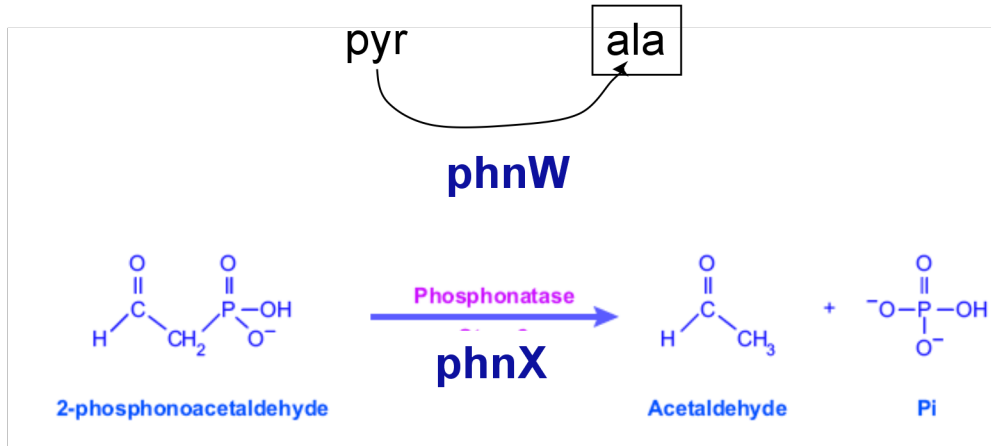


Figure 4. Required genes are boxed

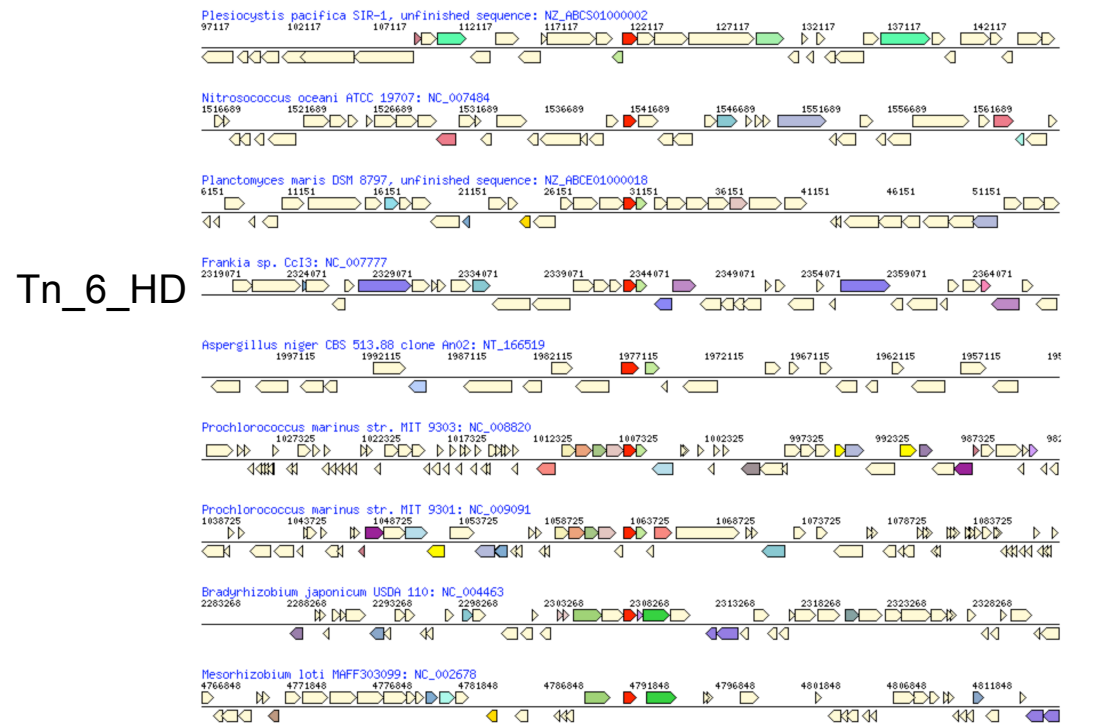
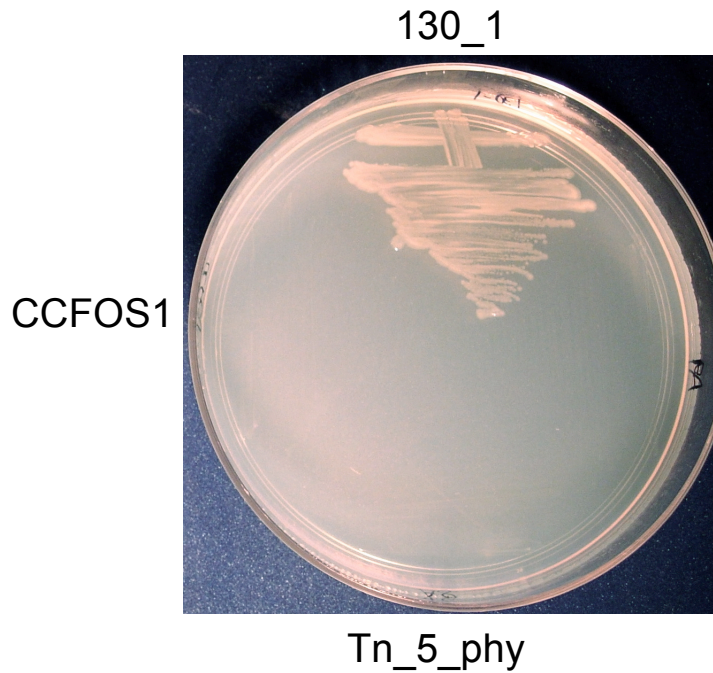
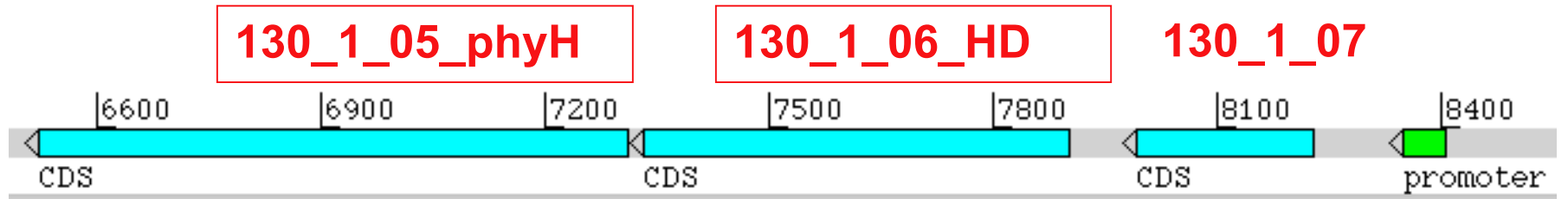


Figure 5. Growth curves on P

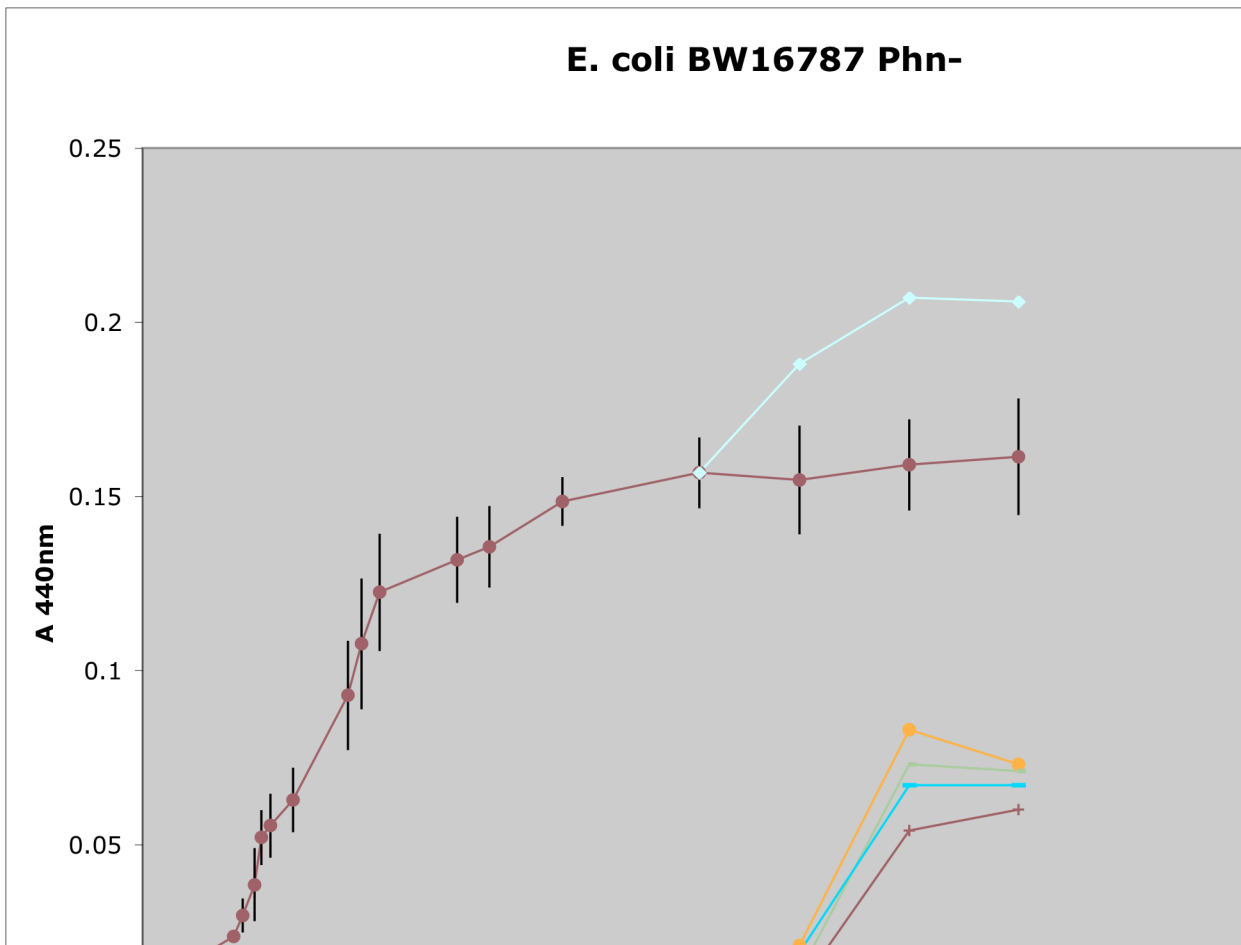


Table 1. Pn gene counts in shotgun libraries

size normalized % 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 normalized % 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)