The ISME Journal (2007) 1, 321–330

www.nature.com/ismej

ORIGINAL ARTICLE

Distribution and diversity of phytate-mineralizing bacteria

Boon Leong Lim[1], Pok Yeung[1], Chiwai Cheng[1] and Jane Emily Hill[2]
[1]*School of Biological Sciences, The University of Hong Kong, Pokfulam, Hong Kong, China and* [2]*Department of Civil and Environmental Engineering, University of Vermont, Burlington, VT, USA*

Phytate, the most abundant organic phosphorus compound in soil, dominates the biotic phosphorus input from terrestrial runoffs into aquatic systems. Microbial mineralization of phytate by phytases is a key process for recycling phosphorus in the biosphere. Bioinformatic studies were carried out on microbial genomes and environmental metagenomes in the NCBI and the CAMERA databases to determine the distribution of the four known classes of phytase in the microbial world. The β-propeller phytase is the only phytase family that can be found in aquatic environments and it is also distributed in soil and plant bacteria. The β-propeller phytase-like genes can be classified into several subgroups based on their domain structure and the positions of their conserved cysteine residues. Analysis of the genetic contexts of these subgroups showed that β-propeller phytase genes exist either as an independent gene or are closely associated with a TonB-dependent receptor-like gene in operons, suggesting that these two genes are functionally linked and thus may play an important role in the cycles of phosphorus and iron. Our work suggests that β-propeller phytases play a major role in phytate-phosphorus cycling in both soil and aquatic microbial communities.
The ISME Journal (2007) **1,** 321–330; doi:10.1038/ismej.2007.40; published online 7 June 2007
Subject Category: integrated genomics and post-genomics approaches in microbial ecology
Keywords: β-propeller phytase; phosphorus; phytase; phytate; TonB-dependent receptor

Introduction

Most of the phosphorus (P) in terrestrial ecosystems is located in the soil. Globally, terrestrial biota holds 2.6×10^9 mg P, which is far less than that contained in the soil ($96–160 \times 10^9$ mg of P) (Stevenson and Cole, 1999). The major input of P from biota into soil is organic P compounds from plants, animals and microbes, which account for 30–65% of total soil P. Organic P (Po) compounds are diverse and their mineralization in soil allows the recycle of P back to biota. Various Po compounds have very different rate of mineralization. For example, Po from microbes – predominantly nucleic acids (30–50% P in RNA and 5–10% P in DNA) and phospholipids (<10% P), are rapidly mineralized in soil environments (Macklon *et al.*, 1997). Other Po compounds, however, are not mineralized easily and can accumulate in soil to a substantial amount. The most significant of these is myo-inositol hexakisphosphate (phytate), which may constitute up to 80% of Po in soil (Turner *et al.*, 2002).

Correspondence: BL Lim, School of Biological Sciences, University of Hong Kong, Pokfulam, Hong Kong, China.
E-mail: bllim@hku.hk
Received 16 February 2007; revised and accepted 27 April 2007; published online 7 June 2007

Phytate contributes up to 65–80% of the total P in grains (Lott *et al.*, 2000) and up to 80% of total P in manures from monogastric animals (Barnett, 1994). Owing to its high negative charge, phytate is tightly adsorbed to various soil components once it is released from plant residues or manures (Lung and Lim, 2006). The accumulation of phytate in soil is attributed to its invulnerability to phytase hydrolysis (Tang *et al.*, 2006). The breakdown of phytate requires the binding of free phytate to the substrate-binding pocket of phytases; hence, once it is tightly bound to soil components, it is insusceptible to direct enzyme hydrolysis.

Phytate from terrestrial runoff is the major external source of Po to aquatic systems (Suzumura and Kamatani, 1995a). Phytate can be detected in preconcentrated lake waters (Eisenreich and Armstrong, 1977), river waters (Stevens and Stewart, 1982), marine waters (Matsuda *et al.*, 1985), lake sediments (Golterman *et al.*, 1998) and marine sediments (Suzumura and Kamatani, 1995b). In addition, it constitutes approximately 20% of total organic P in soil leachates (Toor *et al.*, 2003). However, unlike terrestrial environments, phytate is rapidly mineralized upon introduction to a marine environment (Suzumura and Kamatani, 1995a), possibly owing to its enhanced solubility. Phytases play a crucial role in the recycling of Po in the aquatic biosphere. Up to 50% of the filterable,

nonreactive P harvested from preconcentrated lake water is exclusively hydrolyzed by phytase (Herbes *et al.*, 1975). For sediment P, up to 34% of residual Po can be hydrolyzed by phytase, while less than 5% of Po can be hydrolyzed by alkaline phosphatase (deGroot and Golterman, 1993).

While phytate is a major source of P input into the aquatic system and its life is ephemeral (Suzumura and Kamatani, 1995b), the identity of the organisms and their phytases that are responsible for this important biological process remain obscure. To date, four classes of phytases have been characterized in terrestrial organisms: histidine acid phosphatase (HAP), cysteine phytase (CPhy), purple acid phosphatase (PAP) and β-propeller phytase (BPP) (Mullaney and Ullah, 2003; Chu *et al.*, 2004). Although the biochemical properties of these four classes of phytases have been ascertained, their distribution in the natural environment, including aquatic systems, is poorly understood. To gain insight into the distribution of these four classes of phytases in various habitats, representative genes from the four classes of phytases were used to probe all the microbial and environmental sequence databases available in the National Center for Biotechnology Information (NCBI) and the Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis (CAMERA, Seshadri *et al.*, 2007).

Materials and methods

Distribution of phytase genes in microbial genomes
Representative genes from the four classes of phytases, including HAP (*Escherichia coli AppA*, Genbank accession no. P07102; *Aspergillus niger* PhyA and PhyB, P34752 and P34754), BPP (*Bacillus subtilis 168PhyA*, CAB13871; *Shewanella oneidensis* PhyS, AAN55555; *Xanthomonas oryzae* PhyA, YP_201138), PAP (*Burkholderia cepacia*, ZP_00215284; Soybean GmPhy, AAK49438) and CPhy (*Selenomonas ruminantium*, AAQ13669), were used as probes to BLAST microbial genome databases from the NCBI and the Moore Marine Microbial Genome Sequencing Project (www.moore.org/microgenome/). All of the genes that were used as probes were shown to exhibit phytase activity, except the microbial PAP from *Burkholderia* sp, which has yet to be characterized. An expected value threshold (*E* value) at $<E^{-10}$ over the entire sequence length and a minimal number of 30 amino-acid sequence identity were employed as the criteria for the BLAST searches.

Biochemical characterization of recombinant PAP from Burkholderia cenocepacia *J2315*
The PAP gene from *B. cenocepacia* J2315 (http://www.sanger.ac.uk/Projects/B_cenocepacia/) was amplified from its genomic sequence by PCR and subcloned into the expression vector pGex2T (GE Healthcare Life Sciences, Hong Kong, China). Recombinant plasmid pGEX–PAP was transformed and expressed in *E. coli* BL21. After IPTG induction, the GST–PAP fusion protein in the soluble fraction was purified by glutathione-sepharose 4B affinity column chromatography (GE Healthcare Life Sciences). Recombinant PAP was cleaved from its GST fusion partner by thrombin and was purified to homogeneity by Hi-trap Q anion exchange chromatography (GE Healthcare Life Sciences). Phytase and phosphatase activity of the enzyme was measured using phytate and *p*-nitrophenol phosphate as substrates, respectively, as described previously (Lung and Lim, 2006).

Phylogenetic analysis
Positive hits with expected value thresholds (*E* value) at $<E^{-10}$ were retrieved and analyzed. Only bacterial sequences that carry the conserved motifs of each class of phytases are included, thus excluding fungal, animal and plant genomes from the search. PAP sequences were excluded from the subsequent analysis, as overexpressed *Burkholderia* PAP did not exhibit phytase activity (data not shown). Sequence results were aligned using ClustalW program (Thompson *et al.*, 1994) and a phylogenetic tree for each phytase class was constructed using the MEGA 3.1 program (Kumar *et al.*, 2004). Robustness of the trees was evaluated by 1000 bootstrap repetitions. Protein sequence identity was calculated by the Smith–Waterman algorithm (Smith and Waterman, 1981).

Distribution of phytase genes in environmental sequence databases
All the verified sequences employed for the construction of HAP, CPhy and BPP phylogenetic trees (Figures 1–3) were used as probes to BLAST the environmental sequence databases from the NCBI and the CAMERA (http://camera.calit2.net/). Two of the bacterial databases in CAMERA, the Global Ocean Sampling (GOS) Expedition data set and the Microbial Community Genomics at the Hawaii Ocean Time-series (HOT) station, were selected for analysis. The GOS data set presents the results of the first phase of the GOS expedition, which collected microbes from 41 sites located in and around North and Central America and produced in total 7.7 million sequencing reads by shotgun sequencing (Rusch *et al.*, 2007). The database collected from the HOT station contained the end sequences of seven fosmid libraries, which were constructed from microbial species collected from surface waters to 4 km deep at the North Pacific subtropical gyre (DeLong *et al.*, 2006).

To validate and benchmark the BLAST results, the alkaline phosphatase gene (Genebank accession no. AAA83893) from *E. coli* was also included in each BLAST search. Alkaline phosphatase is regarded as ubiquitous in living organisms and is able to dephosphorylate a wide range of Po compounds but not generally phytate (Vincent and Crowder, 1995).

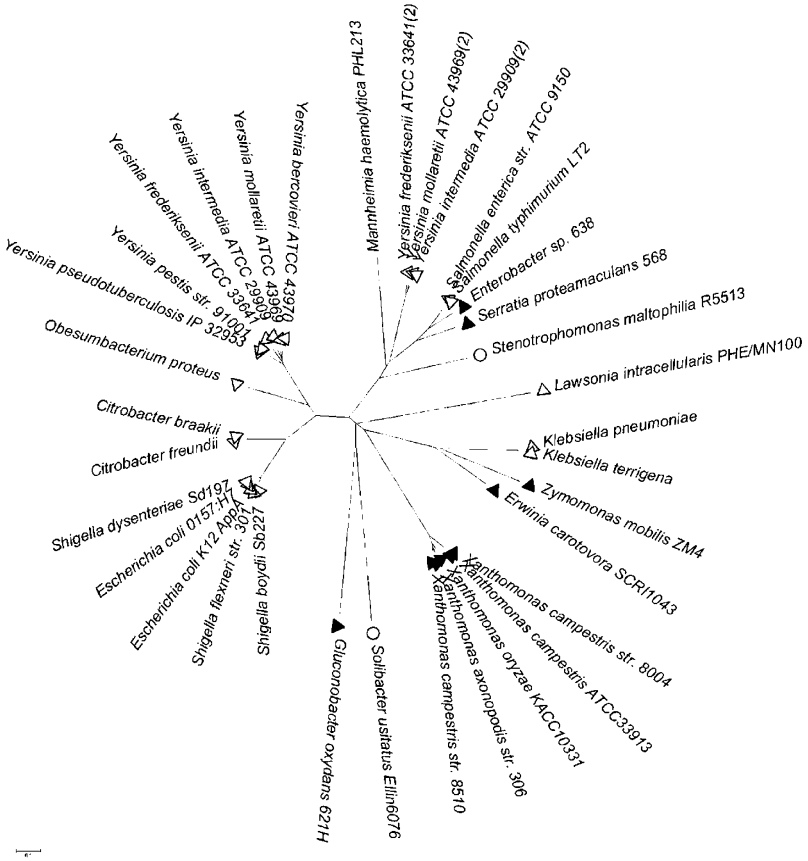


Figure 1 Phylogenetic analysis of prokaryotic HAP-like sequences. The symbols in front of the name of the bacterial strains represent the habitats: plant (▲) gastrointestinal tract (△) and free-living (○): *Citrobacter braakii*, AAS45884; *C. freundii*, AAR89622; *Enterobacter* sp 638, ZP_01590263; *E. coli* K12, AAN28334; *Erwinia carotovora* SCRI1043, YP_052285; *E.* 0157:H7, NP_309163; *Gluconobacter oxydans* 621H, YP_191501; *Klebsiella pneumoniae*, AAM23271; *K. terrigena*, CAE01322; *Lawsonia intracellularis* PHE/MN100, YP_594814; *Mannheimia_haemolytica*_PHL213; *Obesumbacterium proteus*, AAQ90419; *Salmonella enterica* ATCC9150, YP_150966; *S. typhimurium* LT2, NP_460090; *Serratia proteamaculans* 568, ZP_01534140; *Shigella boydii* Sb227, YP_408643; *S. dysenteriae* Sd197, YP_402619; *S. flexneri* str. 301, NP_706903; *Solibacter usitatus* Ellin6076, ZP_00521993; *Stenotrophomonas maltophilia* R5513 ZP_01643788; *Xanthomonas axonopodis* str. 306, AAM35446; *X. campestris* str. 8004, YP_244537; *X. campestris* str. 85–10, CAJ22221; *X. campestris* ATCC 33913, NP_636151; *X. oryzae* KACC10331, YP_202431; *Yersinia bercovieri* ATCC 43970, ZP_00821164; *Y. frederiksenii* ATCC 33641, ZP_00831300; *Y. frederiksenii* ATCC 33641 (2), ZP_00831059; *Y. intermedia* ATCC 29909, ZP_00832361; *Y. intermedia* ATCC 29909 (2), ZP_00834492; *Y. mollaretii* ATCC 43969, ZP_00824387; *Y. mollaretii* ATCC 43969 (2), ZP_00824103; *Y. pestis* str. 91001, NP_993128; *Y. pseudotuberculosis* IP 32953, YP_070934; *Zymomonas mobilis* ZM4, YP_161796.

Molecular modeling of BPP domains

For molecular modeling, the folds of the BPP-like proteins were predicted using the 3D-PSSM protein-fold recognition server (Kelley *et al.*, 2000). The top-ranking model, BaPhy (Protein data bank ID, 1POO), was used as a template to model the three-dimensional structures of the BPP domains using the optimized mode of the Swiss-Model comparative protein modeling server (Schwede *et al.*, 2003) and based on the structural alignments between the BaPhy and these BPP domains that were obtained from the 3D-PSSM server. The models were viewed using the Swiss-PdbViewer (Schwede *et al.*, 2003).

Genomic organization of microbial phytase genes

The genes that are flanking the BPP-like genes in the microbes shown in Figure 3 were identified by analyzing the NCBI and the Kyoto Encyclopedia of Genes and Genomes (KEGG) databases (http://www.genome.jp/kegg/). The destinations of the phytase expression in the bacteria were predicted by the PSORT server (Nakai and Horton, 1999).

Results and discussion

Distribution of three classes of phytases in bacterial genomes

In April 2007, there were 800 and 75 microbial genomes available from the NCBI and the Moore Marine Microbial Genome Sequencing Project, respectively. To be recorded as a positive count, the recovered sequences from the databases must carry the conserved motifs of each class of phytase and have an E value smaller than E^{-10}. The conserved motifs for HAP and CPhy are RH(G/N) XRXP (Mullaney and Ullah, 2003; Herter *et al.*,

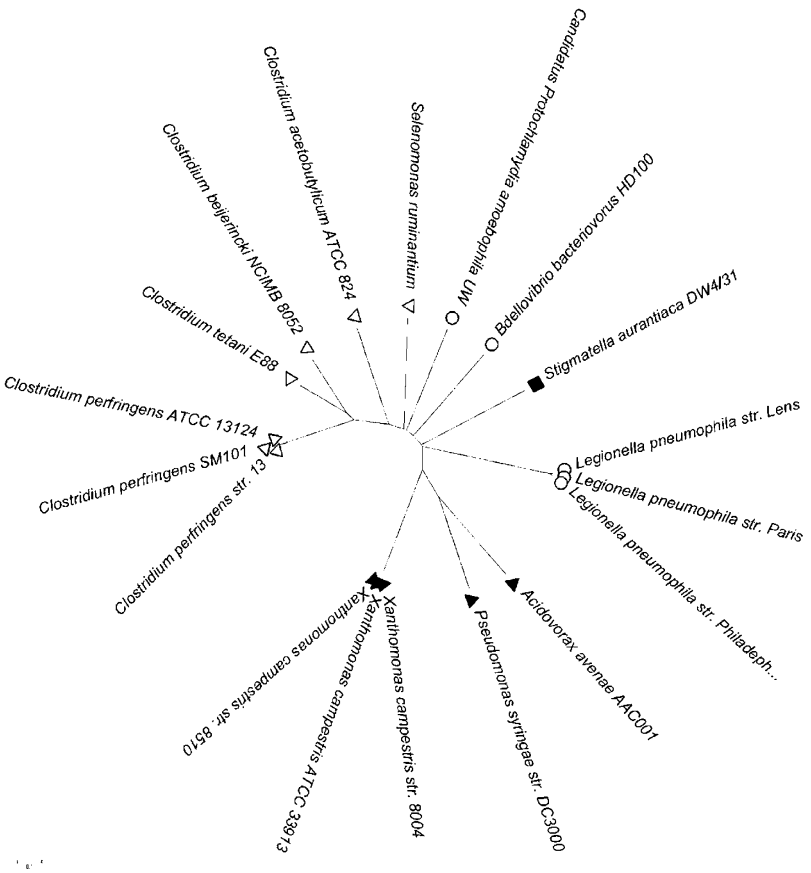


Figure 2 Phylogenetic analysis of prokaryotic CPhy-like sequences. The symbols in front of the name of the bacterial strains represent the habitats: plant (▲), gastrointestinal tract (△) and free-living (○): *Acidovorax avenae* AAC00-1, YP_971831; *Bdellovibrio bacteriovorus* HD100, NP_968118; *Clostridium acetobutylicum* ATCC 824, NP_149178; *C. beijerincki* NCIMB 8052, ZP_00910765; *C. perfringens* str. 13, NP_562440; *C. perfringens* ATCC13124, YP_696211; *C. perfringens* SM101, NC_008262; *C. tetani* E88, NP_782216; *Legionella pneumophila* str. Lens, YP_128063; *L. pneumophila* str. Paris, YP_125176; *L. pneumophila* str. Philadephia, YP_096814; *Candidatus Protochlamydia amoebophila* UWE25, YP_008827; *Pseudomonas syringae* str. DC3000, NP_794465; *Selenomonas ruminantium*, AAQ13669; *Stigmatella aurantiaca* DW4/3-1, EAU66412; *Xanthomonas campestris* str. 8004, YP_243159; *X. campestris* str. 85-10, YP_361664; *X. campestris* ATCC 33913, NP_637463.

2006) and HCXXGXXR(T/S) (Chu *et al.*, 2004), respectively. In BPPs, amino-acid residues that constitute the catalytic sites (three calcium ion-binding sites) are scattered over the protein sequence; hence, no conserved motif could be used as a probe.

Multiple HAP, CPhy and BPP sequences were retrieved from the NCBI microbial genome database, but only BPP sequences could be identified from the Moore Marine Microbial Genome Sequencing Project. Phylogenetic trees of these three classes of phytases were then produced from the positive hits. As shown in Figure 1, HAP-like sequences were mainly found in plant pathogenic and enteric bacteria. These habitats often have ample phytate present and to access the phytate and possibly the carbon in the phytate, bacteria require a phytate-degrading enzyme. Similarly, CPhy-like sequences were found in plant pathogenic, enteric and free-living bacteria (Figure 2). No HAP or CPhy-like sequences were found in the genomes of sequenced aquatic bacteria. In contrast, BPP-like sequences were found in aquatic bacteria as well as soil and plant bacteria (Figure 3).

BPP is the dominant class of phytase in aquatic systems

All the sequences shown in Figures 1–3 were used as probes in BLAST search. No positives hits of HAP, CPhy and BPP were identified from the HOT database, presumably owing to the small size of the database – only 7–11 Mbp of microbial genome sequence from each depth was available. In contrast, the GOS database, which contained 7.7 million independent DNA sequences extracted from microbes collected at 41 aquatic habitats, produced 6, 15 and 589 hits of HAP-, CPhy- and BPP-like sequences, respectively. However, only one of the HAP-like sequences and none of the CPhy-like sequences carried the conserved active sites of the enzymes. Therefore, whether these sequences encode functional phytases is an open question.

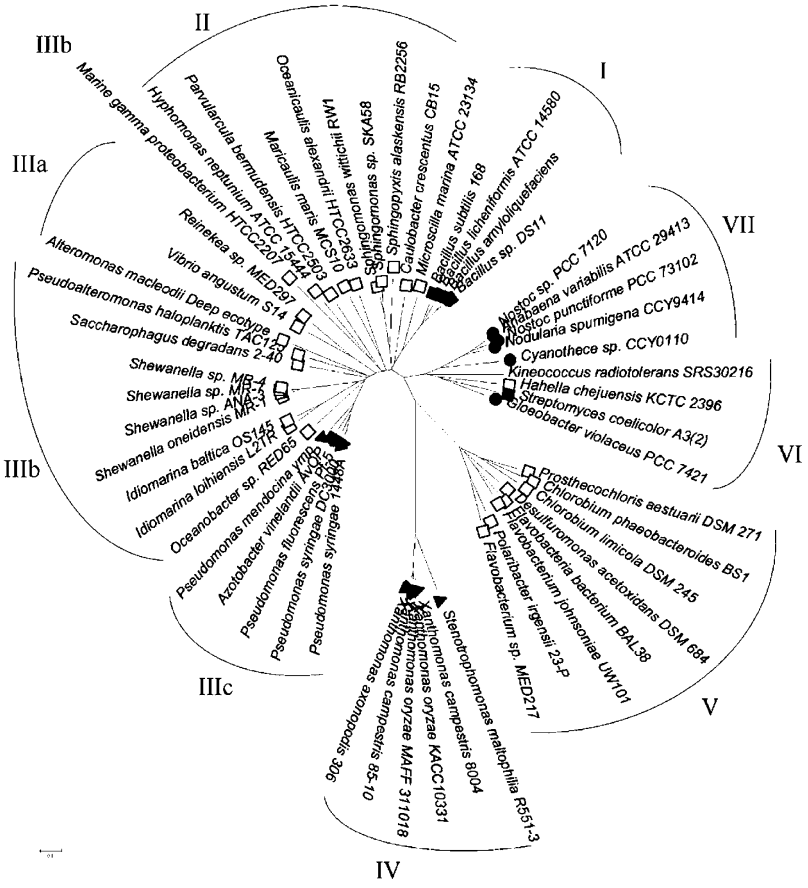


Figure 3 Phylogenetic analysis of prokaryotic BPP-like sequences. The symbols in front of the name of the bacterial strains represent the habitats: plant (▲), soil (■), aquatic systems (□) and cyanobacteria (●): *Alteromonas macleodii* Deep ecotype, EAR06739; *Anabaena variabilis* ATCC 29413, ZP_00158726; *Azotobacter vinelandii* AvOP, ZP_00418781; *Bacillus amyloliquefaciens*, AAW28542; *B. licheniformis* ATCC 14580, YP_090097; *Bacillus* sp DS11, AAC38573; *B. subtilis* 168PhyA, NP_389861; *Caulobacter crescentus* CB15, NP_420108; *Chlorobium limicola* DSM 245, ZP_00512507; *C. phaeobacteroides* BS1, ZP_00532419; *Cyanothece* sp CCY0110, EAZ90225; *Desulfuromonas acetoxidans* DSM 684, ZP_00549889; *Flavobacteria bacterium* BAL38, EAZ95593; *Flavobacterium johnsoniae* UW101, EAS61177; *Flavobacterium* sp MED217, ZP_01062174; *Gloeobacter violaceus* PCC 7421, NP_925045; *Hahella chejuensis* KCTC 2396, YP_434827; *H. neptunium* ATCC 15444, YP_758910; *Idiomarina baltica* OS145, ZP_01042413; *I. loihiensis* L2TR, YP_154494; *Kineococcus radiotolerans* SRS30216, EAM76705; *Maricaulis maris* MCS10, YP_757598; marine *gamma proteobacterium* HTCC2207, EAS47070; *Microscilla marina* ATCC 23134, EAY24393; *Nostoc punctiforme* PCC 73102, ZP_00108697; *Nostoc* sp PCC 7120, NP_488278; *Nodularia spumigena* CCY9414, EAW44492; *Oceanicaulis alexandrii* HTCC2633, ZP_00953252; *Oceanobacter* sp RED65, EAT12633; *Parvularcula bermudensis* HTCC2503, ZP_01017268; *Polaribacter irgensii* 23-P, EAR12508; *Prosthecochloris aestuarii* DSM 271, ZP_00591274; *Pseudoalteromonas haloplanktis* TAC125, CAI85536; *Pseudomonas fluorescens* Pf-5, YP_260816; *P. mendocina ymp*, ZP_01528237; *P. syringae* 1448, YP_275170; *P. syringae* str. DC3000, NP_793025; *Reinekea* sp MED297, EAR10111; *Saccharophagus degradans* 2–40, AABI03000001; *Shewanella oneidensis* MR-1, NP_718111; *Shewanella* sp ANA-3, ZP_00851245; *Shewanella* sp MR-4, ZP_00881318; *Shewanella* sp MR-7, ZP_00855136; *Sphingomonas* sp SKA58, EAT09404; *S. wittichii* RW1, EAW03600; *S. alaskensis* RB2256, EAN45496; *S. maltophilia* R551-3, EAX23120; *Streptomyces coelicolor* A3, NP_631736; *Vibrio angustum* S14, EAS63574; *Xanthomonas axonopodis* str. 306, NP_642834; *X. campestris* str. ATCC 33913, NP_637738; *X. campestris* str. 8004, YP_242815; *X. campestris* str. 85-10, YP_364432; *X. oryzae* KACC10331, YP_201138; *X. oryzae* MAFF 311018, YP_451391.

When using the *Burkholderia* PAP as a probe to blast the GOS database, 59 PAP-like sequences were identified in a single data set (GS000a, Sargasso station 13, 1106 Mbp) but not in the other 55 subsets (11 564 Mbp). Out of these 59 sequences, 33 have an *E* value $<E^{-100}$, which means that they are highly homologous to the *Burkholderia* PAP. The closest sequence has an *E* value $<E^{-176}$ and exhibits 99% sequence identity to the PAP sequence in *Burkholderia* sp 383 genome. To verify whether these bacterial PAP-like genes encode for a phytase, the PAP-like protein from *B. cenocepacia J2315* was overexpressed in *E. coli* as a GST fusion protein. After cleavage of the fusion partner, the purified PAP-like protein was found to exhibit phosphatase activity towards *p*-nitrophenol phosphate, but not phytase activity, when phytate was supplied as a substrate (data not shown). Since this protein shares high protein sequence identity (52–96%) with the PAP-like sequences from the GS000a data set, it is unlikely that these proteins exhibit phytase activity. Furthermore, the appearance of *Burkholderia* PAP sequences in only one (GS000a, Sargasso station 13) out of the 56 GOS data sets of the expedition

indicates that the *Burkholderia* PAP sequences are not widespread in aquatic samples. Shipboard contamination of the sample with *Burkholderia* was suggested for the unexpectedly high population of this strain in one out of the seven Sargasso sea samples, (DeLong, 2005) and therefore, the data retrieved from this data set (GS000a) should be examined with caution.

About 589 BPP-like sequences were found in the GOS database (Supplementary Table S1). BPP-like sequences were found in almost all large data sets, but not in those (GS038–46, GS048–50) with small genetic information (0.8–1.2 Mbp), indicating that BPP-like sequences are essentially ubiquitous in aquatic environments. The average number of hits per Gbp total read length was 46.5, which was higher than the average number for AP-like sequences (28.2, Supplementary Table S2). Five data sets with 89–234 Mbp read length exhibited exceptional low hits (0–6.4 hits/Gbp), including the two estuary data sets (GS011 and GS012), two coastal data sets (GS007 and GS021), and a sample from fringing reef (GS025).

The BPP-like sequences with the lowest *E* values in most oceanic data sets (GS000b–GS006, GS008–GS010, GS013–GS017, GS019, GS022–GS032, and the positive data sets beyond GS034) are usually closest to the BPP-like sequence from marine *gamma proteobacterium* HTCC 2207 (*E* values $<E^{-100}$). Besides, in GS000a and GS018, the dominant BPP-like sequences have the highest homology to that of *Shewanella* sp and *Idiomarina* sp, respectively. In addition, BPP-like sequences with higher homology to that of *Microscilla marina* were commonly seen as the second dominant hits in samples (GS000c/d, GS001a/c, GS017 and GS019) collected from open ocean (with ocean depths >3000 m). An X-ray structure of the BPP from *B. amyloliquefaciens* (Ha *et al*., 2000) identified the specific amino-acid residues responsible for two phosphate- and six calcium-binding sites. The BPP-like sequences from marine *gamma proteobacterium* HTCC 2207 and *M. marina* share $\sim$30 and $\sim$50% sequence identity to the *Bacillus* BPPs; however, most, if not all, of the residues that are responsible for the calcium- and the phosphate-binding sites are conserved. Hence, it is likely that these environmental BPP-like sequences encode for functional phytases.

The environments in the Lake Gatun (GS020) and the Punta Cormorant Lagoon (GS033) are very different from the ocean. All 19 hits retrieved from the Lake Gatun database were homologous to the BPP domains of the five cyanobacteria (Group VII) in Figure 3. No marine BPP-like sequences were seen in this freshwater database. Of the 33 BPP-like sequences retrieved from the lagoon (GS033), 13, 6, 6, 2 and 2 sequences have the highest homology to the BPP-like sequences from *Xanthomonas, Bacillus, Pseudomonas, Flavobacteria and Hyphomonas Neptunium*, respectively (Figure 3). However, BPP-like sequence from marine *gamma proteobacterium* HTCC 2207 was not found in this database.

The GOS database contains 7.7 million independent clones, covering 6325 Mbp (clear range vs 12 670 Mbp total read length) of microbial DNA sequences. Assuming that the average size of a bacterial genome is 4 Mbp, 6325 Mbp is equivalent to $\sim$1582 genomes. Each of the microbial genomes in Figure 3 carries only one copy of BPP-like gene. Hence, the prevalence of BPP-like sequences seems to be high (589/1582 or $\sim$37% of the microbial population). Thus, BPP-like genes are present in a substantial population of the bacteria in the aquatic environments, slightly higher than the incidence of *E. coli* alkaline phosphatase-like genes (357 hits; $E<10^{-10}$).

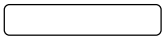
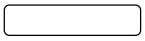
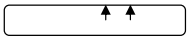
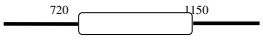
Distribution of BPP-like sequences in the other environmental databases

A total of 21 smaller environmental databases from NCBI (Supplementary Table S3) were probed for phytase-like genes. No hits were obtained for HAP, PAP and CPhy families. BPP search returned six hits from the soil database (AAFX*)* and one hit from a whale fall database (AAFZ). As the sizes of the soil database (100 Mbp) and the whale fall database (25 Mbp) were approximately 1/126 and 1/500 of that of the GOS database, respectively, the small number of hits found in these two databases seems to be in proportion (589/126 and 589/500), implying that BPP-like gene may be found in a population of the bacteria in these databases, as in the GOS database. Around 2000–18 000 distinct prokaryotic genomes were estimated to be present in each gram of soil and $\sim$100 Gbp sequences are required to represent all of the different genomes (Daniel, 2005). The current soil environmental database is clearly too small to provide adequate coverage.

Distribution of BPP-like proteins in bacterial genomes

The BPP-like sequences in bacterial genomes can be classified into several subgroups based on their domain structure and the position of their conserved cysteine residues (Table 1). Proteins in subgroups III contain two BPP domains, whereas proteins in the other subgroups have a single BPP domain. All five cyanobacterial proteins in subgroup VII carry one single BPP domain flanked by two different domains with unknown functions. The three-dimensional structures of the BPP domains in these subgroups were produced from the Swiss-Model comparative protein modeling server (Schwede *et al*., 2003). The models showed that the two conserved cysteine residues in the proteins of Gram-negative bacteria (subgroups II, III and IV) are in close proximity, and therefore, the formation of an intra-molecular disulfide bond is possible. Putative disulfide bonds may stabilize the protein (Cheng *et al*., 2007), and the conservation of these cysteine residues

Table 1 Putative proteins that carry BPP domains

Group	*Taxonomy*	*Domain structure*	*Amino-acid residues*	*No. of conserved cysteine residues*	*Sequence identity to BS168 BPP*[a] (%)
I	Gram+ve bacilli		381–383	None	70–100
II	α-Proteobacteria		342–350	2 (161, 212)[b]	36–40
III[c]	γ-Proteobacteria		597–738	4 (161, 212)[b]	33–46
IV	γ-Proteobacteria; Xanthomonadales		365–373	3 (12, 266, 366)[b]	23–25
V	Flavobacteria, Chlorobia, δ-proteobacteria		340–364	None	28–30
VI	Miscellaneous, including Actinobacteria		435–474	2 (206–207)[d] (251–252)[d]	33–37
VII	Cyanobacteria		1417–2698	None	40–43

Abbreviation: BPP, β-propeller phytase.
The arrows indicate the position of the cysteine residues.
[a]Sequence identity was calculated by the Smith–Waterman algorithm at the EMBOSS website.
[b]The corresponding positions of these cysteine residues on *B. Subtilis* 168 phytase.
[c]The marine γ-proteobacterium HTCC2207 has only a single BPP domain.
[d]The two cysteine residues are located on two additional stretches of amino-acid residues inserted in the corresponding positions on *B. Subtilis* 168 phytase.

indicated that the proteins in the same subgroups are evolutionarily related.

Among these proteins, only those from *Bacillus* spp (Tye *et al.*, 2002) and *S. oneidensis* MR-1 (Cheng and Lim, 2006) have demonstrated phytase activity. While the BPP-like proteins in the other subgroups share only low-protein sequence identity (20–40%), most of the essential amino acids that are responsible for calcium- and phosphate-binding (Ha *et al.*, 2000) are conserved in these sequences (data not shown). Hence, these proteins are likely to function as phytases, especially for subgroups I to VI, which are represented only by one or two BPP domains. In contrast, the role of proteins in subgroup VII is obscure. These large cyanobacterial proteins carry a single BPP domain (Table 1). The N and the C-terminal sequences show homology only to domains with unknown function. However, phytase activity in cyanobacteria has been demonstrated previously. Specifically, in a growth study of 50 cyanobacterial strains, 35 were able to grow on phytate, indicating that they must express a phytate-hydrolyzing enzyme (Whitton *et al.*, 1991). The possibility of subgroup VII proteins having phytase activity is, therefore, a subject for further investigation.

Subgroup IV is BPP-like proteins derived from *Xanthomonas* spp. A BPP protein from *X. oryzae* was reported to be a virulence factor for the pathogen and its deletion mutant strains exhibited lower growth rate when phytate was supplied as a sole P source (Chatterjee *et al.*, 2003). *Xanthomonas* spp is the only species that carries multiple phytase genes in its genome. In addition to BPP, HAP- and CPhy- like genes are also present in the genomes of *Xanthomonas* spp (Figures 1 and 2). Five *Xanthomonas* genomes are available in Genbank (Figure 3). All of these genomes carry one copy of a BPP-like gene and a copy of a HAP-like gene. For CPhy-like genes, there was one hit in each of three strains (8004, 85–10 and ATCC 33913).

Genomic organization of microbial phytase genes – phosphorus and iron uptake

The genetic context of the phytase genes was assessed by analysis of flanking genes using the NCBI and Kyoto encyclopedia of genes and genomes (KEGG) databases. Genes in the same orientation and in close proximity to BPP-like genes were analyzed. The resulting genome structures can be divided into several subgroups (Table 2). In general, phytase genes in subgroups I, IIIa, VI and VII have individual promoters and are independent of the flanking genes. The phytase genes of the other subgroups are all located downstream of a TonB-dependent receptor-like gene. TonB-dependent receptors are involved in the high-affinity binding and energy-dependent uptake of organometallic cofactors or carriers into the periplasm such as ferrisiderophores, porphyrins and cobalamins (Ferguson and Deisenhofer, 2004). TonB and two additional cytoplasmic membrane proteins, ExbB and ExbD, form an energy-transduction complex. ExbB and ExbD couple cytoplasmic membrane proton motive force to a conformation change of TonB, which in turn provides the driving force for cargo import (for example, siderophore) by the TonB-dependent receptor (Koebnik, 2005). The phytase gene proximity to this complex for subgroups IIIb, IIIc, IV, V and some of the members of subgroup II is significant in that they are likely to be transcribed as an operon with the common promoter being upstream of the TonB-dependent receptor gene. The gaps between phytase and TonB-dependent receptor-like genes are

Table 2 Genomic organizations around the microbial BPP phytase genes

Genetic organization	*Strains*
	Groups I, IIIa, VI and VII *C. crescentus*, *H. neptunium* and *M. maris* in group II
	Other bacteria in group II[a] Group V except *D. acetoxidans* *X. oryzae KACC10331* in group IV *M. marina ATCC 23134*
	Group IIIb
	Group IIIc
	Group IV except *X. oryzae KACC10331*
	*D. acetoxidans*

The gray, stripped and black boxes represent a single or double *β*-propeller phytase (BPP) domain, a TonB-dependent receptor and a putative phosphatase, respectively. X and Y are two unknown genes and AP represents an alkaline phosphatase gene.
[a]The genes for BPP and the TonB-dependent receptor are in back-to-back direction in *Oceanicaulis alexandrii* and *Sphingopyxis alaskensis*.

small. For example, the distances are only 0, 2, 27 and 40 bp in *Xanthomonas campestris 8004*, *Flavobacterium* sp *MED217, S. oneidensis MR-1, Pseudomonas fluorescens Pf-5*, respectively. Expectedly, these phytases are predicted to be exported to the membrane or to the periplasmic space by the PSORT program (Nakai and Horton, 1999). Hence, the Gram-negative bacteria in subgroups II, IIIb, IIIc, IV and V are probably taking up phytate through a TonB system and depositing it in the cell's periplasm in an energy-dependent process. The ubiquity of putative TonB receptor-phytase links suggests that this is a fertile area for future research.

Reports of phytate siderophore activity in a *Pseudomonas aeruginosa* strain indicated that iron uptake is substantially enhanced when the form is Fe-phytate (Smith *et al.*, 1994) as opposed to the free, ferric form. Iron, an essential nutrient for microorganisms, is normally insoluble in terrestrial and aquatic systems and is thus not bioavailable. In another study, myo-inositol tri- and tetra-phosphate were also able to facilitate Fe transport into *Pseudomonas aeruginosa* (Hirst *et al.*, 1999). This implies that phytate may be useful to microorganisms as a source of P and carbon, and also, iron.

Importance of phytases in nutrient cycle

In terrestrial systems, phytate incidence correlates with organic C composition (Turner *et al.*, 2002). In what are likely overestimates, 30–48% of soil and rhizosphere microbial populations may have a phytate-degrading ability (Hill and Richardson, 2006). Our work suggests that BPPs play a major role in phytate hydrolysis in both soil and aquatic microbial communities. While a vast amount of phytate is transferred to the aquatic ecosystem from the terrestrial (Suzumura and Kamatani, 1995a), it is rapidly broken down into inorganic phosphate in the aquatic system (Suzumura and Kamatani, 1995b). Our work indicates that BPP class phytases play important roles in this transformation. The recycling of inorganic P from organic phytate-P is important for the aquatic ecosystem. For example, the fixation of CO_2 and N_2 by the dominant N_2-fixing marine cyanobacteria in oceans, *Trichodesmium*, is thought to be constrained by P concentration (Sanudo-Wilhelmy *et al.*, 2001; Dyhrman *et al.*, 2006). In addition, in the Northern Atlantic, Fe and P were found to co-limit N_2 fixation (Mills *et al.*, 2004). Our work shows that BPP-like genes in the genome of Gram-negative marine bacteria are commonly linked to TonB-dependent receptor-like genes. The operon context implies that Fe-complexed phytate is imported through a TonB receptor; whereupon P, C and Fe are then available to the organism.

Acknowledgements

This work was supported by the University Research Committee of the University of Hong Kong, China (No. 10205120). We thank Rekha Seshadri and Aaron Halpern of CAMERA for answering our enquiry about the GOS database.

References

Barnett GM. (1994). Phosphorus forms in animal manure. *Bioresour Technol* **49**: 139–147.

Chatterjee S, Sankaranarayanan R, Sonti RV. (2003). PhyA, a secreted protein of *Xanthomonas oryzae pv. oryzae*, is required for optimum virulence and growth on

phytic acid as a sole phosphate source. *Mol Plant Microbe Interact* **16**: 973–982.
Cheng C, Lim BL. (2006). Beta-propeller phytases in the aquatic environment. *Arch Microbiol* **185**: 1–13.
Cheng C, Wong KB, Lim BL. (2007). The effect of disulfide bond on the conformational stability and catalytic activity of beta-propeller phytase. *Protein Pept Lett* **14**: 175–183.
Chu HM, Guo RT, Lin TW, Chou CC, Shr HL, Lai HL *et al.* (2004). Structures of *Selenomonas ruminantium* phytase in complex with persulfated phytate: DSP phytase fold and mechanism for sequential substrate hydrolysis. *Structure (Camb)* **12**: 2015–2024.
Daniel R. (2005). The metagenomics of soil. *Nat Rev Microbiol* **3**: 470–478.
deGroot CJ, Golterman HL. (1993). On the presence of organic phosphate in some Camargue sediments – evidence for the importance of phytate. *Hydrobiologia* **252**: 117–126.
DeLong EF. (2005). Microbial community genomics in the ocean. *Nat Rev Microbiol* **3**: 459–469.
DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N *et al.* (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 *et al.* (2006). Phosphonate utilization by the globally important marine diazotroph *Trichodesmium*. *Nature* **439**: 68–71.
Eisenreich SJ, Armstrong DE. (1977). Chromatographic investigation of inositol phosphate esters in lake waters. *Environ Sci Technol* **11**: 497–501.
Ferguson AD, Deisenhofer J. (2004). Metal import through microbial membranes. *Cell* **116**: 15–24.
Golterman H, Paing J, Serrano L, Gomez E. (1998). Presence of and phosphate release from polyphosphates or phytate phosphate in lake sediments. *Hydrobiologia* **364**: 99–104.
Ha NC, Oh BC, Shin S, Kim HJ, Oh TK, Kim YO *et al.* (2000). Crystal structures of a novel, thermostable phytase in partially and fully calcium-loaded states. *Nat Struct Biol* **7**: 147–153.
Herbes SE, Allen HE, Mancy KH. (1975). Enzymatic characterization of soluble organic phosphorus in lake water. *Science* **187**: 432–434.
Herter T, Berezina OV, Zinin NV, Velikodvorskaya GA, Greiner R, Borriss R. (2006). Glucose-1-phosphatase (*AgpE*) from *Enterobacter cloacae* displays enhanced phytase activity. *Appl Microbiol Biotechnol* **70**: 60–64.
Hill JE, Richardson A. (2006). Isolation and assessment of microorganisms that utilize phytate. In: Turner B, Richardson A, Mullaney E. (eds). *Inositol Phosphates-Linking Agriculture and the Environment*. CABI: Oxford, UK, ISBN: 1845931521.
Hirst PH, Riley AM, Mills SJ, Spiers ID, Poyner DR, Freeman S *et al.* (1999). Inositol polyphosphate-mediated iron transport in *Pseudomonas aeruginosa*. *J Appl Microbiol* **86**: 537–543.
Kelley LA, MacCallum RM, Sternberg MJ. (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J Mol Biol* **299**: 499–520.
Koebnik R. (2005). TonB-dependent trans-envelope signalling: the exception or the rule? *Trends Microbiol* **13**: 343–347.
Kumar S, Tamura K, Nei M. (2004). MEGA3: integrated software for Molecular Evolutionary Genetics analysis and sequence alignment. *Brief Bioinform* **5**: 150–163.
Lott JNA, Ockenden I, Raboy V, Batten GD. (2000). Phytic acid and phosphorus in crop seeds and fruits: a global estimate. *Seed Sci Res* **10**: 11–33.
Lung SC, Lim BL. (2006). Assimilation of phytate-phosphorus by the extracellular phytase activity of tobacco (*Nicotiana tabacum*) is affected by the availability of soluble phytate. *Plant Soil* **279**: 187–199.
Macklon AE, Grayston SJ, Shand CA, Sim A, Sellars S, Ord BG. (1997). Uptake and transport of phosphorus by *Agrostis capillaris* seedlings from rapidly hydrolysed organic sources extracted from 32P-labelled bacterial cultures. *Plant Soil* **190**: 163–167.
Matsuda O, Ohmi K, Sasada K. (1985). Dissolved organic phosphorus in sea water, its molecular weight fractionation and availability to phytoplankton. *J Fac Appl Biol Sci Hiroshima Univ* **24**: 33–42.
Mills MM, Ridame C, Davey M, La RJ, Geider RJ. (2004). Iron and phosphorus co-limit nitrogen fixation in the eastern tropical North Atlantic. *Nature* **429**: 292–294.
Mullaney EJ, Ullah AH. (2003). The term phytase comprises several different classes of enzymes. *Biochem Biophys Res Commun* **312**: 179–184.
Nakai K, Horton P. (1999). PSORT: a program for detecting the sorting signals of proteins and predicting their subcellular localization. *Trends Biochem Sci* **24**: 34–35.
Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Willamson S, Yooseph S *et al.* (2007). The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* **5**: 398–431.
Sanudo-Wilhelmy SA, Kustka AB, Gobler CJ, Hutchins DA, Yang M, Lwiza K *et al.* (2001). Phosphorus limitation of nitrogen fixation by *Trichodesmium* in the central Atlantic Ocean. *Nature* **411**: 66–69.
Schwede T, Kopp J, Guex N, Peitsch MC. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* **31**: 3381–3385.
Seshadri R, Kravitz SA, Smarr L, Gilna P, Frazier M. (2007). CAMERA: a community resource for Metagenomics. *PLoS Biol* **5**: 394–397.
Smith AW, Poyner DR, Hughes HK, Lambert PA. (1994). Siderophore activity of *myo*-inositol hexakisphosphate in *Pseudomonas aeruginosa*. *J Bacteriol* **176**: 3455–3459.
Smith TF, Waterman MS. (1981). Identification of common molecular subsequences. *J Mol Biol* **147**: 195–197.
Stevens RJ, Stewart BM. (1982). Concentration, fractionation and characterization of soluble organic phosphorus in river water entering Lough Neagh. *Water Res* **16**: 1507–1519.
Stevenson FJ, Cole MA. (1999). Phosphorus. In: Stevenson FJ, Cole MA. (eds). *Cycles of Soil: Carbon, Nitrogen, Phosphorus, Sulfur, Micronutrients*. John Wiley & Sons: New York, pp 279–329.
Suzumura M, Kamatani A. (1995a). Origin and distribution of inositol hexaphosphate in estuarine and coastal sediments. *Limnol Oceanogr* **40**: 1254–1261.
Suzumura M, Kamatani A. (1995b). Mineralization of inositol hexaphosphate in aerobic and anaerobic marine sediments – implications for the phosphorus cycle. *Geochim Cosmochim Acta* **59**: 1021–1026.
Tang J, Leung A, Leung C, Lim BL. (2006). Hydrolysis of precipitated phytate by three distinct families of phytases. *Soil Biol Biochem* **38**: 1316–1324.
Thompson JD, Higgins DG, Gibson TJ. (1994). CLUSTAL W: improving the sensitivity of progressive multiple

sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.

Toor GS, Condron LM, Di HJ, Cameron KC, Cade-Menun BJ. (2003). Characterization of organic phosphorus in leachate from a grassland soil. *Soil Biol Biochem* **35**: 1317–1323.

Turner BL, Paphazy MJ, Haygarth PM, McKelvie ID. (2002). Inositol phosphates in the environment. *Philos Trans R Soc Lond B Biol Sci* **357**: 449–469.

Tye AJ, Siu F, Leung TYC, Lim BL. (2002). Molecular cloning and the biochemical characterization of two novel phytases from *B. subtilis* 168 and *B. licheniformis*. *Appl Microbiol Biotechnol* **59**: 190–197.

Vincent JB, Crowder MW. (1995). Metal-containing phosphatases part A: alkaline phosphatases. In: Vincent JB, Crowder MW. (eds). *Phosphatases in Cell Metabolism and Signal Transduction: Structure, Function and Mechanism of Action*. RG Landes: Austin, pp 99–110.

Whitton BA, Grainger SLJ, Hawley GRW, Simon JW. (1991). Cell-bound and extracellular phosphatase activities of cyanobacterial isolates. *Microbial Ecol* **21**: 85–98.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)