

# Depth distributions of alkaline phosphatase and phosphonate utilization genes in the North Pacific Subtropical Gyre

Haiwei Luo<sup>1,\*,\*\*</sup>, Hongmei Zhang<sup>2,\*\*</sup>, Richard A. Long<sup>1,3</sup>, Ronald Benner<sup>1,3</sup>

<sup>1</sup>Department of Biological Sciences, University of South Carolina, Columbia, South Carolina 29208, USA

<sup>2</sup>Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, South Carolina 29208, USA

<sup>3</sup>Marine Science Program, University of South Carolina, Columbia, South Carolina 29208, USA

**ABSTRACT:** The biochemical composition of dissolved organic phosphorus (DOP) in the ocean is dominated by phosphoesters (C-O-P and C-O-P-O-C bonds), which are hydrolyzed by a diverse group of alkaline phosphatases (PhoA, PhoD, PhoX), and by phosphonates (C-P bond), which are degraded by C-P lyases and hydrolases. We designed a bioinformatics pipeline and a statistical approach to recover and analyze the alkaline phosphatase and phosphonate utilization genes from a metagenomic database derived from water samples collected from 7 depths (between 10 and 4000 m) in the oligotrophic North Pacific Subtropical Gyre. The alkaline phosphatase genes *phoD* and *phoX* were more abundant than *phoA* in the euphotic zone (10–130 m) and in deep waters (500–4000 m). The C-P lyase genes were most abundant in the euphotic zone at 70 m and were rare in deep water ( $\geq 500$  m) where phosphate concentrations were relatively high. These observations indicate that phosphonates are utilized primarily as a phosphorus source by bacterial C-P lyases; this is consistent with the observation that C-P lyase genes are part of the *pho* regulon which is expressed upon phosphorus limitation. In contrast, C-P hydrolase and alkaline phosphatase genes were often more abundant in deep waters, indicating that DOP serves mainly as a carbon and energy source in phosphate-rich deep waters which are depleted in bioavailable dissolved organic matter (DOM). The observed differences in depth distributions and presumed functions of C-P lyase and hydrolase genes indicate variability in the chemical composition of phosphonates between the euphotic zone and deep waters.

**KEY WORDS:** Phosphonate hydrolase · C-P lyase · Alkaline phosphatase · Microbial phosphorus metabolism

Resale or republication not permitted without written consent of the publisher

## INTRODUCTION

Phosphorus (P) occurs in a wide variety of forms in the ocean. Inorganic phosphate ( $P_i$ ) is a preferred source of P for marine bacteria (Karl 2000), but  $P_i$  concentrations are very low in surface waters of the vast subtropical gyres and often limit biological production, respiration and diazotrophic dinitrogen ( $N_2$ ) fixation in many oceanic regions (Michaels et al. 1996, Cotner et al. 1997, Sanudo-Wilhelmy et al. 2001, Sala et al. 2002, Obernosterer et al. 2003, Thingstad et al. 2005, Webb et al. 2007). Dissolved organic phosphorus (DOP) dom-

inates the dissolved phosphorus pool in the surface ocean (Wu et al. 2000). The high-molecular-weight (HMW) DOP reservoir consists primarily of phosphoesters (75%) and phosphonates (25%) (Clark et al. 1998, Kolowitz et al. 2001), which are substrates of microbial alkaline phosphatases (APases) and C-P utilization enzymes, respectively. Marine microbial APases and C-P utilization enzymes are involved in the dominant pathways of DOP utilization in the ocean.

The alkaline phosphatases (APases) include several families (PhoA, PhoX, PhoD) which differ in their metal cofactor requirements, substrate specificities and sub-

\*Email: hluo2006@gmail.com

\*\*These authors contributed equally to this article

cellular localization (Luo et al. 2009). For instance, PhoA uses zinc (Zn) and magnesium (Mg) as cofactors (Sone et al. 1997) and monoesters as substrates (Coleman 1992), whereas PhoX and PhoD require calcium (Ca) (Yamane & Maruo 1978b, Wu et al. 2007) and hydrolyze both monoesters and diesters (Yamane & Maruo 1978a,b, Eder et al. 1996, Wu et al. 2007). In addition, PhoX is predominantly extracellular, whereas a substantial number of the PhoA and PhoD enzymes are distributed in the cytoplasm (Luo et al. 2009). Recent studies have shown that *phoA* and *phoX* co-occur in a marine photosynthetic diazotrophic cyanobacterium, *Trichodesmium*, and that *phoX* has a higher expression than *phoA* during phosphate limitation (Orchard et al. 2009). While *phoD* is more abundant than other APase gene families near the ocean's surface (Luo et al. 2009), the relative abundance of these different APase families in the deeper ocean is unknown.

Phosphonates contain a highly stable carbon-phosphorus (C-P) bond, making them more resistant to microbial degradation than phosphoesters (Quinn et al. 2007). Two major enzyme systems, C-P lyase and C-P hydrolase, that degrade phosphonates have been described (Quinn et al. 2007, White & Metcalf 2007). The C-P lyase system is a multienzyme complex consisting of PhnG, PhnH, PhnI, PhnJ, PhnK, PhnL and PhnM (White & Metcalf 2007). The C-P lyase genes are part of the bacterial *pho* regulon, which is expressed upon P<sub>i</sub> limitation. Utilization of phosphonates through C-P lyase has been reported in *Trichodesmium* (Dyhrman et al. 2006). Three C-P hydrolase enzymes have been reported, including phosphonoacetate hydrolase (PhnA), phosphonoacetaldehyde hydrolase (phosphonatase, PhnX), and phosphonopyruvate hydrolase (PaA) (Quinn et al. 2007). Microbes utilizing C-P hydrolases can acquire carbon (C), nitrogen (N), phosphorus (P) and energy from phosphonates (McMullan & Quinn 1994, McGrath et al. 1997, O'Loughlin et al. 2006, Gilbert et al. 2009, Kulakova et al. 2009), while those using C-P lyase can derive only P from such compounds (Wanner 1994, Dyhrman et al. 2006). Laboratory studies indicate that phosphate limitation induces expression of C-P lyase genes and some alkaline phosphatase genes (Jiang et al. 1995, Dyhrman et al. 2006, Orchard et al. 2009, Sebastian & Ammerman 2009), whereas expression of C-P hydrolase genes is independent of phosphate concentrations (Kulakova et al. 2009). The only known exceptions are when *phnX* is part of the *pho* regulon in *Enterobacter aerogenes* and *Salmonella typhimurium* (Quinn et al. 2007).

The most pronounced concentration gradients of phosphate in the subtropical gyre are found between surface waters of the euphotic zone and the oxygen minimum layer in the lower mesopelagic zone. While

cell-specific APase activity is greater in the deep ocean than in surface waters (Koike & Nagata 1997, Hoppe & Ullrich 1999, Baltar et al. 2009), no studies have addressed the depth-dependent variability of C-P lyase and C-P hydrolase activities. The recent establishment of metagenomic data containing thousands of sequences at each of 7 depths in the North Pacific Subtropical Gyre (NPSG), generated using bidirectional sequencing of fosmid inserts, provides an opportunity to examine the vertical distribution of DOP utilization genes in the free-living bacterial community (DeLong et al. 2006). However, the current dataset is not without limitations. It sampled only a small number of prokaryotic genomes; hence, any result drawn from the bioinformatic analysis requires rigorous statistical argument. In addition, the current dataset does not include gene expression and enzyme activity data. Gene richness is not equivalent to gene expression, and our results are suggestive of a metabolic potential rather than measurements of activity. In the present study, we use these data to test the primary hypothesis that C-P lyase genes are most prevalent in phosphate-depleted waters whereas C-P hydrolase genes are most prevalent in phosphate-replete waters. The depth distribution of C-P utilization genes is also compared with that of APase genes.

## MATERIALS AND METHODS

### Identification of C-P utilization and APase genes from NPSG depth profile metagenomic databases.

The NPSG depth profile metagenomic peptide database was downloaded from Community Cyberinfrastructure for Advanced Marine Microbial Ecology (CAMERA) web site (<http://camera.calit2.net/>) in April 2009. A similarity search of C-P lyases (PhnG, PhnH, PhnI, PhnJ, PhnM, PhnJ, PhnK), C-P hydrolases (PhnA, PhnX), APases (PhoA, PhoD, PhoX), and *recA* against the NPSG depth profile metagenomic peptide database (DeLong et al. 2006) was done using PSI-BLAST software (Altschul et al. 1997) with an E cutoff of 0.1. The query sequence information is listed in Table S1 in the supplement at [www.int-res.com/articles/suppl/a062p061\\_supp.pdf](http://www.int-res.com/articles/suppl/a062p061_supp.pdf), and the recovered sequence ID and depth information of the above gene homologs are listed in Tables S2, S3 & S4 in the supplement. The retrieved sequences were used as input in RPS-BLAST (Altschul et al. 1997) searches (E = 0.1) against CDD, which consists of multiple protein domain databases including Pfam, COG, SMART, Protein Clusters, and internally established database by the CDD research group at NCBI (Marchler-Bauer et al. 2009). The conserved domains of these proteins are curated in CDD. The sequences having the top hits, known as

proteins of interest, were retained. All of the identified homologs were searched against the NCBI non-redundant database by BLAST, and the queries with the best hits were treated as the true homologs.

PhnA and PhnX are not annotated in CDD, and their homologs in the metagenomic databases were identified by a BLAST search against the NCBI non-redundant database. Only the top hit known as PhnA or PhnX was treated as the true homolog. PaA is another C-P hydrolase; it resembles phosphoenolpyruvate mutase (PEPM, an enzyme responsible for C-P bond synthesis) at the sequence and structural level (Chen et al. 2006), and both proteins are annotated as PEPM in CDD. PaA is distinguished from PEPM by the presence of Thr118 in the mobile loop (Chen et al. 2006, Kulakova et al. 2009). It is important to note that in some PaA peptides this Thr is replaced by Asn, which is used by PEPM, making it indistinguishable from PEPM in this case. Thus, we did not include PaA in the analysis. The abundance of the genes was normalized to the genome equivalents as described in previous studies (Karl et al. 2008, Martinez et al. 2010).

**Statistical analysis of C-P lyase gene abundance among different depths in the NPSG.** Statistical analyses were applied to test the abundance of 5 C-P lyase genes: *phnG*, *phnH*, *phnI*, *phnJ* and *phnM*. We used the term 'executor genes' to denote these 5 genes. Two genes, *phnK* and *phnL*, were not considered in the statistical analyses as they are transporter genes (Chen et al. 1990) and have different stoichiometry from the other 5 genes. Data from the 7 sampling depths were divided into 2 groups based on the concentrations of phosphate measured in these water samples (DeLong et al. 2006). Surface water samples ( $\leq 70$  m) had phosphate concentrations  $< 50$  nmol kg<sup>-1</sup>, and samples from  $\geq 130$  m had concentrations  $> 50$  nmol kg<sup>-1</sup>. This grouping of samples is designed to test the hypothesis that the existence or absence of executor genes is influenced by phosphate concentrations.

The basic method applied was the one-sample test on proportions. Specifically, a 95% confidence interval was set up to indicate the range of possible true proportions (or population proportions) of executor genes, which was defined as:

$$p \pm (1.96 \sqrt{p(1-p)/N}) \quad (1)$$

in which  $N$  is the number of observed genes in total. The symbol  $p$  denotes the sample proportion. In this context, sample proportions are the proportions of the executor genes among all the genes collected and mathematically defined as the ratio between the number of observed executor genes and the number of observed genes in total (i.e. sample size  $N$ ) at each depth in the water column. The confidence interval was set up based on the normality assumption of proportions when

the sample size (number of all genes collected) is large (Pagano & Gauvreau 2000), which was certainly a valid assumption for our data. If zero is not included in the interval, then the existence of the executor genes can be confirmed with 95% of confidence.

The above method would not be applicable if sample proportions of the executor genes were zeros because any 95% confidence intervals would include zero. This indicated that it was unlikely to see any executor genes at the designated water depths where these genes were not found. However, the absence of the executor genes could indicate that these genes were rare and the sample size was not large enough or the number of observed executor genes might have been miscounted, i.e. misclassification might have occurred.

To take these concerns into account, we considered the following 2 directions for further investigation: (1) increasing the sample size so that it was large enough to collect 1 executor gene; or (2) increasing the number of observed executor genes to correct possible misclassifications. It can be easily shown, based on the definition of confidence intervals, that if direction 2 cannot lead to the confirmation of the true existence of executor genes, then the existence cannot be confirmed through direction 1 either (see supplement). In the 'Results' section we thus explore the possibility of direction 2 first.

In our analyses, we assumed that the water sample at each sampling depth was a random sample, i.e. each gene at each depth had the same *a priori* probability of being included in the sample. Table 1 lists the number of executor genes observed in 2 different sample groups ( $\leq 70$  m and  $\geq 130$  m). All the water samples were different because each sample was collected at a different water depth. We used  $N_{70}$  to denote the number of observed genes at water depths in the upper 70 m of the water column and  $N_{130}$  to denote the number of observed genes at depths  $\geq 130$  m.

**Simulation study of sample size effect on sample proportion.** To demonstrate that the sample size (the total number of genes collected) is sufficiently large, we conducted a simulation study to see how sample proportion varies with sample size. The simulation

Table 1. Number of observed C-P lyase executor genes in surface ( $\leq 70$  m) and deep ( $\geq 130$  m) waters in the NPSG metagenomic peptide database (DeLong et al. 2006)

Genes collected	$\leq 70$ m	$\geq 130$ m
Five executor genes <sup>a</sup>	7	0
Other genes	121338	327741
Observed genes in total	121345	327741

<sup>a</sup>Five executor genes include *phnG*, *phnH*, *phnI*, *phnJ* and *phnM*

study was based on the assumption that the presence of a C-P lyase gene follows a binomial distribution and that the true proportion  $p$  is given by:

$$p = p_{70} = 7/121345 = 5.77 \times 10^{-5} \quad (2)$$

The simulation result is presented in Fig. S1 in the supplement.

## RESULTS

### C-P lyase executor genes

We found that the genes *phnK* and *phnL* had a depth distribution different from those of *phnG*, *phnH*, *phnI*, *phnJ* and *phnM* (Table 2, Tables S2 & S5 in the supplement), suggesting that the former proteins may have different ecological functions from the latter proteins. BLAST analysis using *PhnK* and *PhnL* as queries to search the NCBI non-redundant database showed that these 2 proteins had significant similarity with ABC transporter proteins ( $E < 10^{-128}$ ). In addition, we found sequence similarity between *PhnC*, an ABC transporter protein, and both *PhnK* and *PhnL*. Furthermore, the Pfam database annotates *PhnK* and *PhnL* as phosphonate transporters rather than as phosphonate lytic executors, which strengthened the hypothesis that *PhnK* and *PhnL* are members of the ABC transporter superfamily (Chen et al. 1990).

Because *phnK* and *phnL* did not have the same distribution profile as other essential genes in the C-P lyase complex, they cannot represent the true abundance of the protein complex in the ocean. Therefore, *phnK* and *phnL* were excluded from the following statistical analysis. Hereafter, we designate the term 'executor genes' to denote the other 5 genes: *phnG*, *phnH*, *phnI*, *phnJ* and *phnM*.

### Vertical distribution pattern of C-P lyase, hydrolase and alkaline phosphatase genes

The number of genome equivalents at each depth was normalized by the number of *recA* genes, a single-copy gene in bacterial genomes (Karl et al. 2008, Luo et al. 2009, Martinez et al. 2010), identified at each depth. There were 4, 8, 6, 6, 13, 13 & 6 genome equivalents sampled at depths of 10, 70, 130, 200, 500, 770 & 4000 m, respectively. The C-P lyase executor genes were found exclusively at a depth of 70 m (Table 2 & Table S5 in the supplement). Of the 5 executor genes, *phnM* and *phnJ* were absent from surface waters and all 5 executor genes were missing from 10 m (Table 2 & Table S5). Among the 2 C-P hydrolase genes, *phnX* was most abundant in deep waters, whereas *phnA* was more uniformly distributed throughout the water column (Table 2 & Table S6). Among the 3 APase families, *phoA* was found in deep waters but was absent in surface waters, whereas *phoD* and *phoX* genes were present in both surface and deep waters (Table 2 & Table S7). In addition, *phoD* and *phoX* were more abundant than *phoA* throughout the water column (Table 2 & Table S5).

### Statistical analysis results of C-P lyase executor gene distribution data in the NPSG depth profile

To validate the pattern of C-P lyase executor genes (*phnG*, *phnH*, *phnI*, *phnJ* and *phnM*) being present in the surface ocean metagenomes but absent in deeper samples, a statistical approach was designed. Testing the significance of the existence or absence of executor genes in the 2 depth regions is equivalent to testing the following 2 hypotheses: (1) executor genes exist in sur-

Table 2. The depth distributions of C-P lyase, C-P hydrolase and alkaline phosphatase genes, expressed as a percentage of genome equivalents that contain the gene of interest. The genome equivalent is represented by *recA*, provided that *recA* is a single-copy gene in bacterial genomes

Depth (m)	'Executors' of C-P lyase					Transporters of C-P lyase		C-P hydrolases		Alkaline phosphatase		
	<i>phnM</i>	<i>phnJ</i>	<i>phnH</i>	<i>phnG</i>	<i>phnI</i>	<i>phnK</i>	<i>phnL</i>	<i>phnA</i>	<i>phnX</i>	<i>phoA</i>	<i>phoD</i>	<i>phoX</i>
10	0					0		0		0		
70	21~40%					41~60%		61~80%		81~100%		
130	0					0		0		0		
200	0					0		0		0		
500	0					0		0		0		
770	0					0		0		0		
4000	0					0		0		0		

Colour scheme:

0	1~20%	21~40%	41~60%	61~80%	81~100%	>100%
---	-------	--------	--------	--------	---------	-------

face waters ( $\leq 70$  m) of the NPSG, and (2) executor genes were absent at depths  $\geq 130$  m.

We first tested the hypothesis that these genes exist in surface waters. Let  $p_{70}$  denote the sample proportion (see 'Materials and methods') in the first depth interval; we had:

$$p_{70} = 7/121345 = 5.77 \times 10^{-5} \quad (3)$$

Based on the confidence interval formula given in 'Materials and methods', a 95% confidence interval was given as  $(1.00 \times 10^{-4}, 1.50 \times 10^{-5})$ , which clearly excluded the number zero. This implies that executor genes exist in the upper 70 m, but such genes are rare and account for only a very small proportion of the existing genes in the upper 70 m. An intuitive understanding of this conclusion is given as follows. It is true that a very small number of executor genes were observed in the sample. However, the sample size was large ( $N_{70} = 121345$ ), implying that the sample proportion is close to the true proportion (or population proportion) and reflects the true existence of such genes. Note that sample size means the total number of genes sampled, rather than the number of genome equivalents sampled. The simulation study suggested that, as long as the sample size gets sufficiently large ( $>100\,000$ ), the estimated sample proportion approaches the true proportion (Fig. S1 in the supplement).

Next, we tested the hypothesis that these executor genes are absent at depths  $\geq 130$  m. Because the observed proportion was zero, any 95% confidence intervals would include zeros. A conclusion could be drawn that there are no executor genes  $\geq 130$  m based on the current available data. Nevertheless, it is possible that the current sequence similarity-based querying approaches missed the distantly related true homologs of the executor genes. Following the idea presented in 'Materials and methods', we considered that the different numbers of executor genes were miscounted among the  $N_{130} = 327\,741$  genes in deeper waters. Specifically, we took the numbers of miscounted genes as  $N_{\text{mis}} = 1, 2, 3, 4, 5, 6$ . Again, the sample proportions and the 95% confidence intervals were calculated for each case.

The calculated confidence intervals together with the sample proportions were plotted in Fig. 1. It can be seen that zero was always included in the 95% confidence intervals unless we misclassified 4 or more executor genes, which is unlikely to happen in practice due to the characteristic of the rarity of these executor genes. We conclude that the pattern observed in the metagenomic dataset is valid, i.e. C-P lyase genes are present in the upper 70 m, whereas it is very unlikely that these genes occurred at depths  $\geq 130$  m.

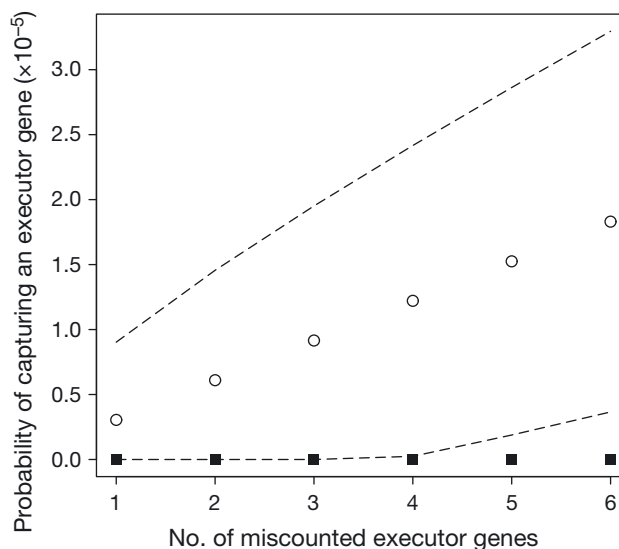


Fig. 1. Probability of capturing an executor gene (O) and 95% confidence intervals (dashed lines). x-axis: number of assumed miscounted executor genes among  $N_{130} = 327\,741$  genes at depth  $\geq 130$  m. (■) locations of zeros

## DISCUSSION

Phosphate concentrations are typically very low and often limiting for biological process in surface waters of the North Pacific Gyre (Karl et al. 1995, Karl & Tien 1997, Karl et al. 2001), but concentrations increase rapidly below the euphotic zone and should no longer limit biological activity. The C-P lyase genes are part of the *pho* regulon, and the expression of these genes is regulated by  $P_i$  concentrations (Dyhrman et al. 2006). It is thought that C-P lyase-expressing bacteria only derive P from phosphonates (Wanner 1994, Jiang et al. 1995, Dyhrman et al. 2006, Quinn et al. 2007), so it is intriguing to investigate the depth-dependent distribution of C-P lyase genes in contrasting nutrient regimes of surface water, being  $P_i$ -limited and with relatively abundant reactive dissolved organic matter (DOM), vs. the ocean's interior, which is replete with  $P_i$  and has less bioavailable DOM.

The 7 genes *phnG* to *phnM* encode the membrane-associated C-P lyase (Metcalf & Wanner 1993). It has been suggested that *phnK* and *phnL* encode phosphate transporters (Chen et al. 1990), which was verified *in silico* in the present study. We thus called the enzymes encoded by the remaining 5 genes, *phnG*, *phnH*, *phnI*, *phnJ* and *phnM* 'executor' proteins for C-P degradation. The non-stoichiometric distribution of transporter and executor proteins among the depth profile suggests that PhnK and PhnL might transport substrates other than phosphonates. There are lines of evidence from other transporter proteins to support this hypothesis. For example, the functional exchange-

ability of transporter proteins has been observed for UgpC and MalK transporters (Hekstra & Tommassen 1993). In addition, the glycine betaine transporter is able to transport dimethylsulfoniopropionate (DMSP) (Holtmann & Bremer 2004, Sowell et al. 2009). Likewise, although the C-P lyase pathway occurs in a limited number of cyanobacterial taxa (Dyhrman et al. 2006), the phosphonate transporter genes, such as *phnC*, *phnD* and *phnE*, are more widespread (Moore et al. 2005). It appears that the executor genes are more representative of C-P lyase than are the transporter genes *phnK* and *phnL*. The executor genes were found only at a depth of 70 m, indicating that they could be specific to the euphotic zone. The absence of C-P lyase executor genes at 10 m was probably due to the small number of genome equivalents ( $n = 4$ ) sampled compared with 8 genome equivalents sampled at 70 m. Given the limitations of the data, we only explored whether these genes occur in the  $P_i$ -depleted surface ocean ( $\leq 70$  m), the  $P_i$ -replete deep ocean ( $\geq 130$  m), or both. Using this approach, we combined the metagenomic databases from 10 m and 70 m depths into one set, and pooled the remaining databases (130, 200, 500, 770 & 4000 m) into another set.

Statistical analyses indicated that the C-P lyase executor genes exist only in surface waters (Fig. 2). Based on the convincing evidence that biological activity is often limited by  $P_i$  in surface waters and by bioavailable C in the deep ocean (Hoppe 2003), our analysis from the depth-profile metagenomic datasets supports the findings from bacterial cultures that marine bacteria primarily use C-P lyase for acquisition of P (Dyhrman et al. 2006, Quinn et al. 2007). In contrast, the depth distribution of C-P hydrolase genes spanned surface and deep waters, which is consistent with laboratory studies which show that specific phosphonates can serve as C, N, P and energy sources when processed by C-P hydrolases (McMullan & Quinn 1994, Quinn et al. 2007, Gilbert et al. 2009, Martinez et al. 2010). These observations indicate that marine microorganisms are adapted to the depth-variable resources in the ocean. It is important to note that the statistical model indicates that C-P lyase executor genes are present in surface waters of the euphotic zone, but it does not deny the possibility of observing these genes in samples from deep waters. Surface-dwelling bacteria are transported to deeper waters through vertical mixing events. Moreover, the seawater samples used to generate this dataset were pre-filtered with glass fiber (GFA) filters (DeLong et al. 2006) which excluded most of the particle-attached bacteria. It is likely that particle-attached bacteria carry these genes and sink to the ocean's interior. Given the limited dataset, the observed gene stoichiometry could be subject to considerable stochastic error. Therefore, the data analyses in the present study

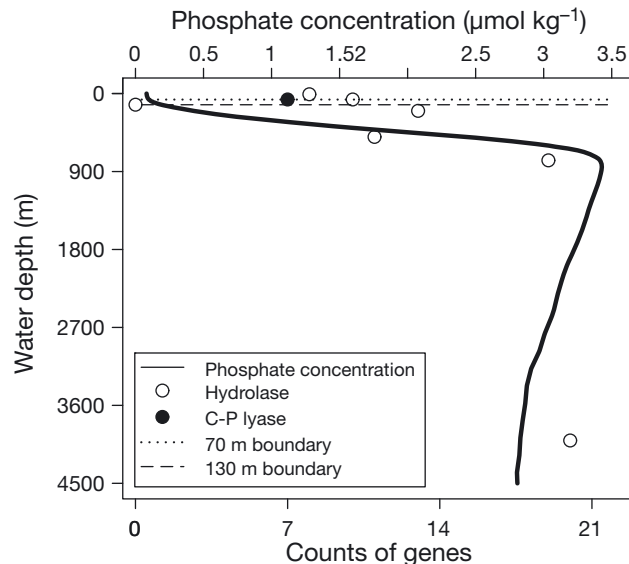


Fig. 2. Phosphate concentration, and the distribution of C-P lyase and hydrolase (in terms of their gene counts) along the water depths in the North Pacific Subtropical Gyre. C-P lyase genes include *phnG*, *phnH*, *phnI*, *phnJ* and *phnM*; C-P hydrolase genes include *phnA* and *phnX*. The phosphate concentration data were downloaded from the Hawaii Ocean Time-series (HOTs) website: <http://hahana.soest.hawaii.edu/index.html>

form an intriguing hypothesis rather than a definitive conclusion about the depth distribution of C-P lyase. Future studies need to construct larger metagenomic datasets of the prokaryotic community throughout the ocean water column.

Unlike C-P lyase, which cleaves the C-P bond of a wide range of substituted and unsubstituted phosphonates, C-P hydrolases appear to have strict substrate specificities (Quinn et al. 2007). For instance, PhnX is known to degrade only 2-aminoethylphosphonate (2-AEP) (Martinez et al. 2010) and PhnA is only known to degrade phosphonoacetate (Gilbert et al. 2009). Since *phnX* is more abundant in deep waters than in surface waters, it is likely that its substrate, 2-AEP, occurs in higher concentrations in deep waters. In contrast, *phnA* is more homogeneously distributed throughout the water column (Table 2 & Table S6 in the supplement), suggesting that its substrate, phosphonoacetate, occurs throughout the water column. It has been suggested that 2-AEP and its N-methylated compounds are the most abundant and prevalent biogenic phosphonates (Quinn et al. 2007, White & Metcalf 2007). In contrast, phosphonoacetate is a widely used synthetic antiviral agent, but its natural source remains unknown (Quinn et al. 2007). We found that *phnA* is much more abundant than *phnX* throughout the water column, indicating that phosphonoacetate could be a natural product that is more abundant and ubiquitous

than 2-AEP and its derivatives. The *phnA* gene also occurs in soil bacteria, suggesting that terrestrial bacteria could be a natural source of phosphonoacetate (Panas et al. 2006). Phosphonates can be present in various biomolecules, including lipids, proteins and antibiotics (Kolowitz et al. 2001, Sannigrahi et al. 2006). The relative abundances of individual components of phosphonates remain unknown because it is difficult to identify and quantify phosphonates in seawater. The present study provides indirect evidence for the depth distribution of specific phosphonate compounds.

In the NPSG, the upper euphotic zone has been in a decades-long transition from N-limitation to P-limitation (Karl et al. 2001). APase activity is detectable in this region, and the occurrences of *phoA*, *phoX* and *phoD* in the upper ocean (Table 2 & Table S7 in the supplement) underlies these field observations.  $P_i$  concentrations were higher ( $66 \text{ nmol kg}^{-1}$ ) at the base of the deep chlorophyll maximum (~130 m) than in the upper euphotic zone (Fig. 2; DeLong et al. 2006), but APase activity is still detectable (Hoch & Bronk 2007) at this location. The occurrence of *phoD* and *phoX* at this depth (Table 2 & Table S7) is consistent with APase activity at the base of the deep chlorophyll maximum.  $P_i$  concentrations increase below the euphotic zone and reach a maximum in the core of the oxygen minimum layer at 770 m (Fig. 2) where *phoD* and *phoX* were also relatively abundant (Table 2 & Table S7). The occurrence of APase genes in the oxygen minimum layer is consistent with the rapid decomposition of sinking particles and increasing concentrations of phosphate and nitrate in the mesopelagic zone (200–1000 m). The occurrence of *phoA*, *phoX* and *phoD* at 4000 m (Table 2 & Table S7) is consistent with reports of APase activity at this depth in the NPSG (Koike & Nagata 1997).

Previous studies demonstrated that cell-specific APase activity is higher in the deep ocean than in surface waters of the NPSG (Koike & Nagata 1997), subtropical North Atlantic (Baltar et al. 2009) and Indian Ocean (Hoppe & Ullrich 1999). Similarly, a positive correlation was found between  $P_i$  concentration and APase activity in the deep chlorophyll maximum layer of the Eastern Tropical North Pacific (Hoch & Bronk 2007) and in the benthic nepheloid layer of a large mesotrophic lake (Kim et al. 2007). In the mesopelagic and abyssal zones,  $P_i$  and other nutrients are abundant, but bioavailable C and energy sources become limiting for heterotrophic bacteria (Karl & Letelier 2009) and it is possible that APase is used to sequester C and energy (Hoppe & Ullrich 1999, Hoppe 2003). Laboratory studies showed that bacterial APase activity was repressed when cultures were supplemented with bioavailable carbon (e.g. glucose, cellobiose, amino

acids), indicating that APase can be used to supply bioavailable carbon for energy and bacterial growth (Chrost 1991, Siuda & Chrost 2001). On the other hand, APase hydrolyzes the P-O bond and is not known to alter the chemical structure of the C skeleton of DOP substrates. It is likely that some unknown coupling of metabolic pathways or non-specific APase activity occurs in deep waters for transforming biorefractory to bioavailable DOM. Our finding of abundant APase genes in the deep NPSG corroborated previous field studies indicating that the  $P_i$ -repressible APase activity is not regulated by the end product,  $P_i$ , in the deep ocean. This in turn lends support to the hypothesis that APases play other roles, possibly in the utilization of otherwise refractory organic matter, in the deep ocean.

Previous studies demonstrated that the concentrations of high-molecular-weight (HMW) DOP decrease with depth, but the relative proportion of phosphoesters and phosphonates remains fairly constant throughout the water column (Clark et al. 1998, Kolowitz et al. 2001). These observations indicate that degradation of HMW DOP in the deep ocean is non-selective with regard to phosphoesters and phosphonates (Kolowitz et al. 2001). Phosphoesters occur in numerous forms with varying reactivities (Karl & Yanagi 1997), and it appears that some APases utilize refractory organic matter in the deep ocean. No studies have compared the capabilities of APases from surface and deep waters in hydrolyzing refractory phosphoesters. Moreover, the decreasing concentrations of DOP and bacterial cells with depth indicate that the encounter rate between bacteria and DOP compounds would be reduced in the deep ocean. Because bacteria in the deep ocean are limited by bioavailable carbon, enzymatic degradation of DOP could provide an important source of bioreactive carbon in the deep ocean (Koike & Nagata 1997, Hoppe 2003, Quinn et al. 2007). The energy and carbon used for enzyme production could be returned by the acquisition of bioavailable carbon through the hydrolysis of DOP. This in turn suggests that APases in the deep ocean might target the C moiety of DOP in the deep ocean.

Only *phoD* was present at 10 m depth, and on average each bacterium possessed 1 copy of *phoD*, whereas both *phoX* and *phoA* were absent. This is consistent with our previous findings that *phoD* is significantly more abundant than *phoX* and *phoA* in surface waters (<10 m) of the global oceans, whereas *phoX* and *phoA* abundances are not statistically different (Luo et al. 2009). In deeper waters of the euphotic zone, *phoX* was more abundant than *phoA*, and the abundances of *phoD* and *phoX* were comparable, suggesting that the depth distributions of *phoD*, *phoX* and *phoA* are variable in the euphotic zone as is the microbial community composition (DeLong et al. 2006).

It is important to point out that the experimental procedures from sample collection to sample sequencing are relevant to the *in silico* predictions made in the present study. For instance, a clone library was used prior to Sanger sequencing (DeLong et al. 2006), and certain genes and taxa could be under-represented in the metagenomic clone library (DeLong 2009). The random sampling nature would theoretically sample the dominant genomes present, and the low sequencing depth suggests that few recovered sequences would be from uncommon taxa. As in the case of P-metabolizing genes, the less abundant *Gamma-proteobacteria* produce more APase genes than the most dominant *Alphaproteobacteria* in the global surface oceans (Luo et al. 2009). Because community structure varies with water depth (DeLong et al. 2006), low sequencing depth might miss some less abundant but important functional taxa. While tentative due to the limited data, this study demonstrates that depth profiles of metagenomic data are an excellent resource for understanding the metabolic potential and stoichiometry of DOP metabolism genes within the ocean biome and for providing novel insights about the molecular composition of DOP.

**Acknowledgements.** We thank Dr. Toshi Nagata (University of Tokyo) for helpful discussion. H.Z. was supported by the National Institute of Health (Grant 1R03HL095429-01, NIH, NHLBI and Grant 1 R01 HL082925-01A2, NIH, NHLBI); R.B. was supported by the National Science Foundation (Grant NSF 0850653). All computations were conducted on a 128-core shared memory computer of the High Performance Computing Group at the University of South Carolina.

#### LITERATURE CITED

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Baltar F, Aristegui J, Sintes E, Van Aken HM, Gasol JM, Herndl GJ (2009) Prokaryotic extracellular enzymatic activity in relation to biomass production and respiration in the meso- and bathypelagic waters of the (sub)tropical Atlantic. *Environ Microbiol* 11:1998–2014
- Chen CM, Ye QZ, Zhu ZM, Wanner BL, Walsh CT (1990) Molecular biology of carbon-phosphorus bond cleavage. Cloning and sequencing of the *phn* (*psiD*) genes involved in alkylphosphonate uptake and C-P lyase activity in *Escherichia coli* B. *J Biol Chem* 265:4461–4471
- Chen CCH, Han Y, Niu W, Kulakova AN and others (2006) Structure and kinetics of phosphonopyruvate hydrolase from *Vorriovorax* sp. Pal2: new insight into the divergence of catalysis within the PEP mutase/isocitrate lyase superfamily. *Biochemistry* 45:11491–11504
- Chrost RJ (1991) Environmental control of the synthesis and activity of aquatic microbial ectoenzymes. In: Chrost RJ (ed) *Microbial enzymes in aquatic environments*. Springer-Verlag, Berlin, p 29–59
- Clark LL, Ingall ED, Benner R (1998) Marine phosphorus is selectively remineralized. *Nature* 393:426
- Coleman JE (1992) Structure and mechanism of alkaline phosphatase. *Annu Rev Biophys Biomol Struct* 21:441–483
- Cotner J, Ammerman J, Peele E, Bentzen E (1997) Phosphorus-limited bacterioplankton growth in the Sargasso Sea. *Aquat Microb Ecol* 13:141–149
- DeLong EF (2009) The microbial ocean from genomes to biomes. *Nature* 459:200–206
- DeLong EF, Preston CM, Mincer T, Rich V and others (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 311:496–503
- Dyhrman ST, Chappell PD, Haley ST, Moffett JW, Orchard ED, Waterbury JB, Webb EA (2006) Phosphonate utilization by the globally important marine diazotroph *Trichodesmium*. *Nature* 439:68–71
- Eder S, Shi L, Jensen K, Yamane K, Hulett FM (1996) A *Bacillus subtilis* secreted phosphodiesterase/alkaline phosphatase is the product of a *Pho* regulon gene, *phoD*. *Microbiology* 142:2041–2047
- Gilbert JA, Thomas S, Cooley NA, Kulakova A and others (2009) Potential for phosphonoacetate utilization by marine bacteria in temperate coastal waters. *Environ Microbiol* 11:111–125
- Hekstra D, Tommassen J (1993) Functional exchangeability of the ABC proteins of the periplasmic binding protein-dependent transport systems Ugp and Mal of *Escherichia coli*. *J Bacteriol* 175:6546–6552
- Hoch MP, Bronk DA (2007) Bacterioplankton nutrient metabolism in the Eastern Tropical North Pacific. *J Exp Mar Biol Ecol* 349:390–404
- Holtmann G, Bremer E (2004) Thermoprotection of *Bacillus subtilis* by exogenously provided glycine betaine and structurally related compatible solutes: involvement of opu transporters. *J Bacteriol* 186:1683–1693
- Hoppe HG (2003) Phosphatase activity in the sea. *Hydrobiologia* 493:187–200
- Hoppe HG, Ullrich S (1999) Profiles of ectoenzymes in the Indian Ocean: phenomena of phosphatase activity in the mesopelagic zone. *Aquat Microb Ecol* 19:139–148
- Jiang W, Metcalf W, Lee K, Wanner B (1995) Molecular cloning, mapping, and regulation of *Pho* regulon genes for phosphonate breakdown by the phosphonate pathway of *Salmonella typhimurium* LT2. *J Bacteriol* 177:6411–6421
- Karl DM (2000) Phosphorus, the staff of life. *Nature* 406:31–32
- Karl DM, Letelier R (2009) Marine habitats and conditions for microbial growth. In: Schaechter M (ed) *Encyclopedia of microbiology*. Elsevier, Oxford, p 258–277
- Karl DM, Tien G (1997) Temporal variability in dissolved phosphorus concentrations in the subtropical North Pacific Ocean. *Mar Chem* 56:77–96
- Karl DM, Yanagi K (1997) Partial characterization of the dissolved organic phosphorus pool in the oligotrophic North Pacific Ocean. *Limnol Oceanogr* 42:1398–1405
- Karl DM, Letelier R, Hebel D, Tupas L, Dore J, Christian J, Winn C (1995) Ecosystem changes in the North Pacific subtropical gyre attributed to the 1991–92 El Niño. *Nature* 373:230–234
- Karl DM, Bjorkman KM, Dore JE, Fujieki L, Hebel DV, Houlihan T, Letelier RM, Tupas LM (2001) Ecological nitrogen-to-phosphorus stoichiometry at station ALOHA. *Deep Sea Res II* 48:1529–1566
- Karl DM, Beversdorf L, Bjorkman KM, Church MJ, Martinez A, Delong EF (2008) Aerobic production of methane in the sea. *Nat Geosci* 1:473–478
- Kim C, Nishimura Y, Nagata T (2007) High potential activity of alkaline phosphatase in the benthic nepheloid layer of a



- large mesotrophic lake: implications for phosphorus regeneration in oxygenated hypolimnion. *Aquat Microb Ecol* 49:303–311
- Koike I, Nagata T (1997) High potential activity of extracellular alkaline phosphatase in deep waters of the central Pacific. *Deep Sea Res II* 44:2283–2294
- Kolowith LC, Ingall ED, Benner R (2001) Composition and cycling of marine organic phosphorus. *Limnol Oceanogr* 46:309–320
- Kulakova AN, Kulakov LA, Villarreal-Chiu JF, Gilbert JA, McGrath JW, Quinn JP (2009) Expression of the phosphonoalanine-degradative gene cluster from *Variovorax* sp. Pal2 is induced by growth on phosphonoalanine and phosphonopyruvate. *FEMS Microbiol Lett* 292:100–106
- Luo H, Benner R, Long RA, Hu J (2009) Subcellular localization of marine bacterial alkaline phosphatases. *Proc Natl Acad Sci USA* 106:21219–21223
- Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK and others (2009) CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res* 37: D205–D210
- Martinez A, Tyson GW, DeLong EF (2010) Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses. *Environ Microbiol* 12:222–238
- McGrath JW, Ternan NG, Quinn JP (1997) Utilization of organophosphonates by environmental micro-organisms. *Lett Appl Microbiol* 24:69–73
- McMullan G, Quinn JP (1994) In vitro characterization of a phosphate starvation-independent carbon-phosphorus bond cleavage activity in *Pseudomonas fluorescens* 23F. *J Bacteriol* 176:320–324
- Metcalf WW, Wanner BL (1993) Mutational analysis of an *Escherichia coli* fourteen-gene operon for phosphonate degradation, using *TnphoA'* elements. *J Bacteriol* 175: 3430–3442
- Michaels A, Olson D, Sarmiento J, Ammerman J and others (1996) Inputs, losses and transformations of nitrogen and phosphorus in the pelagic North Atlantic Ocean. *Biogeochemistry* 35:181–226
- Moore LR, Ostrowski M, Scanlan DJ, Feren K, Sweetsir T (2005) Ecotypic variation in phosphorus-acquisition mechanisms within marine picocyanobacteria. *Aquat Microb Ecol* 39:257–269
- O'Loughlin SN, Graham RL, McMullan G, Ternan NG (2006) A role for carbon catabolite repression in the metabolism of phosphonoacetate by *Agromyces fucosus* Vs2. *FEMS Microbiol Lett* 261:133–140
- Obernosterer I, Kawasaki N, Benner R (2003) P-limitation of respiration in the Sargasso Sea and uncoupling of bacteria from P-regeneration in size-fractionation experiments. *Aquat Microb Ecol* 32:229–237
- Orchard ED, Webb EA, Dyrman ST (2009) Molecular analysis of the phosphorus starvation response in *Trichodesmium* spp. *Environ Microbiol* 11:2400–2411
- Pagano M, Gauvreau K (2000) Principles of biostatistics, 2nd edn. Brooks/Cole, Belmont, CA
- Panas P, Ternan NG, Dooley JSG, McMullan G (2006) Detection of phosphonoacetate degradation and phnA genes in soil bacteria from distinct geographical origins suggest its possible biogenic origin. *Environ Microbiol* 8:939–945
- Quinn JP, Kulakova AN, Cooley NA, McGrath JW (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
- Sala MM, Peters F, Gasol JM, Pedrós-Alió C, Marrasé C, Vaqué D (2002) Seasonal and spatial variation in the nutrient limitation of bacterioplankton growth in the north-western Mediterranean. *Aquat Microb Ecol* 27:47–56
- Sannigrahi P, Ingall ED, Benner R (2006) Nature and dynamics of phosphorus-containing components of marine dissolved and particulate organic matter. *Geochim Cosmochim Acta* 70:5868–5882
- Sanudo-Wilhelmy SA, Kustka AB, Gobler CJ, Hutchins DA and others (2001) Phosphorus limitation of nitrogen fixation by *Trichodesmium* in the central Atlantic ocean. *Nature* 411:66–69
- Sebastian M, Ammerman JW (2009) The alkaline phosphatase PhoX is more widely distributed in marine bacteria than the classical PhoA. *ISME J* 3:563–572
- Siuda W, Chrost RJ (2001) Utilization of selected dissolved organic phosphorus compounds by bacteria in lake water under non-limiting orthophosphate conditions. *Pol J Environ Stud* 10:475–483
- Sone M, Kishigami S, Yoshihisa T, Ito K (1997) Roles of disulfide bonds in bacterial alkaline phosphatase. *J Biol Chem* 272:6174–6178
- Sowell SM, Wilhelm LJ, Norbeck AD, Lipton MS and others (2009) Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *ISME J* 3:93–105
- Thingstad TF, Krom MD, Mantoura RFC, Flaten GAF and others (2005) Nature of phosphorus limitation in the ultraoligotrophic Eastern Mediterranean. *Science* 309:1068–1071
- Wanner BL (1994) Phosphate-regulated genes for the utilization of phosphonates in members of the family Enterobacteriaceae. In: Torriani-Gorini A, Yagil E, Silver S (eds) Phosphate in microorganisms: cellular and molecular biology. American Society of Microbiology, Washington DC, p 215–221
- Webb EA, Jakuba RW, Moffett JW, Dyrman ST (2007) Molecular assessment of phosphorus and iron physiology in *Trichodesmium* populations from the western Central and western South Atlantic. *Limnol Oceanogr* 52:2221–2232
- White AK, Metcalf WW (2007) Microbial metabolism of reduced phosphorus compounds. *Annu Rev Microbiol* 61: 379–400
- Wu J, Sunda W, Boyle EA, Karl DM (2000) Phosphate depletion in the Western North Atlantic Ocean. *Science* 289: 759–762
- Wu JR, Shien JH, Shieh HK, Hu CC, Gong SR, Chen LY, Chang PC (2007) Cloning of the gene and characterization of the enzymatic properties of the monomeric alkaline phosphatase (PhoX) from *Pasteurella multocida* strain X-73. *FEMS Microbiol Lett* 267:113–120
- Yamane K, Maruo B (1978a) Alkaline phosphatase possessing alkaline phosphodiesterase activity and other phosphodiesterases in *Bacillus subtilis*. *J Bacteriol* 134:108–114
- Yamane K, Maruo B (1978b) Purification and characterization of extracellular soluble and membrane-bound insoluble alkaline phosphatases possessing phosphodiesterase activities in *Bacillus subtilis*. *J Bacteriol* 134:100–107