Phosphorus (P) is an essential macronutrient for all living cells (1). Despite its relative abundance in soils, P is one of the main limiting nutrients for terrestrial organisms (2). P is present in organic and inorganic forms in soil, but only the inorganic orthophosphate ions in soil solutions are readily available for plants (3). To sustain crop productivity, large amounts of P fertilizers are therefore used in agriculture, both as inorganic fertilizers (e.g., triple super phosphate) and organic fertilizers (e.g., manure). After application, some of the inorganic P is rapidly taken up by plants and microorganisms, while the remaining P is immobilized as insoluble and bound P forms in the soil. Microorganisms can access and recycle P from these recalcitrant P forms by solubilization of inorganic P and by mineralization of organic P via enzymatic processes mediated primarily by phosphatases, which hydrolyze the orthophosphate group from organic compounds (3). When facing P scarcity, microorganisms upregulate expression of functional genes coding for phosphatases (phosphomonoesterases, phosphodiesterases, phytases), high-affinity phosphate transporters, and enzymes for phosphonate utilization, which together constitute the Pho regulon (4). The phosphomonoesterases which are hydrolyzed by phosphatases are generally the dominant fraction of organic P and can represent up to 90% of the organic P in soil (3).

Prokaryotic alkaline phosphatases have been grouped into three distinct families, PhoA, PhoD, and PhoX (5–7), which are classified in COG1785, COG3540, and COG3211, respectively, of the Cluster of Orthologous Groups (COG) categorization. PhoA was the first alkaline phosphatase to be characterized. It is a homodimeric enzyme that hydrolyzes phosphomonoesters and is activated by Mg$^{2+}$ and Zn$^{2+}$ (7). PhoD and PhoX are monomeric enzymes that hydrolyze both phosphomonoesters and phosphodiesters and are activated by Ca$^{2+}$ (5, 6). Enzymes of all three families are predominantly periplasmic, membrane bound, or extracellular (8). PhoD and PhoX are exported by the twin-arginine translocation pathway (5, 6), while PhoA is secreted via the Sec protein translocation pathway (9). There is high sequence variability in the PhoA, PhoD, and PhoX proteins, not only between the families but also within each family (5, 9). PhoD is widespread in both terrestrial and aquatic ecosystems (8, 10).

Until recently, our knowledge of the phosphatase-encoding genes in prokaryotes was based on traditional culture-dependent methods. Advances in culture-independent techniques have provided new tools for the study of microbial communities in the environment. The first functional gene probes to target alkaline phosphatase genes were the primers developed by Sakurai et al. named ALPS primers (11). They were based on phosphatase gene sequences from seven isolates and first used to examine the differences in soil alkaline phosphatase community structures resulting from mineral and organic fertilization. Alkaline phosphatase genes belonging to the Actinobacteria, Alpha-, Beta-, and Gamma-proteobacteria, and Cyanobacteria classes were identified by clon-
ing, giving the first insight into alkaline phosphatase diversity in soil (11).

Subsequently, the ALPS primers were demonstrated to be specific to the phoD alkaline phosphatase gene (10). They were used to assess alkaline phosphatase gene diversity and structure in several soils by PCR-denaturing gradient gel electrophoresis (DGGE) (12–15) and by 454 sequencing (10, 16). These studies showed that crop management, application of organic and conventional fertilizers, and vegetation all affect the phoD alkaline phosphatase gene diversity. Tan et al. (10) examined the effects of three mineral P fertilization intensities (zero, medium, and high input) in grassland soil on the composition and diversity of alkaline phosphatase and found a change in the phoD bacterial community compositions between unfertilized and fertilized treatments, with the dominant phoD alkaline phosphatase genes affiliated with Alpha- and Gammaproteobacteria, Actinobacteria, and Cyanobacteria. However, they pointed out that the ALPS primers are likely to have an amplification bias, resulting in an overrepresentation of Alphaproteobacteria, and that new primers are therefore required to provide better coverage of the phoD diversity.

In this study, we assessed the diversity and environmental distribution of the phoD gene based on current genome and metagenome databases, and we present a new set of improved primers which targets the large diversity of phoD genes in soil microorganisms. These primers can be used as a tool both to identify PhoD-producing bacteria and to study phoD bacterial community diversity and composition in the environment. The newly designed primers were tested in a gene-targeted metagenomic approach using 454 sequencing in a range of soils collected from two continents with different climates and soil properties. Finally, we compared them to the previously published ALPS primers (11), using the same samples and methodology.

MATERIALS AND METHODS

Taxonomic and environmental distribution of phoD alkaline phosphatase genes across microbial genomes and metagenomes. The distribution of phoD genes was assessed using the Integrated Microbial Genomes and Metagenomes (IMG/M) database, a dedicated system for annotation of whole genomes and metagenomes (17). Draft and complete genome data sets were used to evaluate the distribution of phoD across kingdoms and phyla, and metagenome data sets were used to evaluate the prevalence of phoD in the environment (data accessed on 13 July 2015). Metagenome data sets were categorized as “air,” “engineered and waste” (bioreactor and waste treatment), “extreme environments” (saline, alkaline, hot spring, brine, and black smokers), “fresh water,” “marine environment,” “plant-associated” (leaves and wood), “animal-associated” (associated with humans, arthropods, molluscs, and sponges), and “soil” (rhizosphere and bulk soil). These categories were chosen based on the environment-type classification of the IMG/M database. The relative abundance of phoD gene counts per environment type was calculated as the gene count number normalized by the total number of bases sequenced per metagenome data set.

Soil sampling and general soil characteristics. Four grassland soils were collected in Australia in July 2013 (samples 1 to 4 [S1 to S4]), and two grassland soils were sampled in Switzerland in September 2012 (S5 and S6) (Table 1). These represent a broad range of soil types, vegetation, and climatic conditions, varying from hot semi-arid to continental temperate climates. At each site, five soil cores from the top 5 cm were randomly collected and homogenized by sieving (4 mm). A subsample was stored at −80°C for molecular analysis. The remaining composite soil was air dried and used to determine the basic soil properties, including pH, texture, and total carbon (C) and P. Methods used to determine the soil properties are

<table>
<thead>
<tr>
<th>Sample Site</th>
<th>Soil typea</th>
<th>Geographical coordinates</th>
<th>Climate (Köppen-Geiger)</th>
<th>Soil typeb</th>
<th>Climate (climatic category)c</th>
<th>Soil typec</th>
<th>Climate (climatic category)c</th>
<th>Soil typec</th>
<th>Climate (climatic category)c</th>
<th>Soil typec</th>
<th>Climate (climatic category)c</th>
<th>Soil typec</th>
<th>Climate (climatic category)c</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1 Kia-Ora (Australia)</td>
<td>Cambisol</td>
<td>34°49'18&quot;S, 148°33'00&quot;E</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
</tr>
<tr>
<td>S2 Narrabri (Australia)</td>
<td>Cambisol</td>
<td>30°15'00&quot;S, 147°04'14&quot;E</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
</tr>
<tr>
<td>S3 Nyngan (Australia)</td>
<td>Cambisol</td>
<td>31°25'00&quot;S, 147°04'14&quot;E</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
</tr>
<tr>
<td>S4 Mutawintji (Australia)</td>
<td>Cambisol</td>
<td>31°16'00&quot;S, 147°14'44&quot;E</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
</tr>
<tr>
<td>S5 Watt (Switzerland)</td>
<td>Arrhenaterion elatioris</td>
<td>47°25'45&quot;N, 008°29'31&quot;E</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
</tr>
<tr>
<td>S6 Watt (Switzerland)</td>
<td>Arrhenaterion elatioris</td>
<td>47°25'45&quot;N, 008°29'31&quot;E</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
<td>Planosol</td>
<td>Warm temperate climate, fully humid (Cfb)</td>
</tr>
</tbody>
</table>
described in Table 1. The sampled soils covered a range of textures, with clay contents varying between 12 and 38%. Soil pH ranged between 4.2 and 6.8. Total C varied between 5 and 34 g kg⁻¹ soil, and total P varied between 193 and 705 mg kg⁻¹ soil. The vegetation densities were similar at sampling sites S5 and S6 but very different at the other sites, ranging from dense to scarce, depending on the location.

DNA extraction from soil. All DNA samples were extracted in duplicate. Nucleic acids were extracted from the Australian samples using a DNA PowerSoil isolation kit (Mo Bio, Carlsbad, CA, USA), according to the manufacturer’s instructions, with an initial bead-beating step of 2 cycles of 3 min at 30 Hz using a TissueLyser II (Qiagen, CA). Nucleic acids were extracted from the Swiss samples (2 g of frozen soil) using an RNA PowerSoil isolation kit (Mo Bio) according to the manufacturer’s instructions, with an additional homogenizing step using an Omni Bead Ruptor homogenizer (Omni International, Kennesaw, GA) (2.8-mm zirconium beads for 1 min at 5 m s⁻¹) prior to isolation. DNA was eluted from the RNA/DNA capture column using 4 mL of DNA elution solution (1 M NaCl, 50 mM morpholinepropanesulfonic acid [MOPS], 15% isopropanol [pH7.0]). DNA was precipitated using isopropanol and resuspended in diethyl pyrocarbonate (DEPC)-treated H₂O. Only the DNA extracts were used in this study.

Primer design and in silico testing. Gene sequences annotated as phoD and/or associated with COG3540 (Clusters of Orthologous Groups; http://www.ncbi.nlm.nih.gov/COG/), which corresponds to phoD alkaline phosphatase, were retrieved from the European Nucleotide Archive (ENA) and UniProt Knowledgebase (UniProtKB) databases. They were then clustered at 97% similarity using CD-HIT (18), resulting in a total of 315 sequences used as references for the primer design (see the list in Table S1 and the taxonomic tree in Fig. S1 in the supplemental material). The reference sequences were affiliated with 11 phyla, including Actinobacteria (59 sequences), Bacteroidetes (22 sequences), Cyanobacteria (22 sequences), Deinococcus-Thermus (2 sequences), Ignavibacteriae (1 sequence), Firmicutes (13 sequences), Gemmatimonadetes (1 sequence), Spirochaetes (16 sequences), Planctomycetes (4 sequences), Proteobacteria (173 sequences), and Verrucomicrobia (2 sequences).

The gene sequences were aligned using MUSCLE (19), and the alignment was manually reviewed by comparison with the aligned translated sequences, using Geneious 6.1.2 (Biomatters, Australia) and the alignment of the COG3540 group available on the NCBI website (Conserved Domain Protein Family, http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) as the amino acid reference alignment. The most suitable conserved regions for primer design were identified using PrimerProspector (20). Forward and reverse candidate primers were then manually designed to reach the maximum coverage of the reference sequences. Candidate primers were paired to target an amplicon length of 250 to 500 bp, which represents the best compromise length for next-generation sequencing and quantitative PCR studies. They were then tested in silico using De-MetaST-BLAST (21) to identify potential primer pairs with an appropriate product size and coverage of the reference sequences.

Optimization and validation of phoD-targeting primers. Candidate primers (21 forward primers and 23 reverse primers) were tested in a gradient PCR using a mixture of soil genomic DNA (S5 and S6) (Table 1) as the template. PCRs were performed in a 25-µL volume containing 1× MyTaq reaction buffer (including MgCl₂ and deoxyadenosine triphosphates [dNTPs]), 0.5 µM each primer, and 0.6 U of MyTaq polymerase (Bioline, NSW, Australia) with 1 to 2 ng DNA as the template in a 1X S1000 thermocycler (Bio-Rad Laboratories, CA). The amplification reaction included an initial denaturation step of 5 min at 95°C, followed by 35 cycles of a denaturation step of 30 s at 95°C and an annealing step of 30 s at the calculated annealing temperature of each candidate primer pair (gradient of ±3°C), and an extension step of 30 s at 72°C. A final extension step was performed for 5 min at 72°C. Amplicon size and intensity and the presence of primer dimers were assessed visually after electrophoresis on a 1.5% (wt/vol) agarose gel and staining with ethidium bromide.

The amplicon specificity was evaluated for selected primer pairs by cloning and sequencing. The PCR products were ligated at 4°C overnight using pGEM-T vector systems (Promega, Madison, WI) and transformed into chemically competent E. coli cells [α-select; F⁻ deor endA1 recA1 relA1 gyrA96 hisD17 rpsL22 supE44 thi-1 phoA Δ(lacZYA-argF)U169 φ80lacZΔM15 λ−] according to the manufacturer’s instructions (Bio- line). Restriction fragment length polymorphism (RFLP) profiling of clones with the expected insert size was done using Hhal (0.2 U/µl for 3 h at 37°C; Promega), and profiles were visualized by electrophoresis on a 2% (wt/vol) agarose gel. Representative inserts of unique RFLP profiles were then sequenced (Macrogen Inc., Seoul, South Korea). The resulting sequences were used to evaluate the coverage and specificity of the candidate primer pairs using BLAST (22).

Amplification diversity was examined for three candidate primer pairs by 454 GS-FLX + sequencing (Roche 454 Life Sciences, Branford, CT) using barcoded primers. The barcoded primer design, sequencing, and initial quality filtering were performed by Research and Testing Laboratory (Lubbock, TX) using standard protocols. Briefly, sequences with a quality score of <25 were trimmed, and chimeras were removed using USEARCH, with clustering at a 4% divergence (23). Denoising was performed with the Research and Testing Denoiser algorithm, which uses the nonchimeric sequences and the quality scores to create consensus clusters from aligned sequences. Within each cluster, the probability of prevalence of each nucleotide was calculated, and a quality score was generated, which was then used to remove noise from the data set.

The primer pair phoD-F373 (5′-TGGGAYGATCAYGARGT-3′)/phoD-R1083 (5′-CTGSGCSAKSACRTTCCA-3′) provided the highest phoD diversity and coverage (numbers indicate the respective positions in the reference phoD gene of Mesorhizobium loti MAFF303099). phoD-F373 anneals to the conserved region that consists of the amino acid residues WDHDE, which contribute to the coordination of two Ca²⁺ cations (24). In addition, the fragment targeted by phoD-F373/phoD-R1083 includes two conserved arginine residues. Nevertheless, the variable part of the amplified region also allows a good identification of taxonomy. This primer pair was named PHOD and used further in this study.

454 sequencing by PHOD and ALPS primers. For comparative analysis of PHOD and ALPS primers ALPS-F730/ALPS-R110 (5′-CA GTGGGACGACCACGAGGT-3′)/ALPS-MAFF303099 (25) were trimmed, and chimeras were removed using USEARCH, with clustering at a 4% divergence (23). Denoising was performed with the Research and Testing Denoiser algorithm, which uses the nonchimeric sequences and the quality scores to create consensus clusters from aligned sequences. Within each cluster, the probability of prevalence of each nucleotide was calculated, and a quality score was generated, which was then used to remove noise from the data set.

Sequence analysis. Sequencing data sets amplified by PHOD and ALPS primers were analyzed separately using mothur (25). Sequences were analyzed as nucleic acid sequences to keep the maximum information, allow accurate identification, and avoid artifacts due to framershifts and errors during back-translation (26). After demultiplexing, reads containing ambiguities and mismatches with either the specific primers or the barcode were removed. Reads with an average quality score of <20 were then filtered out. The remaining reads were trimmed at 150 bp and 450 bp as the minimum and maximum lengths, respectively. Across all samples, 92% of the sequences had a length between 320 and 380 bp.

The resulting PHOD- and ALPS-amplified data sets were merged and aligned using the Needleman-Wunsch global alignment algorithm as implemented in mothur, using 6-mer searching and the aligned reference sequences as the template. The pairwise distance matrix was calculated from the alignment, and sequences were clustered using the k-furthest method as implemented in mothur, with a similarity cutoff at 75% to define the operational taxonomic units (OTUs), as calculated by Tan et al. (10). OTU matrices were normalized to the smallest library size using the normalized.shared command in mothur to allow comparison between samples. The relative abundance of each OTU was normalized by the total number of reads per sample. The normalized values were then rounded to the nearest integer. The taxonomy assignment was performed using
Alkaline phosphatase activity has not previously been described based on the IMG/M database. A total of 63 archaeal, 6,469 bacterial, and 73 eukaryotic draft or complete genomes containing at least one copy of the phoD gene were found. In bacteria, the phoD gene was spread across 20 phyla (Fig. 1a). More than half of the genomes of Actinobacteria, Gemmatimonadetes, Spirochaetes, and Verrucomicrobia contained at least one copy of the phoD gene. Among the Proteobacteria, the phoD gene occurred in 52, 30, and 34% of the Alpha-, Beta-, and Gammaproteobacteria, respectively. The number of phoD copies per genome varied between 1 and 9, but the majority of sequenced genomes (71%) carried only a single copy. Although phoD is widespread across the bacterial phyla, it is important to note that the microbial genome sequence database contains the genomes of cultured strains almost exclusively, which creates a general bias in databases (32). Proteobacteria was the most recurrent phylum in the database, as the Gammaproteobacteria and more particularly the Pseudomonas genus are among the most intensively studied taxa (32) and thus are the genomes found most frequently in databases. Given the presence of the phoD gene in the less represented phyla, such as Chloroflexi, Deinococcaceae, and Planctomycetes, phoD-targeting primers represent an important tool to study these less easily culturable phyla.

Additionally, phoD genes were found in archaea, affiliated almost entirely with Euryarchaeota (Halobacteriaceae), and in eukaryotes, mainly in Ascomycetes. Alkaline phosphatase activity in archaea has only rarely been reported, e.g., from extreme environments (33, 34), while in eukaryotes it has been reported in Basidiomycetes (35) and in eukaryotic phytoplanktonic cells (36); in mammals, it is widely used as an indicator for liver disease (37). However, alkaline phosphatase activity has not previously been associated with the phoD gene in these taxa.

Environmental distribution of phoD alkaline phosphatase—a meta-analysis. The prevalence of phoD in the environ-
ment was investigated by analysis of 3,011 available metagenome data sets in the IMG/M database. The *phoD* gene was found in a range of environments (Fig. 1b), with the greatest abundance in soil, followed by marine and air environments.

Metagenomic studies focusing on phosphatases in marine environments have shown that *phoD* and *phoX* are more common than *phoA* in these samples (8,38). The high diversity and relative abundance of the *phoD* gene found in soil metagenomes (Fig. 1b) suggest that *phoD* may also be particularly relevant in terrestrial ecosystems, although the relative abundances of the three alkaline phosphatase families in soil have not yet been studied on the meta-
genome level. The fact that organic P represents between 30% and 80% of the total P in grassland and agricultural soils, mainly in the form of diverse phosphonomonooesters and diesters (3), may promote the diversity of *phoD* in terrestrial ecosystems.

**Performance of PHOD and ALPS primers.** A key aim of this work was to design a new set of PHOD primers targeting the bacterial *phoD* alkaline phosphatase for studying the *phoD* bact-
erial community diversity and composition in soil. The PHOD primers were tested on six soils that represent a range of contrast-
ing soil properties, collected in Australia and Switzerland, and the results were compared with those obtained with the same samples using the ALPS primers.

Generally, amplification using PHOD primers resulted in fewer filtered reads than that with ALPS primers, with 2,309 ± 1,148 (mean ± standard deviation) and 7,778 ± 3,107 reads and average read lengths of 380 ± 33 bp and 364 ± 35 bp for PHOD- and ALPS-amplified samples, respectively (Table 2). The difference in the number of filtered reads per library was directly linked to primer design, more particularly to the degree of degeneracy of the PHOD primers. Increasing degeneracy in primers generally increases the risk of unspecific annealing during the PCR but increases the probability of ampli-
fying yet-unknown *phoD* gene sequences by allowing all coding possibilities for an amino acid residue in the nucleic acid se-
quencies (40). When used appropriately, degenerate primers, such as the PHOD primers, represent a great advantage in studies on genetic diversity by targeting known and unknown sequences in environmental samples (41).

By filtering out redundant sequences, the number of reads decreased remarkably in the ALPS-amplified samples, leading to more-similar numbers of unique reads for the two sets of primers, which averaged 1,893 ± 885 bp (mean ± standard deviation) and 2,297 ± 659 bp for PHOD- and ALPS-amplified samples, respectively. This showed that ALPS-amplified samples consisted of a greater number of redundant reads than did PHOD-amplified samples. Finally, normalization of the library size in order to compare the two primer sets resulted in an average library size of 1,013 ± 31 bp. Our results suggest that the ALPS primers target a narrow spectrum of sequences which represent a large fraction of the reads after amplification.

**Species richness and alpha diversity of the *phoD* gene in six soils.** Amplification with PHOD primers revealed a 2-fold varia-
tion in species richness among the six samples (Table 2). *S*<sub>obs</sub> was lowest in S4 and highest in S1, with 148 and 290 OTUs, respec-
tively. Chao1 and *S*<sub>est</sub> indices, derived from the rarefaction curves, showed a similar trend. The difference in species richness between samples is well illustrated by the rarefaction curves (Fig. 2a). The rarefaction curve of S1 had the steepest slope, showing the greatest increase of species with the number of reads, while that of S4 reached the asymptote with the fewest reads (ca. 3,000).

Compared to amplification with PHOD primers, amplification with ALPS primers resulted in significantly lower species richness and alpha diversity (Table 2). In ALPS-amplified samples, the rarefaction curves always reached the asymptote with fewer reads than in the corresponding PHOD-amplified samples (Fig. 2a and b). The rarefaction curves of S1 when amplified using PHOD and ALPS primers contrasted particularly strongly, leading to a 7-fold difference in Chao1 and *S*<sub>est</sub>. Likewise, H' was always greater in PHOD- than in ALPS-amplified samples. This

---

**TABLE 2 Data obtained with PHOD and ALPS primers based on normalized data**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sample</th>
<th>No. of filtered reads</th>
<th>No. of unique reads</th>
<th>No. of reads after normalization</th>
<th>Species richness index</th>
<th>Good's coverage</th>
<th>No. of:</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHOD</td>
<td>S1</td>
<td>1,915</td>
<td>1,763</td>
<td>1,088</td>
<td>290</td>
<td>685</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>2,170</td>
<td>1,820</td>
<td>963</td>
<td>201</td>
<td>293</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>S3</td>
<td>3,090</td>
<td>2,709</td>
<td>1,001</td>
<td>227</td>
<td>458</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>S4</td>
<td>1,042</td>
<td>829</td>
<td>1,037</td>
<td>148</td>
<td>214</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>S5</td>
<td>4,399</td>
<td>3,296</td>
<td>977</td>
<td>191</td>
<td>359</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>S6</td>
<td>1,240</td>
<td>937</td>
<td>1,039</td>
<td>199</td>
<td>318</td>
<td>0.89</td>
</tr>
<tr>
<td>ALPS</td>
<td>S1</td>
<td>5,958</td>
<td>2,097</td>
<td>1,017</td>
<td>78</td>
<td>100</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>12,619</td>
<td>3,168</td>
<td>998</td>
<td>168</td>
<td>209</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>S3</td>
<td>3,730</td>
<td>1,276</td>
<td>1,027</td>
<td>139</td>
<td>217</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>S4</td>
<td>5,025</td>
<td>2,097</td>
<td>995</td>
<td>123</td>
<td>181</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>S5</td>
<td>9,482</td>
<td>3,110</td>
<td>1,012</td>
<td>107</td>
<td>143</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>S6</td>
<td>9,854</td>
<td>2,038</td>
<td>999</td>
<td>195</td>
<td>238</td>
<td>0.98</td>
</tr>
</tbody>
</table>

P value (Student’s t test) <0.05* <0.05* NS <0.1* <0.05* <0.05* <0.01* <0.05* <0.01* <0.01* <0.01* <0.01* NS

* Data obtained with PHOD and ALPS primers based on normalized data.*

**Good’s coverage, alpha diversity (Shannon-Wiener index, H’), and taxonomy (numbers of phyla, classes, orders, families, and genera).**, *, statistically significant result; NS, nonsignificant.
suggests that PHOD primers target a broader diversity of \textit{phoD}-bearing bacteria than ALPS primers.

Using ALPS primers, Tan et al. (10) found between 450 and 548 OTUs in soils fertilized with zero, medium, or high P input, with a sequencing depth between 14,279 and 16,140 reads. In contrast, Fraser et al. (16) reported lower numbers, which are in the same range as in the six soils analyzed in this study. They found between 137 and 163 OTUs in soils from organic and conventional cropping systems and prairie, with a sequencing depth of 11,537 to 54,468 reads. Thus, the number of OTUs seems to be quite variable between studies and/or soils. By applying both primers on the same soils, we found that PHOD primers targeted a larger species spectrum than ALPS primers.

**Dominant phyla harboring \textit{phoD} in six soils.** Taxonomy was assigned to most sequences using BLAST+ (27) (Fig. 3; see also Table S2 in the supplemental material). The remainder, 5,052 reads representing between 0.1 and 22% of the total filtered read number, were not assigned a taxonomic identity. In theory, the primers could amplify \textit{phoD} in archaea and eukaryotes also, as \textit{phoD} has been found in several archaeal and eukaryotic species in the IMG/M database. In the six soils studied here, both ALPS and PHOD primers amplified \textit{phoD} from bacteria only, based on identification using BLAST+.

PHOD primers targeted \textit{phoD} genes from 13 phyla (\textit{Acidobacteria}, \textit{Actinobacteria}, \textit{Bacteroidetes}, \textit{Chloroflexi}, \textit{Cyanobacteria}, \textit{Deinococcus-Thermus}, \textit{Firmicutes}, \textit{Gemmatacomadetes}, \textit{Nitrospirae}, \textit{Planctomycetes}, \textit{Proteobacteria}, \textit{Sphingobacteria}, and \textit{Verrucomicrobia}). They covered 22 classes, 38 orders, 71 families, and 113 genera. The dominant orders were \textit{Actinomycetales} (13 to 35%), \textit{Bacillales} (1 to 29%), \textit{Gloebobacteria} (1 to 18%), \textit{Rhizobiales} (18 to 100%).
showed that the phoD bacterial communities in the different soils were significantly different from each other (P < 0.001). S1 was characterized by 25% Pseudomonadales and 10% Xanthomonadales. The highest relative abundance of Bacillales (29%) was found in S2. S3 had particularly high abundances of Caulobacteriales (19%), Deinococcales (14%), and Xanthomonadales (11%). Planctomycetes were especially abundant in S4 and S6, with 18 and 19%, respectively, while S5 showed a high abundance of Gloeobacteriales (18%).

ALPS primers amplified phoD genes from 6 phyla (Actinobacteria, Bacteroidetes, Chloroflexi, Cyanobacteria, Gemmatimonadales, and Proteobacteria). In more detail, ALPS primers covered 13 classes, 22 orders, 42 families, and 64 genera. The most prevalent class was Alphaproteobacteria (55 to 92%). Rhizobiales was the dominant taxon in this class, with an overrepresentation of Methyllobacterium sp., which represented 60 to 95% of the abundance of Rhizobiales. A libshuff analysis showed that the structures of the phoD bacterial communities in the different samples were also significantly different from each other (P < 0.001).

This taxonomy analysis highlights the fact that the phoD gene is widespread across phyla and that the PHOD primers covered the phoD diversity well. PHOD primers targeted phoD genes in 13 out of the 20 phyla known to carry the phoD gene, based on the IMG/M database. PHOD primers captured a particularly large diversity of Actinobacteria, including the common soil genera Actinomyces, Arthrobacter, Kineococcus, Kitasatospora, Micrococcus, and Streptosporangium (42), and of Proteobacteria, including Azorhizobium, Rhodospirillum, Caulobacter, Geobacter, and Variovorax (43). Both Actinobacteria and Proteobacteria are known to be important for mineralization of soil organic matter and in composting processes (44, 45). Our sequencing results for soils, in accordance with the IMG/M analysis, show that a greater diversity of microorganisms than previously thought contributes to organic P mineralization by secreting PhoD.

PHOD primers amplified many sequences belonging to phyla with low abundances in the IMG/M database. These sequences were affiliated with the phyla Deinococcus-Thermus (e.g., Deinobacter sp.), Nitrosoparaeae (e.g., Nitrosirpa sp.), Spirochaetes (e.g., Spirochaeta sp.), Planctomycetes (e.g., Isosphaera sp. and Planctomyces sp.), and Verrucomicrobia (e.g., Opititus sp.). The ALPS primers did not amplify phoD genes from these phyla. Moreover, compared to the PHOD primers, the ALPS primers failed to detect phoD genes from many genera, including, e.g., Anaabaeus, Chroococcidiopsis, and Chroococcus, belonging to the Cyanobacteria. Our results support the conclusion of Tan et al. (10) that the ALPS primers have an amplification bias, restraining the amplification to a limited number of microbial taxa and overrepresenting Alphaproteobacteria, probably because of the few sequences used to design the primers (7 sequences from 4 phyla used, compared with 315 sequences from 11 phyla used here for the primer design).

Soil pH is the main driver of the phoD bacterial community. Redundancy analysis (RDA) of the PHOD-amplified samples indicated that 49.1% of the variation was explained by the two main RDA components (Fig. 4). Variance partitioning analysis showed that soil pH explained 23.7% and total P 18.3% of the variance among the communities. However, soil pH was the only environmental variable that was significantly correlated with the distribution of the samples (P = 0.03). The most divergent samples along the first RDA component axis were S1 and S4. The observed differences between these samples are likely due to the very contrasting soil and environmental properties between the sampling sites. S1 was taken from an oceanic and temperate climatic region with dense vegetation, while S4 was collected in a hot semiarid climatic region with only scattered vegetation, where lower soil microbial biomass and diversity are expected (46). S1 and S4 also exhibited the biggest difference in soil pH, which is regarded as the main environmental driving force that affects total microbial communities and activities (47, 48). Soil pH has previously been observed to be an important driver of the phoD bacterial community in the rhizosphere of wheat grown in different soils (15). Phosphatase activity can respond to changes in soil pH within days, e.g., after a lime treatment in agricultural soils (49). The second RDA component was linked mainly to total P. The phoD communities of S5 and S6 clustered together along the second-component axis, probably because these two samples were both collected in Switzerland and had high total carbon and other similar soil properties. In contrast, S1, S3, and S4 had low total C and P values.

Previous studies using the ALPS primers have reported an effect of the application of organic and conventional fertilizers, crop management, vegetation, and pH (10, 12–16, 50). The plant community has been reported to have an impact on phoD diversity and community structure in monocultures (14, 15). P fertilization has been reported to either increase (10) or reduce (12) the diversity of the phoD gene. Jorquera et al. (13) observed that P fertilization alone did not affect the phoD bacterial community structure in a Chilean Andisol pasture, while combined N and P fertilization did change the phoD bacterial community structure. While all these studies have provided some insights into the environmental driv-
ers affecting \textit{phoD} bacterial communities, they need to be interpreted with caution due to the amplification bias of the ALPS primers toward \textit{Alphaproteobacteria} described above. PHOD primers should now be applied to a wider range of soils to verify whether pH is the main driver of the \textit{phoD} bacterial community.

In conclusion, evaluation of metagenomic data sets revealed that the \textit{phoD} gene is found primarily in bacteria and is spread across 20 bacterial phyla. \textit{phoD} has been found to be ubiquitous in the environment, with terrestrial ecosystem metagenomes containing the highest relative abundance of \textit{phoD}. The newly designed PHOD primers reported here covered the large diversity of the \textit{phoD} gene better than previously published primers and amplified sequences affiliated with 13 bacterial phyla. The most prevalent \textit{phoD} genes identified in six diverse soils from Europe and Australia were affiliated with the orders Actinomycetales, Bacillales, Gloeobacterales, Rhizobiales, and Pseudomonadales. Soil pH was found to be the main environmental driver affecting the \textit{phoD} bacterial community. PHOD primers can be used as a tool to study \textit{phoD} bacterial community diversity and composition and to identify and quantify microorganisms that carry and express \textit{phoD} in the environment.

**ACKNOWLEDGMENTS**

We thank Stefan Zoller for the Perl scripts for the taxonomic analysis and the Genetic Diversity Center (Zurich, Switzerland) for technical assistance. We also acknowledge Agroscope (Switzerland) and the New South Wales Department of Primary Industry (NSW, Australia) for access to the sampling sites. This work was supported by the Swiss National Science Foundation (SNF) and by a research grant from the University of Sydney. We declare no conflicts of interest.

**REFERENCES**