environmental microbiology

Environmental Microbiology (2010) doi:10.1111/j.1462-2920.2010.02310.x

Prevalence of a calcium-based alkaline phosphatase associated with the marine cyanobacterium *Prochlorococcus* and other ocean bacteria

Satish Kathuria[1] and Adam C. Martiny[1,2]*

Departments of [1]Earth System Science and [2]Ecology and Evolutionary Biology, University of California, Irvine, CA 92697, USA.

Summary

Phosphate plays a key role in regulating primary productivity in several regions of the world's oceans and here dissolved organic phosphate can be an important phosphate source. A key enzyme for utilizing dissolved organic phosphate is alkaline phosphatase and the *pho*A-type of this enzyme has a zinc cofactor. As the dissolved zinc concentration is low in phosphate depleted environments, this has led to the hypothesis that some phytoplankton may be zinc-P co-limited. Recently, it was shown that many marine bacteria contain an alternative form of alkaline phosphatase called *pho*X, but it is unclear which marine lineages carry this enzyme. Here, we describe the occurrence in low phosphate environments of *pho*X that is associated with uncultured *Prochlorococcus* and SAR11 cells. Through heterologous expression, we demonstrate that *pho*X encodes an active phosphatase with a calcium cofactor. The enzyme also functions with magnesium and copper, whereas cobalt, manganese, nickel and zinc inhibit enzyme activity to various degrees. We also find that uncultured SAR11 cells and cyanophages contain a different alkaline phosphatase related to a variant present in several *Prochlorococcus* isolates. Overall, the results suggest that many bacterial lineages including *Prochlorococcus* and SAR11 may not be subject to zinc-P co-limitation.

Introduction

Phosphate (P) plays a key role in regulating primary productivity in several regions of the world's oceans (Wu *et al*., 2000; Sanudo-Wilhelmy *et al*., 2001; Moore *et al*., 2002; Thingstad *et al*., 2005). In surface waters from these areas, dissolved inorganic phosphate (DIP) is found in very low concentrations (commonly below 10 nM). Instead, phytoplankton may support a significant fraction of primary production using dissolved organic phosphate (DOP) (Lomas *et al*., 2010). In some low P regions like the Sargasso Sea, this DOP pool can account for more than 80% of the total dissolved phosphorus pool (Wu *et al*., 2000; Cavender-Bares *et al*., 2001; Lomas *et al*., 2010). The DOP pool consists of a diverse range of compounds including phosphoesters (75–85%) and phosphonates (10–25%) (Kolowith *et al*., 2001; Karl and Björkman, 2002; Paytan and McLaughlin, 2007; Young and Ingall, 2010).

Phosphoesters are readily hydrolysed by the enzyme alkaline phosphatase produced by bacteria and phytoplankton. Culture studies have demonstrated that gene expression and activity of alkaline phosphatase are commonly regulated by P availability via the two-component regulatory system *pho*BR (e.g. Torriani-Gorini, 1987; Wanner, 1996; Moore *et al*., 2005; Martiny *et al*., 2006; Sebastian and Ammerman, 2009; Tetu *et al*., 2009; Zaheer *et al*., 2009). Many different families of phosphatases involved in P assimilation have been characterized, including the alkaline phosphatase *pho*A and *pho*X families. Commonly, alkaline phosphatases in bacteria are located in the periplasm, where they can hydrolyse phosphoesters passing through the cell wall without damaging intracellular phosphorylations (Wanner, 1996). *pho*A is exported to the periplasm via the sec pathway, whereas *pho*X is exported via the twin-arginine translocation pathway (Zaheer *et al*., 2009). Another difference between *pho*A and *pho*X is metal dependence, where *pho*A is activated by zinc and magnesium ions and *pho*X by calcium ions (Roy *et al*., 1982; Coleman, 1992). This metal requirement can be important as the total zinc concentration in some low P regions like the North Atlantic Ocean can be drawn down to below 0.12 nM (Jakuba *et al*., 2008), and may restrict alkaline phosphatase activity and DOP assimilation here (Shaked *et al*., 2006). Thus, cells containing a Ca-dependent *pho*X may not be subject to zinc-P co-limitation.

A recent study has shown the widespread presence of *pho*X among marine bacteria (Sebastian and

Received 9 April, 2010; accepted 13 June, 2010. *For correspondence. E-mail amartiny@uci.edu; Tel. (+1) 949 824 9713; Fax (+1) 949 824 3874.

Ammerman, 2009). However, many sequences from the ocean environment grouped with the γ-*Proteobacteria Marinomonas* and *Vibrio*, even though these two lineages are not very common in the open ocean. We hypothesize that these variants of *pho*X could be associated more dominant linages. Therefore, the goal of this study was to test if uncultured *Prochlorococcus* and other abundant bacteria contain *pho*X by analysing the Global Ocean Survey metagenomic libraries from different ocean regions. These libraries allow us to examine flanking genomic regions of *pho*X genes and possibly identify the phylogenetic position of lineages containing *pho*X. We also confirm the enzymatic function and metal requirements of these variants by heterologous expression. The results affect our understanding of the influence of trace metals on macro nutrient uptake and generally phytoplankton community structure and activity.

Results

Identification of phosphatase in Prochlorococcus

To test for the presence of a calcium-based phosphatase in *Prochlorococcus*, we searched metagenomic libraries for genes similar to *pho*X in marine *Synechococcus* and *Marinomonas*. The samples were part of the Global Ocean Survey (GOS) and covered sites from the Atlantic, Pacific and Indian Ocean (Rusch *et al*., 2007). As previously seen, we detected a large cluster related to *pho*X in V. *fischeri*, *Marinomonas* and *Roseobacter* (Sebastian and Ammerman, 2009) (Fig. 1 and Fig. S1). Within this cluster, we observed several clades with no culture representatives. One clade (cluster IV in Fig. 1 and Fig. S1) has a GC content of 37% ($n = 28$) compared with 45% in *Marinomonas* and 39% in V. *fischeri*. Importantly, several paired-end sequences only share high nucleotide similarity with genes from high-light adapted (HL) *Prochlorococcus* (Fig. 1 and Fig. S1, and Table S1). Because the paired-end sequences linked to HL *Prochlorococcus* are located on the same fragments of DNA as the *pho*X sequences, this supports the presence of *pho*X in *Prochlorococcus*. The low GC content supports that these sequences are not associated with *Marinomonas*. Considering the low GC, pair-end linkage to at least 14 genes most similar to *Prochlorococcus*, and the high occurrence in the Sargasso, the most parsimonious interpretation is that *pho*X cluster IV sequences are associated with *Prochlorococcus* and we will refer to this as *Prochlorococcus pho*X. We also noticed one sequence (1093015571554) that formed an outgroup to clade IV and had a slightly higher GC content (Fig. S1). This sequence may represent *pho*X in the low-light adapated *Prochlorococcus* eNATL clade although we need more sequences to confirm this. Several marine *Synechococcus* and other cyanobacterial strains also carry *pho*X. However, these all belong to a different subfamily of *pho*X compared with all the entries in Fig. 1 and Fig. S1 (Zaheer *et al*., 2009). Thus, the putative *Prochlorococcus* variant is very distantly related to *pho*X in other cyanobacteria.

Within the putative *Prochlorococcus pho*X group, we also observe two similar but distinct subclades – cluster IVA and IVB (76.1% nucleotide and 82% amino acid similarity). Within these subclades, the *pho*X sequences are close to identical (> 99% nucleotide similarity). This suggests two distinct *pho*X lineages associated with HL *Prochlorococcus*.

The gene neighbourhood of these subclade variants suggests that *pho*X in *Prochlorococcus* is mostly located in a variable genomic island in proximity to other P acquisition and stress response genes (Fig. 1 and Fig. S1, and Table S1). We observe considerable variation among the neighbouring genes, suggesting that the gene neighbourhood is not conserved. Genes include the twin-arginine translocase protein (*tat*A) and several conserved hypothetical proteins and share high nucleotide sequence similarity to genes in a hypervariable genomic island in *Prochlorococcus* (labelled ISL5 in Coleman *et al*., 2006). Interestingly, many genes from ISL5 are upregulated under P stress and are predominantly found in low phosphate environments (Martiny *et al*., 2006; 2009a). We also detected several GOS fragments, where *pho*X was located on the same piece of DNA as genes from the genome region containing the P stress response regulator *pho*B in *Prochlorococcus* (as described in Martiny *et al*., 2006). Overall, it appears that *pho*X in *Prochlorococcus* share gene neighbourhood with other P stress response genes.

Another *pho*X clade (cluster I in Fig. 1 and Fig. S1) also has a low GC content (34.9%, $n = 28$). Here, paired-end sequences only share high nucleotide similarity with genes in SAR11 (*Pelagibacter*) genomes (Fig 1 and Fig. S1, and Table S1). Although no cultured representatives of the SAR11 clade contain *pho*X, the paired-end and GC content information indicate that this lineage also carry a *pho*X type alkaline phosphatase (see also Table S1). From hereon, we will refer to cluster I as SAR11 *pho*X. The other clades (cluster II, III, V and VI) did not have any paired-end matches so at present the origin is unknown.

Several *Prochlorococcus* strains contain a different predicted alkaline phosphatase that it is commonly referred to as *pho*A (e.g. *Prochlorococcus* MED4 PMM0708) (Moore *et al*., 2005; Martiny *et al*., 2006). However, it is important to note that this phosphatase is only distantly related to *pho*A in *E. coli* and we do not know the exact biochemistry of this enzyme (e.g. metal

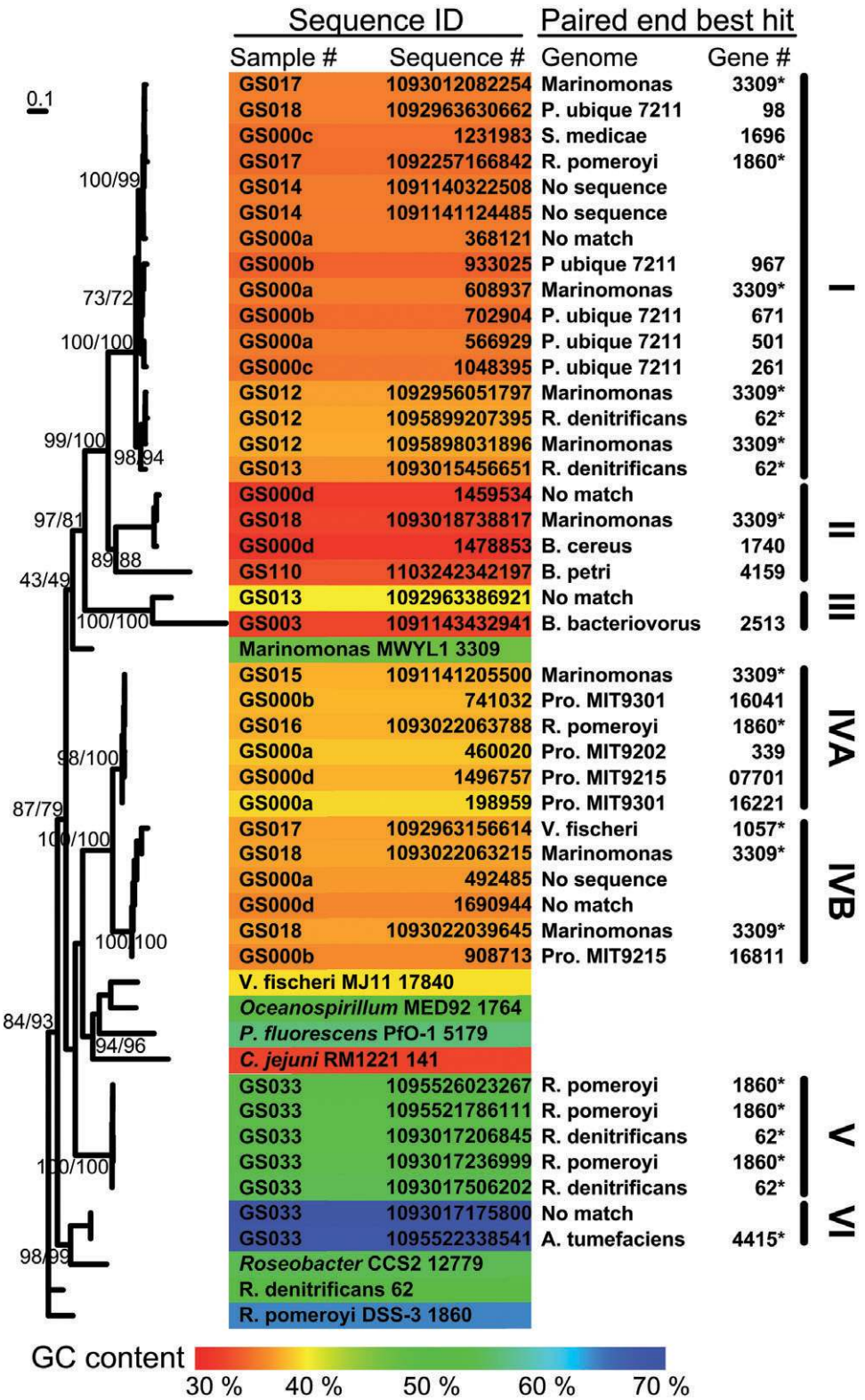

Fig. 1. Phylogeny of partial putative PhoX alkaline phosphate protein sequences inferred from GOS samples. Next to the sequence ID is listed the organism and locus most similar to the paired-end sequence mate. The sequence region corresponds to nucleotide position 814 to 1902 in *Marinomonas* MWYL1 3309. The phylogenetic tree of an additional region is presented in Fig. S1. The tree is calculated using a distance-based neighbour-joining method and bootstrap values are estimated using neighbour-joining and maximum likelihood. *Best match to *pho*X and therefore non-informative in terms of which lineage the gene is situated in.

dependence). Parallel to *pho*X, we also found that uncultured SAR11 cells likely contain this alkaline phosphatase, which was most closely related to the *pho*A type found in *Prochlorococcus* (Fig. 2 and Fig. S2). First, we screened for genes similar to *pho*A from *Prochlorococcus* MED4 and MIT9312. Not surprisingly, we found many sequences clustering with *pho*A from *Prochlorococcus* and with paired-end sequences matching *Prochlorococcus* genomes (cluster III in Fig. 2 and Fig. S2, and Table S1). However, we also observed two other clades closely related to *Prochlorococcus pho*A with an average GC content of 38.3% and 40.0% (cluster II and I in Fig. 2 and Fig. S2 respectively). The paired-end sequences from cluster II had many best matches to the genomes of *Pelagibacter ubique* strain HTCC1002, HTCC1062 and HTCC7211 (see also Table S1). This places a low GC version of *pho*A on genomic fragments associated with uncultured SAR11 cells. The second clade (cluster I) had slightly higher GC content compared with *pho*A in *Prochlorococcus* and *Pelagibacter*. For this clade, the paired-end sequences had best match to cyanophages similar to ones infecting *Prochlorococcus* (see also Table S1). This suggests that SAR11 cells as well as some cyanophages carry a *pho*A type alkaline phosphatase. An analysis of the neighbouring genes suggests that the phage version is located next to the cyanophage version of the phosphate uptake gene *pst*S (GC content approximately 42%). The slightly elevated GC content of *pho*A and neighbouring genes support a phage origin compared with *Prochlorococcus* as host genes involved in P uptake found in cyanophages can have higher GC content (Martiny *et al*., 2009a).

We also found that both *pho*A and *pho*X genes associated with *Prochlorococcus* and SAR11 were primarily found in cells from low P environments. We first enumerated the number of reciprocal best blast hits of each gene in all the GOS samples and normalized to the mean abundance of core genes (Martiny *et al*., 2009a; 2010). The ratio is an estimate of the fraction of cells within a population that carries the gene. For both *Prochlorococcus* and SAR11, we observe that *pho*A and *pho*X are primarily present in cells proliferating in ocean regions characterized by less than approximately 100 nM of DIP. This includes the Sargasso and Caribbean Sea (Fig. S3). Furthermore, *pho*X in *Prochlorococcus* is significantly less abundant compared with *pho*A (paired Student *t*-test, $P < 0.03$, $n = 36$), whereas we saw no significant difference in the occurrence of phosphatases in SAR11 populations. It is important to note that phosphate concentration values are monthly averages retrieved from the World Ocean Atlas and may therefore not represent the exact nutrient concentrations at the time of sampling.

Protein characterization

Next, we characterized the putative *pho*X associated with *Prochlorococcus* in terms of cellular location, pH optimum and metal dependence. First, we examined if the *pho*X GOS sequences contained a signal peptide and thus the protein was exported across the cytoplasmic membrane. This was done using signalP 3.0 (Emanuelsson *et al*., 2007). All GOS sequences covering the N-terminal part of the protein contained a predicted signal peptide for the twin-arginine translocation (tat) system. The same was the case for *pho*X in SAR11. This includes the two adjacent arginine residues in the signal peptide that are indicative of the export system. This result suggests that *pho*X is exported from the cytoplasm in both *Prochlorococcus* and SAR11 – presumably to the periplasm – and is consistent with other analyses of *pho*X from non-marine bacteria (Zaheer *et al*., 2009). Similarly, *pho*A associated with *Prochlorococcus*, SAR11 and cyanophages also contained a predicted sec-based signal peptide.

To confirm that the predicted *pho*X variant in *Prochlorococcus* encoded an active phosphatase, we cloned and heterologously expressed the gene into *E. coli*. As no GOS fragment covered the entire gene, we first estimated the consensus sequence (from cluster IVB) and then synthesized the consensus gene *in vitro*. For comparison, we also cloned and expressed *pho*X from *Synechococcus* WH8102 (SYNW1799). Both variants had strong activity under alkaline conditions with para-nitrophenolphosphate as substrate and thus encoded active alkaline phosphatases (Fig. S4). Specifically, PhoX associated with *Prochlorococcus* had an *in vitro* pH optimum close to 9.5, whereas the optimum of *Synechococcus* PhoX was 7.5. Noteworthy, this may not be the *in vivo* pH optimum for *Prochlorococcus* and *Synechococcus*. In other systems, it was shown that *in vitro* conditions can affect pH optimum (Ross *et al*., 1951).

To examine the metal dependence of PhoX, we first added EDTA to chelate any metal ions, which resulted in strong reduction of the enzyme activity for both phosphatase variants (Fig. 3). As predicted, these enzymes are likely metalloproteins with a strongly bound endogenous metal. Adding calcium back into the enzyme solution containing EDTA restored the enzyme activity. In contrast, adding other metals resulted in much lower activity (Fig. 3). Magnesium and copper additions restored some activity (< 15%), whereas no activity was observed by adding cobalt, manganese, nickel and zinc. This result suggests that PhoX primarily uses calcium ions in the active site, but can function with magnesium and copper. A detailed analysis of the protein sequences of PhoX in *Prochlorococcus* (IVA and IVB), *Synechococcus*, and SAR11 reveals that all proteins contained the proline and aspartate residues involved in binding calcium.

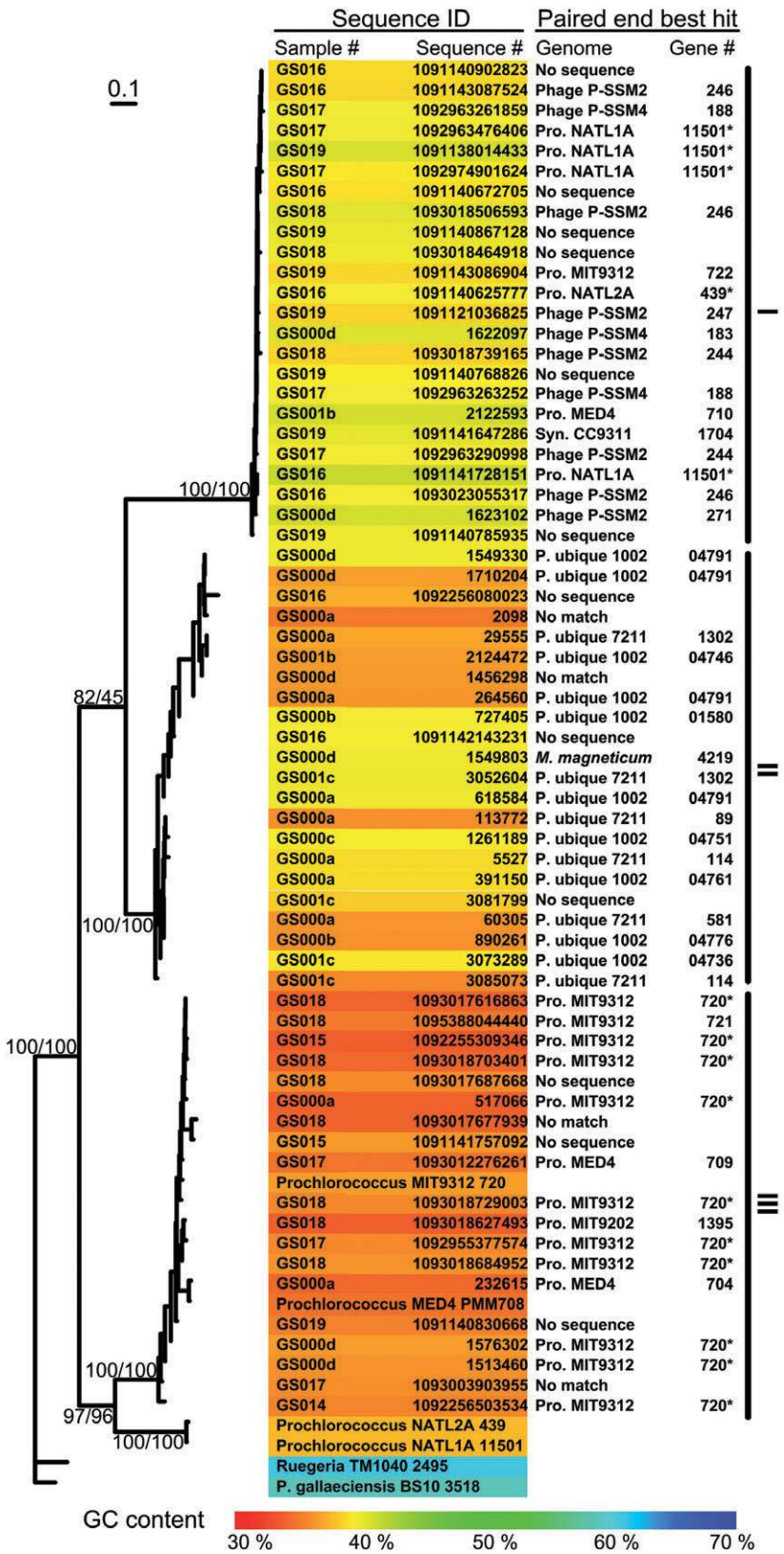

Fig. 2. Phylogeny of partial putative PhoA alkaline phosphate protein sequences inferred from GOS samples. Next to the sequence ID is listed the organism and locus most similar to the paired-end sequence mate. The sequence region corresponds to nucleotide position 1 to 672 in *Prochlorococcus* MED4 PMM708. The phylogenetic tree of an additional region is presented in Fig. S2. The tree is calculated using a distance-based neighbour-joining method and bootstrap values are estimated using neighbour-joining and maximum likelihood. *Best match to *pho*A and therefore non-informative in terms of which lineage the gene is situated in.

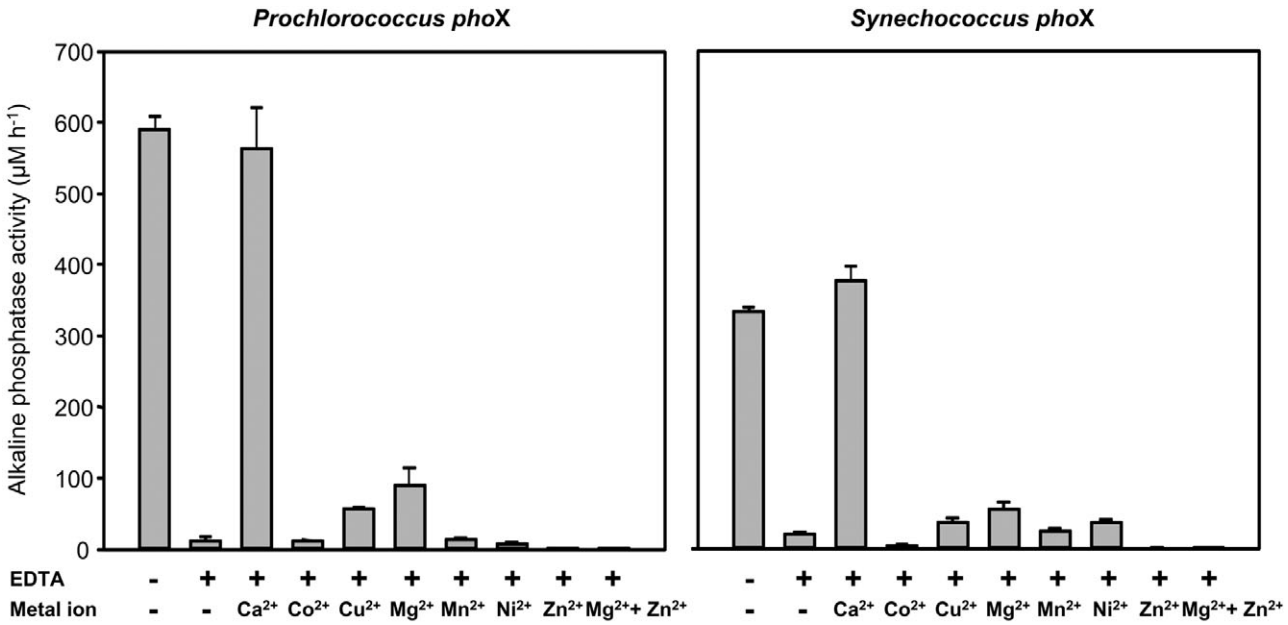

Fig. 3. Effect of various divalent metal cations on *Prochlorococcus* (GOS consensus) and *Synechococcus* WH8102 (SYNW1799) *pho*X alkaline phosphatase activity. Enzyme activity of a crude extract of heterologously expressed PhoX protein was measured colorimetrically using pNPP as substrate.

Previous work showed that certain metals including zinc can inhibit the activity of PhoX (Zaheer *et al*., 2009). To examine this in *Prochlorococcus* PhoX, we mixed various metals (but no EDTA) with the expressed protein (Fig. 4). As before, calcium enhanced the activity, but cobalt, manganese and zinc strongly inhibited activity. Copper and nickel had some inhibitory effect, whereas magnesium had little effect. To examine the relative competition between the metals, we added various ratios of zinc or nickel to calcium (0.1 to 50) (Fig. S5). At a 1:1 ratio, zinc completely inhibited the activity of *pho*X from both *Prochlorococcus* and *Synechococcus* (< 1% residual activity). In contrast, nickel had less effect on the activity. Furthermore, we observed that the inhibitory effects of both zinc and nickel ions were stronger on PhoX from *Prochlorococcus* than that from *Synechococcus*. This may result in a difference in the tolerance of elevated metal concentrations for these two lineages.

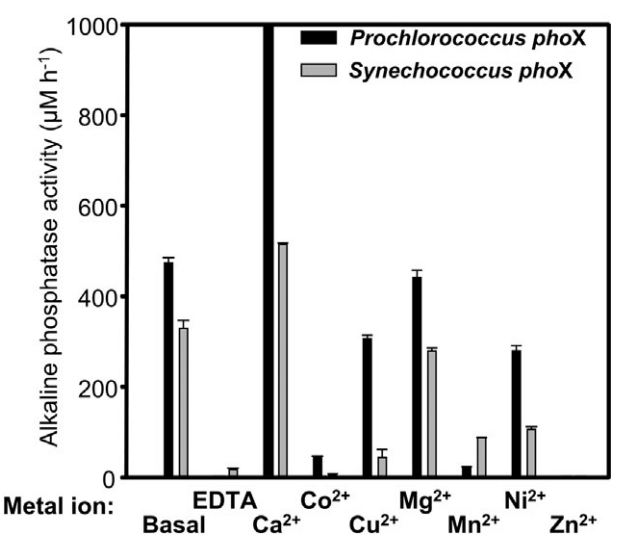

Fig. 4. Inhibition effect of various divalent metal cations on *Prochlorococcus* (GOS consensus) and *Synechococcus* WH8102 (SYNW1799) *pho*X alkaline phosphatase activity. Enzyme activity of a crude extract of heterologously expressed PhoX protein was measured colorimetrically using pNPP as substrate. Except in the treatment with EDTA, no EDTA was added in this assay.

Discussion

Here, the analysis of paired-end metagenomic sequences indicates that uncultured *Prochlorococcus* and SAR11 cells contain an active calcium-based alkaline phosphatase. Furthermore, the low GC content of these *pho*X variants is consistent with the GC content of these organisms. In addition, both lineages carry genes orthologous to *Prochlorococcus* MED4 gene PMM0708. Although no closely related genes have been biochemically characterized, this gene likely encodes an alkaline phosphatase based on (i) AP activity in *Prochlorococcus* strains with the gene and no activity in strains without, (ii) domain analysis, (iii) upregulated under P stress, (iv) always located in vicinity of the P regulators *pho*BR, (v) cellular location (only P stress protein with signal peptide), and (vi) only present in *Prochlorococcus* cells from low P environments (Moore *et al*., 2005; Martiny *et al*., 2006; 2009a; Scanlan

et al., 2009). It is commonly referred to as *pho*A but it is important to note that this phosphatase share little homology to *pho*A in *E. coli*, so it may or may not be zinc dependent. We attempted to heterologously clone and express *pho*A from *Prochlorococcus* MED4 to examine this issue but were unsuccessful.

Recent work suggests that alkaline phosphatases in marine bacteria were found in a variety of subcellular locations – including intracellular and extracellular (Luo *et al*., 2009). Our data clearly suggest that phosphatases in both *Prochlorococcus* and SAR11 carry a signal peptide and are exported from the cytoplasm (and thus not intracellular). This is consistent with other biochemical analyses of these enzymes. However, it is not clear if they get exported outside the cell as the biochemistry of protein export across the cell wall is not well understood in these lineages.

The phylogenetic analysis of both *pho*X and the putative *pho*A showed variants present in *Prochlorococcus* and SAR11 grouping together and related to phosphatases in other *Proteobacteria*. This is despite the distant relationship of these two lineages based on 16S rRNA and other core genes. In contrast, a *pho*X type found in many *Cyanobacteria*, including marine *Synechococcus* and *Trichodesmium*, belongs to a different subfamily of *pho*X and is distantly related to *pho*X in *Prochlorococcus* and SAR11 cells (Zaheer *et al*., 2009). Also, no other marine cyanobacteria strains so far carry a *pho*A-type phosphatase related to the one in *Prochlorococcus* and SAR11, and we did not detect any putative *Synechococcus* GOS fragments with a *pho*A gene. One scenario for these phylogenetic patterns is an ancient lateral gene transfer event for each gene from *Proteobacteria* into *Prochlorococcus*, but now *pho*A and *pho*X are fixed within the *Prochlorococcus* radiation. Evidence of lateral gene transfer of *pho*X has also been noted for other lineages (Zaheer *et al*., 2009). Beyond these ancient transfer events, there also appears to be significant gene gain and loss within the *Prochlorococcus* and SAR11 populations. First, we observe that *pho*A and *pho*X are only present in cells from ocean regions with less than approximately 100 nM of phosphate as we have seen with other P acquisition genes (Martiny *et al*., 2009a; 2010). These regions include the Sargasso and Caribbean Sea where DOP appears to be an important P source. Second, we show that *pho*X associated with *Prochlorococcus* appears to be located next to genes found in the variable genomic island ISL5 (Coleman *et al*., 2006), where many genes are upregulated under P stress and only found in cells from low P environments (Martiny *et al*., 2006; 2009a). Superimposed on the ancient transfer event, the contemporary pattern can either be explained by lateral gene transfer within either SAR11 and eMIT9312 *Prochlorococcus* populations or the loss of the genes in some lineages of SAR11 and *Prochlorococcus*. Either way, it suggests that gene gain or loss is important for adaptation to low phosphate availability in the ocean.

Previously, it was shown that some cyanophages infecting *Prochlorococcus* harbour the phosphate acquisition gene *pst*S – putatively to boost P uptake (Sullivan *et al*., 2005). We find that cyanophages also contain an alkaline phosphatase, suggesting this enzyme may assist in acquiring P during infection. The enzyme contains a predicted signal peptide and is likely exported to the periplasm. Thus, it appears that the phage-derived phosphatase assists in external DOP uptake rather than cleaving off host intracellular phosphate groups.

As described in *Introduction*, low P environments are often characterized by low dissolved zinc concentration (Jakuba *et al*., 2008), presumably due to the high biological demand for zinc as cofactor for the *pho*A type of alkaline phosphatase enzyme. Based on these observations, it has been suggested that phytoplankton may be subject to zinc-P co-limitation in these regions (Shaked *et al*., 2006; Jakuba *et al*., 2008). However, the widespread presence of a calcium-based alkaline phosphatase in abundant ocean bacteria like *Prochlorococcus*, *Synechococcus*, SAR11 and *Roseobacter* suggests that most bacteria here may not be prone to zinc-P co-limitation. *pho*X is also present in some eukaryotic photosynthetic organisms, including *Chlamydomonas*, *Micromonas* and *Volvox* (Quisel *et al*., 1996; Hallmann, 1999). This raises two questions. First, can cells from all major marine phytoplankton and bacterial lineages potentially carry a calcium-based phosphatase or are certain lineages restricted to a zinc-based type? Second, why do phytoplankton carry *pho*A as opposed to *pho*X altogether considering the much higher concentration of available calcium compared with zinc in the ocean? Perhaps *pho*A has a wider substrate spectrum, higher substrate affinity, or is less costly to produce. The answer to these questions will have important implications for our understanding of the role of zinc in regulating (or at least affecting) P uptake and more broadly phytoplankton activity and community structure.

Experimental procedures

Identification of pho*A and* pho*X in GOS samples*

We searched the Global Ocean Survey sequence database for *pho*A and *pho*X matching protein sequences from *Synechococcus* WH8102 (SYNW1799) and *Marinomonas* MWYL1 3309 for *pho*X and *Prochlorococcus* MED4 and MIT9312 for *pho*A using TBLASTN (*e*-value = 1E-5). Sample location and environment conditions are listed in Table S1. Next, each GOS hit and its paired-end sequence mate were compared with a database consisting of all sequenced genomes (as of 03/13/09) using both BLASTX and BLASTN

(*e*-value = 1E-30). Paired-end sequences are end-sequences from the same cloned fragment of DNA. GOS sequences matching phosphatases were exported and matching regions were translated. We aligned the protein sequences from GOS and bacterial genomes using the software Geneious (Biomatters, Australia) and carefully curated the alignments manually. Since GOS sequences were less than 1000 bp long, we split alignments of more than 800 bp to ensure that GOS sequences covered at least 75% of the region. Thus, we have included multiple phylogenetic trees for each gene. Phylogenetic trees of protein sequences (100 bootstraps) were estimated with Phylip v.3.67 using neighbour-joining (JTT matrix) and maximum likelihood (JTT model, 10 jumbles) (Felsenstein, 2006). Phylogenetic trees based on nucleotide alignments gave similar tree topology and supported specific clades associated with *Prochlorococcus*, SAR11 and cyanophages.

Distribution of pho*A and* pho*X*

The abundance of all *Prochlorococcus* and SAR11 genes were determined as previously described (Martiny *et al*., 2009a; 2010). First, we searched all GOS samples (Table S1) with TBLASTN (*e*-value 1E-6 and minimum length of 25 letters) to find hits matching any *Prochlorococcus* or SAR11 strain respectively. The translated GOS sequences of these hits were compared with a reference database of all sequenced genomes to detect reciprocal best hits. Second, we assigned each *Prochlorococcus* or SAR11 hit to a specific protein in MED4 or HTCC7211 respectively. A GOS hit was considered an orthologue if (i) it was assigned to *Prochlorococcus* or SAR11 and (ii) the reciprocal hit matched the original query protein when only the MED4 or HTCC7211 genome was searched. We only assigned one hit per GOS read for each query protein to accommodate errors such as frame shifting of the GOS reads. Finally, we estimated the abundance of *pho*A and *pho*X in *Prochlorococcus*, SAR11 and cyanophages by reciprocal best blast hit to a reference database containing all sequenced genomes (as of 03/31/2009) plus the uncultured variants of *pho*A and *pho*X. The relative abundance of each gene was calculated as the number of hits of a specific gene divided by the mean number of hits of single-copy core genes. All hits were normalized for gene length. A maximum likelihood estimate of the mean abundance of *Prochlorococcus* or SAR11 genes and variance for each sample is calculated by fitting the number of hits of single-copy core genes ($n_{Prochlorococcus} = 1209$ or $n_{SAR11} = 1075$) to a gamma distribution using Matlab (Mathworks, MA).

Heterologous cloning and expression of pho*X*

We examined the function of *pho*X in *Prochlorococcus* and *Synechococcus* WH8102 by cloning and expressing each gene in *E. coli* and testing for phosphatase activity. First, we estimated the consensus sequence of the putative *Prochlorococcus pho*X (cluster IVB) based on GOS sequences and then synthesized the gene *in vitro* (Genscript Corporation, NJ). The *pho*X sequences are provided as supplementary information (File S1). It was then cloned into a pET vector (26b). We also PCR-amplified and cloned *Synechococcus* WH8102 gene SYNW01799 (*pho*X) and cloned the gene into pEcoli-Nterm 6xHN (Clontech, Mountain View, CA, USA). As a control, we used *nar*B cloned into pEcoli-Nterm 6xHN (Martiny *et al*., 2009b). After verifying the inserts by sequencing, we transformed *E. coli* BL21 cells (Invitrogen, Carlsbad, CA, USA) with each plasmid. We initiated expression by adding 1 mM IPTG to *E. coli* BL21D3-pLysS (Novagen, NJ, USA) (including plasmid) growing in LB with 1% glucose, 100 mM phosphate and 100 mg l^{-1} ampicillin at 23°C. A crude extract was made by harvesting 50 ml induced BL21 cells (OD 0.5) containing PhoX and incubated with CelLytic Express solution (Sigma-Aldrich, MO, USA) for 10 min.

To find the pH optimum, we incubated the crude extract in 100 mM phosphate buffer, 0.1 mM calcium and 4 mM para-nitrophenolphosphate (pNPP). At time 0, 15, 30, 45, 60 and 90 min, the reaction was stopped by adding 500 mM NaOH and phosphatase activity was measured colorimetrically at 405 nm. The phosphatase activity was lower in the phosphate buffer (potentially due to some inhibition), but this buffer system covered the full pH range for the experiment. However, in the subsequent experiments we used a high pH carbonate buffer to avoid any inhibition.

To determine the metal dependence of PhoX, we first incubated the crude extract with 0.025 mM EDTA for 2 min and 100 mM carbonate buffer (pH 8.0 for SYNW1799 and pH 9.5 for *Prochlorococcus* PhoX). We also tested *E. coli* with NarB as control but observed very low activity. We next added 4 mM pNPP and 0.5 mM Ca^{2+}, Co^{2+}, Cu^{2+}, Mg^{2+}, Mn^{2+}, Ni^{2+}, Zn^{2+} and $Zn^{2+} + Mg^{2+}$, and compared with the activity of no addition or only EDTA addition. To explore any metal inhibition, we added 0.5 mM Ca^{2+}, Co^{2+}, Cu^{2+}, Mg^{2+}, Mn^{2+}, Ni^{2+}, Zn^{2+} and $Zn^{2+} + Mg^{2+}$ to the crude extract (so no EDTA added), and compared with the activity of no addition or only EDTA added. Finally, we added 0.1 mM Ca^{2+} to the crude extract in 100 mM carbonate buffer (again no EDTA) and varying concentrations of either Zn^{2+} or Ni^{2+} (0.01 mM to 5 mM) to further test inhibition. Phosphatase activity was measured colorimetrically at 405 nm every 2 min for 60 min for all metal assays and converted to $\mu M\ h^{-1}$ using an extinction of coefficient of $1.78 \times 10^4\ M^{-1}\ cm^{-1}$. We always measured pH at the end of each assay as control. All concentrations are final.

Acknowledgements

We would like to thank Gabrielle Rocap and Jennifer Martiny for many helpful comments and the National Science Foundation for supporting the work (OCE- 0928544).

References

Boyer, T.P., Antonov, J.I., Garcia, H.E., Johnson, D.R., Locarnini, R.A., Mishonov, A.V., *et al*. (2006) *World Ocean Database 2005*. Washington, DC, USA: US Government Printing Office.

Cavender-Bares, K.K., Karl, D.M., and Chisholm, S.W. (2001) Nutrient gradients in the western North Atlantic Ocean: relationship to microbial community structure, and comparison to patterns in the Pacific Ocean. *Deep Sea Res I* **48:** 2373–2395.

Coleman, J.E. (1992) Structure and mechanism of alkaline phosphatase. *Annu Rev Biophys Biomol Struct* **21:** 441–483.

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

Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2:** 953–971.

Felsenstein, J. (2006) *PHYLIP (Phylogeny Inference Package)*. Seattle, WA, USA: Department of Genome Sciences, University of Washington.

Hallmann, A. (1999) Enzymes in the extracellular matrix of *Volvox*: an inducible, calcium-dependent phosphatase with a modular composition. *J Biol Chem* **274:** 1691–1697.

Jakuba, R.W., Moffett, J.W., and Dyhrman, S.T. (2008) Evidence for the linked biogeochemical cycling of zinc, cobalt, and phosphorus in the western North Atlantic Ocean. *Global Biogeochem Cycles* **22:** 6B4012. doi: 4010.1029/2007GB003119.

Karl, D.M., and Björkman, K.M. (2002) Dynamics of DOP. In *Biogeochemistry of Marine Dissolved Organic Matter*. Hansell, D.A., and Carlson, C.A. (eds). London, UK: Academic Press, pp. 250–366.

Kolowith, L.C., Ingall, E.D., and Benner, R. (2001) Composition and cycling of marine organic phosphorus. *Limnol Oceanogr* **46:** 309–320.

Lomas, M.W., Burke, A.L., Lomas, D.A., Bell, D.W., Shen, C., Dyhrman, S.T., and Ammerman, J.W. (2010) Sargasso Sea phosphorus biogeochemistry: an important role for dissolved organic phosphorus (DOP). *Biogeosciences* **7:** 695–710.

Luo, H.W., Bennera, R., Long, R.A., and Hu, J.J. (2009) Subcellular localization of marine bacterial alkaline phosphatases. *Proc Natl Acad Sci USA* **106:** 21219–21223.

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

Martiny, A.C., Huang, Y., and Li, W.Z. (2009a) Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions. *Environ Microbiol* **11:** 1340–1347.

Martiny, A.C., Kathuria, S., and Berube, P.M. (2009b) Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proc Natl Acad Sci USA* **106:** 10787–10792.

Martiny, A.C., Huang, Y., and Li, W. (2010) Adaptation to nutrient availability in marine microorganisms by gene gain and loss. In *Handbook of Molecular Microbial Ecology II: Metagenomics in Different Habitats*. de Bruijn, F. (ed.). Oxford, UK: Wiley/Blackwell (in press).

Moore, J.K., Doney, S.C., Glover, D.M., and Fung, I.Y. (2002) Iron cycling and nutrient-limitation patterns in surface waters of the World Ocean. *Deep Sea Res II* **49:** 463–507.

Moore, L.R., Ostrowski, M., Scanlan, D.J., Feren, K., and Sweetsir, T. (2005) Ecotypic variation in phosphorus acquisition mechanisms within marine picocyanobacteria. *Aquat Microb Ecol* **39:** 257–269.

Paytan, A., and McLaughlin, K. (2007) The oceanic phosphorus cycle. *Chem Rev* **107:** 563–576.

Quisel, J.D., Wykoff, D.D., and Grossman, A.R. (1996) Biochemical characterization of the extracellular phosphatases produced by phosphorus-deprived *Chlamydomonas reinhardtii*. *Plant Physiol* **111:** 839–848.

Ross, M.H., Ely, J.O., and Archer, J.G. (1951) Alkaline phosphatase activity and ph optima. *J Biol Chem* **192:** 561–568.

Roy, N.K., Ghosh, R.K., and Das, J. (1982) Monomeric alkaline phosphatase of *Vibrio cholerae*. *J Bacteriol* **150:** 1033–1039.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., *et al*. (2007) The sorcerer II global ocean sampling expedition: Northwest Atlantic through eastern Tropical Pacific. *PLoS Biol* **5:** e77.

Sanudo-Wilhelmy, S.A., Kustka, A.B., Gobler, C.J., Hutchins, D.A., Yang, M., Lwiza, K., *et al*. (2001) Phosphorus limitation of nitrogen fixation by *Trichodesmium* in the central Atlantic Ocean. *Nature* **411:** 66–69.

Scanlan, D.J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W.R., *et al*. (2009) Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev* **73:** 249–299.

Sebastian, M., and Ammerman, J.W. (2009) The alkaline phosphatase PhoX is more widely distributed in marine bacteria than the classical PhoA. *ISME J* **3:** 563–572.

Shaked, Y., Xu, Y., Leblanc, K., and Morel, F.M.M. (2006) Zinc availability and alkaline phosphatase activity in *Emiliania huxleyi*: implications for Zn-P co-limitation in the ocean. *Limnol Oceanogr* **51:** 299–309.

Sullivan, M.B., Coleman, M.L., Weigele, P., Rohwer, F., and Chisholm, S.W. (2005) Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biol* **3:** e144.

Tetu, S.G., Brahamsha, B., Johnson, D.A., Tai, V., Phillippy, K., Palenik, B., and Paulsen, I.T. (2009) Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp WH8102. *ISME J* **3:** 835–849.

Thingstad, T.F., Krom, M.D., Mantoura, R.F., Flaten, G.A., Groom, S., Herut, B., *et al*. (2005) Nature of phosphorus limitation in the ultraoligotrophic eastern Mediterranean. *Science* **309:** 1068–1071.

Torriani-Gorini, A. (1987) The birth and growth of the Pho regulon. In *Phosphate Metabolism and Cellular Regulation in Microorganisms*. Torriani-Gorini, A., Rothman, F.G., Silver, S., Wright, A., and Yagil, E. (eds). Washington, DC, USA: American Society for Microbiology, pp. 3–11.

Wanner, B.L. (1996) Phosphorus assimilation and control of the phosphate regulon. In Escherichia coli *and* Salmonella*: Cellular and Molecular Biology*. Neidhardt, F.C. (ed.). Washington, DC, USA: American Society of Microbiology, pp. 1357–1381.

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

Young, C., and Ingall, E. (2010) Marine dissolved organic phosphorus composition: insights from samples recovered using combined electrodialysis/reverse osmosis. *Aquat Geochem* **16:** 563–574.

Zaheer, R., Morton, R., Proudfoot, M., Yakunin, A., and Finan, T.M. (2009) Genetic and biochemical properties of an alkaline phosphatase PhoX family protein found in many bacteria. *Environ Microbiol* **11:** 1572–1587.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Phylogeny of partial putative PhoX alkaline phosphate protein sequences inferred from GOS samples. Next to the sequence ID is listed the organism and locus most similar to the paired-end sequence mate. The sequence region corresponds to nucleotide position 1 to 835 in *Marinomonas* MWYL1 3309. The tree is calculated using a distance-based neighbour-joining method and bootstrap values are estimated using neighbour-joining and maximum likelihood. *Best match to *pho*X and therefore non-informative in terms of which lineage the gene is situated in.

Fig. S2. Phylogeny of partial putative PhoA alkaline phosphate protein sequences inferred from GOS samples. Next to the sequence ID is listed the organism and locus most similar to the paired-end sequence mate. The sequence region corresponds to nucleotide position 673 to 1338 in *Prochlorococcus* MED4 PMM0708. The tree is calculated using a distance-based neighbour-joining method and bootstrap values are estimated using neighbour-joining and maximum likelihood. *Best match to *pho*A and therefore non-informative in terms of which lineage the gene is situated in.

Fig. S3. Relationship between phosphate concentration and average occurrence in GOS samples of *phoA and phoX* genes putatively associated with *Prochlorococcus* and SAR11. Abundance of individual genes is determined by reciprocal best BLAST hit and normalized against length. The frequency is calculated as the length normalized occurrence of a specific gene divided by the length normalized mean occurrence of single copy core *Prochlorococcus* MED4 and *Pelagibacter ubique* HTCC7211 genes at each site (Martiny *et al*., 2009a; 2010). Phosphate concentration monthly average values for each site are retrieved from the World Ocean Database (Boyer *et al*., 2006).

Fig. S4. pH optimum of *Prochlorococcus* (GOS consensus), *Synechococcus* WH8102 (SYNW1799) *pho*X alkaline phosphatase activity. The activity of *E. coli* containing *nar*B was used as control. Enzyme activity of a crude extract of heterologously expressed protein was measured colorimetrically using pNPP as substrate.

Fig. S5. Relative inhibition effect of zinc and nickel divalent cations on *Prochlorococcus* (GOS consensus) and *Synechococcus* WH8102 (SYNW1799) *pho*X alkaline phosphatase activity. Enzyme activity of a crude extract of heterologously expressed PhoX protein was measured colorimetrically using pNPP as substrate. We added 0.1 mM of calcium and then 0.01–5 mM of zinc and nickel divalent cations to test the relative inhibition effect compared with calcium. No EDTA was added in this assay.

Table S1. List of GOS fragments and their paired-end mates associated with *pho*A and *pho*X genes in *Prochlorococcus*, SAR11 and cyanophages. [#]Genes associated with the genomic island ISL5 in *Prochlorococcus*(*sensu* Coleman *et al*., 2006). [&]Genes associated with the pho regulon in *Prochlorococcus* (Martiny *et al*., 2006).

File S1. GOS consensus sequences of *phoA* and *phoX* in *Prochlorococcus*, SAR11 and cyanophages.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.