Validating novel oligonucleotide primers targeting three classes of bacterial non-specific acid phosphatase genes in grassland soils

Jonathan R. Gaiero · Elizabeth Bent ·
Tandra D. Fraser · Leo M. Condron · Kari E. Dunfield

Abstract

Background and aims Microbially driven mineralization of organic phosphorus forms is of particular importance in the soil environment, where it becomes available to plants as inorganic orthophosphates. In acidic soils, microbes produce non-specific acid phosphatases (NSAPs; E.C. 3.1.3.2) which act on the most common forms of organic P in the soil. Our understanding of phosphorus turnover in soils would greatly benefit from an improvement in research tools targeting these genes. *Methods* Thus, in this study we developed two novel oligonucleotide PCR primer sets, that will enable researchers to target the present and active communities of bacteria with the genetic potential of acid phosphatase production. A total of three primer sets were validated to target the three classes of NSAPs. Utilizing Illumina MiSeq, amplicons from grassland pasture soils were sequenced. *Results* The resulting target specificity was high for all three groups; CAAP (97.2%), CBAP (99.5%), and CCAP (94.8%). Quantification of target genes by qPCR indicated measurable differences between classes, ranging from 5 log to 7.5 log for CAAP, 6 log to 8 log for CBAP, and 4 log to 5 log for CCAP. *Conclusions* The validated primer sets were specific to the target genes and identified potential quantitative differences between the NSAP classes.

Keywords Acid phosphatase · Microbial community diversity · High-throughput sequencing · Oligonucleotide primers · Organic phosphomonoesters · Soil phosphorus

J. R. Gaiero · E. Bent · K. E. Dunfield (✉)
School of Environmental Science, University of Guelph, 50 Stone Road East, Guelph, ON N1G 2W1, Canada
e-mail: dunfield@uoguelph.ca
e-mail: kdunfield@gmail.com

T. D. Fraser
Centre for Agri-Environmental Research, School of Agriculture, Policy and Development, University of Reading, Reading, UK

L. M. Condron
Faculty of Agriculture and Life Sciences, Lincoln University, Lincoln, New Zealand

Abbreviations

BLAST	Basic local alignment search tool
CAAP	Class A acid phosphatase
CBAP	Class B acid phosphatase
CCAP	Class C acid phosphatase
HTS	High-throughput sequencing
NSAP	Non-specific acid phosphatase
OTU	Operational taxonomic unit
qPCR	Quantitative real-time PCR

Introduction

Non-specific acid phosphatases (NSAPs; E.C. 3.1.3.2) are an enzyme class produced by plants and microorganisms,

and capable of hydrolyzing a variety of structurally diverse phosphomonoesters to release inorganic phosphorus (P_i) (Nakas et al. 1987; Nannipieri et al. 2011). Organisms producing NSAPs are thus able to scavenge P_i from organic P sources, which comprises approximately 30–65% of total P (Harrison 1987), and most of which (upwards of 90%) is comprised of phosphomonoesters such as inositol phosphates (phytin), sugar phosphates, phosphoproteins, and mononucleotides (Turner et al. 2002; Condron et al. 2005). In soils, microbe-derived phosphatase may be the dominant source (Rodríguez and Fraga 1999), as this microbially-scavenged P has been shown to be subsequently turned over and is then available for plants or other organisms (Tamburini et al. 2012). However, the relative contributions by microbes, plant roots, and other soil organisms, to the total phosphatase activity is not presently clear.

Bacterial NSAPs are categorized separately into three molecular classes (A, B, and C), based on protein sequence similarity (Fig. 1a) (Thaller et al. 1994; Thaller et al. 1998). All three classes are secreted, class A (CAAP) and B (CBAP) into periplasm and class C (CCAP) is retained as a membrane-bound lipoprotein, and show optimal activity in the acidic to neutral pH range. Based on protein sequence alignments, signature motifs sequences have been denoted that are distinguishing features of the three classes. The signature motif found in the majority of CAAP sequences is G-S-Y-P-S-G-H-T (Thaller et al. 1994) (PROSITE PDOC00891) (Hulo et al. 2006). CBAP contains two signature motif sequences, domain A at the N-terminal moiety (F-D-I-D-D-T-V-L-F-S-S-P) and domain B at the C-terminal moiety (Y-G-D-[A/S]-D-X-D-[I/V]) (Thaller et al. 1998). CCAP was the third novel class to be discovered, with signature motif sequences proposed as ([W]-[V/A/L]-D-[I/L]-D-E-T-[V/M]-L-X-[N/T/I-X(2)-Y), located at the N-terminus and [I/V]-[L/M]-X(2)-C-D-[N/T]-L-X near the C-terminus (Thaller et al. 1998). Phylogenetic analysis revealed CBAPs and CCAPs share sequence similarities (14–22% identity) and to plant-derived acid phosphatases (Thaller et al. 1998). They similarly belong to a superfamily of phosphohydrolases containing four invariant aspartate residues, "DDDD' (Thaller et al. 1998), and the haloacid dehalogenase (HAD) superfamily (Singh et al. 2011). CAAPs are less related to the other two classes (Fig. 1a).

Molecular studies aiming to examine the community structure of microorganisms involved in environmental P cycling have been limited, and not many PCR primer

sequences are available. There have been a few studies on quantification or community diversity analysis for alkaline phosphatase (Sakurai et al. 2008; Fraser et al. 2015a; Ragot et al. 2015; Fraser et al. 2015b; Bergkemper et al. 2016; Ragot et al. 2016), and phytase (Lim et al. 2007; Nakashima et al. 2007; Huang et al. 2009; Jorquera et al. 2014; Bergkemper et al. 2016; Morrison et al. 2016; Sanguin et al. 2016). Near full-length NSAP gene sequences for specific bacterial isolates were identified in relation to the bioremediation of uranium(VI) (Martinez et al. 2007). The lack of sequence conservation among NSAP classes have proven difficult for molecular studies aiming at the design of primers for the general quantification of NSAP gene targets, or in analyzing diversity of NSAP producing communities. Recently, a primer set was developed to target a 159 bp region of CAAP and utilized Illumina MiSeq to sequence PCR amplicons from beech forest soils (Bergkemper et al. 2016). The goal of the current study was to develop and validate primers targeting short segments of the three known bacterial NSAP classes which would be suitable for quantitative PCR and community analyses. These primers will enable researchers to appropriately parse the bacterial-plant rhizosphere interactions regarding organic P cycling which has been strongly demonstrated (Tarafdar and Jungk 1987; Nannipieri et al. 2011).

Materials and methods

Primer design

Bacterial NSAP protein sequences were obtained from NCBI GenBank® (National Center for Biotechnology information: http://ncbi.nlm.nih.gov) based on the curated list of sequences meeting stringent quality controls (Martinez et al. 2007). Sequences were aligned using the online tool Clustal Omega (Sievers et al. 2011) (http://www.ebi.ac.uk/Tools/msa/clustalo/). The multiple sequence alignment file was exported and opened using CEMAsuite (Lane et al. 2015) (https://sourceforge.net/projects/cemasuite/). Nucleotide coding sequences (CDS) were imported to CEMAsuite using the efetch tool (via NCBI). CBAP nucleotide CDS were successfully discovered and downloaded for 6 bacterial NSAPs (from a total of 9 protein sequences): *Klebsiella pneumoniae* (AF453252.1), *Shigella boydii* (CP011511.1), *Escherichia coli* (NC_000913.3), *Salmonella typhimurium* (AY841758.1), *Streptococcus pyogenes* (NC_002737.2),

Fig. 1 Neighbour joining trees of known representative NSAP sequences (**a**) and of inferred amino acid sequences for NSAP OTUs (**b**). CAAP (*green*), CBAP (*blue*), and CCAP (*red*) amino acid sequences were aligned using Clustal W (Larkin et al. 2007) and the resultant Newick file imported to the Interactive tree of life v3 (iTOL) (Letunic et al. 2016) in order to generate the phylogenetic tree. See full table of potential OTU identities in Supplementary (Tables S2-S4). OTUs were submitted to GenBank with accession numbers KY368398 to KY368571
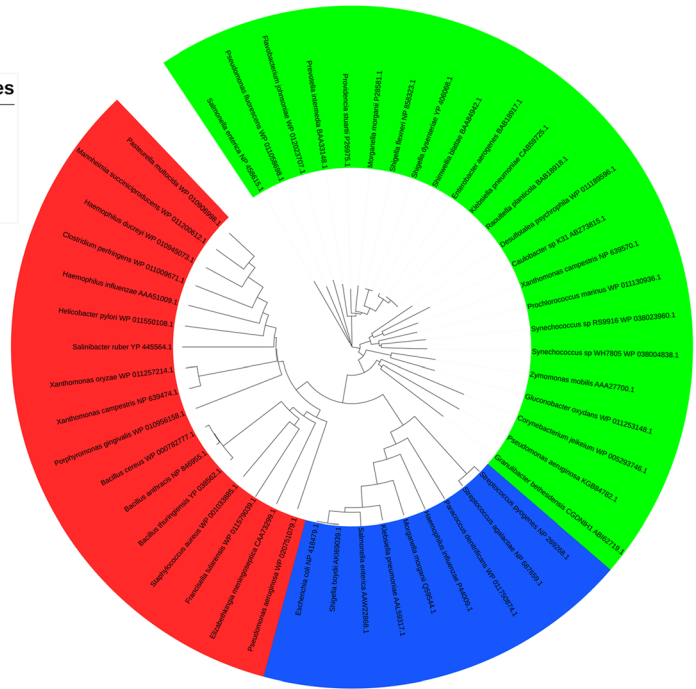
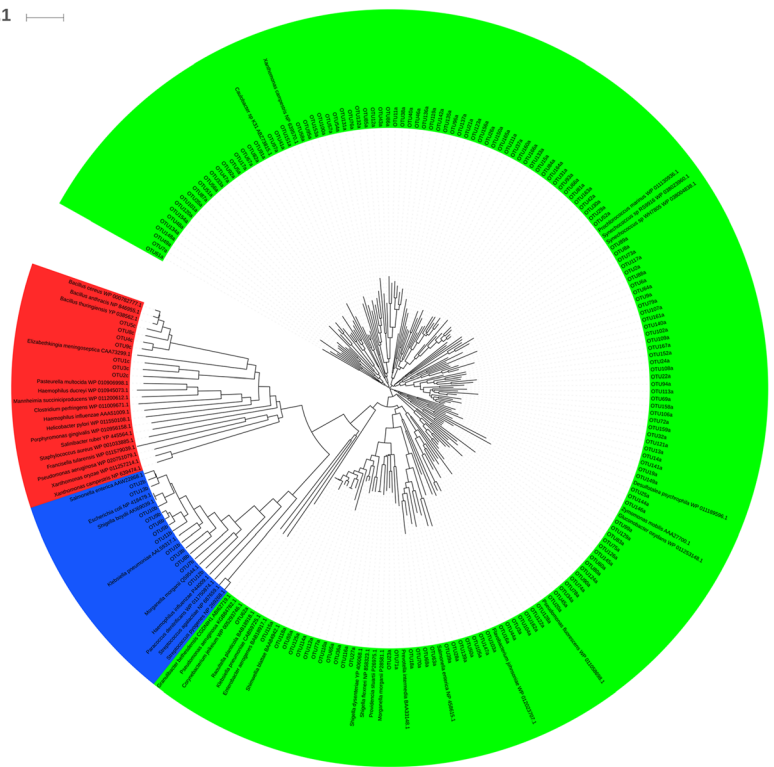and *Streptococcus agalactiae* (NC_004116.1). CCAP nucleotide CDS were successfully discovered and downloaded for 6 bacterial NSAPs (from a total of 17 protein sequences): *Salinibacter ruber* (NC_007677.1), *Xanthomonas campestris* (NC_003902.1), *Elizabethkingia meningoseptica* (Y12759.1), *Bacillus thuringiensis* (NC_005957.1), *Bacillus anthracis* (NC_003997.3), and *Haemophilus influenzae* (M68502.1). CAAP was previously targeted by this method from 15 bacterial NSAPs (Fraser et al. 2017) and was evaluated along with CBAP and CCAP in this study. Coding sequences were aligned using CEMAsuite along with protein sequence information. Primers were identified using CEMAsuite, which includes a version of Primer3 (Koressaar and Remm 2007; Untergasser et al. 2012), and selected based on binding at active site residues. An initial attempt was made to design primers based on sequences from both CBAP and CCAP (which are more similar to each other than to CAAP), that were able to amplify from pure cultures but not soil metagenomic DNA (data not shown), and thus primers were ultimately designed for individual classes (Table 1).

Specificity

Primer selection was supported by in silico hybridization results to target sequences in CEMAsuite, as well as primer-BLAST (Basic local alignment search tool) (Altschul et al. 1990) in the GenBank® NCBI database.

Primer specificity was briefly confirmed by end-point PCR using genomic DNA from pure cultures of selected bacterial species used in the design of the primers. Purified genomic DNA was acquired from DSMZ (https://www. dsmz.de/) for type strains *Bacillus thuringiensis* ATCC 10792 (CCAP), *Xanthomonas campestris* ATCC 33913 (CAAP and CCAP), and *Bacillus cereus* ATCC 14579 (CCAP). CBAP was tested using *Escherichia coli* DNA

(Supplemental Fig. S1). The PCRs were run in individual 0.2 mL thermocycling tubes on an Eppendorf AG 22331 Hamburg Mastercycler® (Eppendorf, Hamburg, Germany). PCRs were run in 25 µl reactions comprising 12.5 µl of 2X GoTaq® Green Mastermix (Promega, Madison, WI), 0.4 µM of each primer, purified genomic DNA, and nuclease free sterile water. Optimized PCR conditions were as follows: an initial 3 min denaturation step at 95 °C was used to activate the Taq polymerase, followed by 25 cycles of 30 s denaturation at 95 °C, 30 s anneal at 58 °C (CAAP) or 55 °C (CBAP and CCAP), followed by a 30 s elongation at 72 °C and then a final elongation step at 72 °C for 10 min.

Soil samples

To test the NSAP primers in natural soils, two long-term research trials in New Zealand were sampled. Trials were chosen based on soil characteristics (e.g. pH, P availability), management (e.g. aerial biomass and fertilizer application), and long history of minimal P inputs that were likely to promote NSAP gene expression. In September 1994, the "Christchurch" field trial was established at Lincoln University, Christchurch, New Zealand (S 43°38′51, E 172°28′05) on a Wakanui silt loam soil (Mottled Immature Pallic [NZ]; Udic Ustochrept [USDA]) to investigate the impacts of grassland management on soil properties (Simpson et al. 2012; Adair et al. 2013; Farrell et al. 2014). The trial area had been under mixed cropping for many years and the P fertility of the site at establishment was considered 'medium' (Olsen P – 28 mg kg^{-1}) for lowland agricultural soils in New Zealand. The "Winchmore" site was established in 1947 to investigate the management of flood irrigation (border dyke) on grazed pasture. It is one of the longest running grazed field trials in the World. The site is on the Canterbury plains near Ashburton

Table 1 List of primers designed to target the three classes of bacterial NSAP genes

Gene target	Primer	Sequence (5′ to 3′)	Anneal temp (°C)	Size (bp)[a]	Ref
phoN, phoC	CAAP-F1	CGGCTCCTATCCGTCCGG	58	155	Fraser et al. 2017
	CAAP-R1	CAACATCGCTTTGCCAGTG			
aphA	CBAP-F0	GTCGGCTTTGATATCGATGA	55	167	This study
	CBAP-R0	CGAGCGACCTCTTTTGGAAT			
olpA, lppC	CCAP-F3	TCTTCGGTGATAACTTAGGTGACTT	55	165	This study
	CCAP-R3	AATCGTATAAAGCGCCTTCC			

[a] Size approximated based on available sequences

(43°47'S, 171°48'E) on a shallow free-draining Lismore stony silt loam (Orthic Brown [NZ]; Udic Ustochrept [USDA]) (Condron and Goh 1989; Condron et al. 2012). Soil samples were collected from eight plots in duplicate, per site. Soil samples (ca. 2 g) were placed immediately into a pre-weighed sterile collection tube containing (1 mL g^{-1} soil) LifeGuard™ Soil Preservation Solution (MoBio Laboratories, Inc. Carlsbad, CA) to stabilize nucleic acids in situ for transportation to the laboratory, as described previously (Yarwood et al. 2013).

Nucleic acid extractions and reverse transcription

Soil samples were centrifuged for 5 min at 2500 x g and the Lifeguard™ (MoBio) solution poured off. The RNA and DNA were co-extracted according to manufacturer's instructions using RNA PowerSoil® Total RNA Isolation Kit (MoBio), followed by the RNA PowerSoil® DNA Elution Accessory Kit (MoBio). Extracted DNA for use in quantitative PCR (qPCR) and high-throughput sequencing (HTS) was stored at −20 °C until analysis. The RNA extracts were immediately subjected to DNase treatment to remove any residual DNA using the RQ1 RNase-Free DNase kit (Promega) according to manufacturer's protocol. The 10 μl reaction included 1 μl RNase-Free DNase, 1 μl RNase-Free 10X Buffer, 3 μl water, and 5 μl of RNA. RNA was converted to cDNA according to manufacturer's instructions using the High Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific, Waltham, MA) with random hexamer primers and stored at −20 °C. The absence of DNA contamination was confirmed by the lack of a PCR product when visualized on an agarose gel and from qPCR analysis with the RT controls.

Quantitative PCR

NSAP genes and RNA transcripts (cDNA) were quantified in the metagenomic samples. qPCRs were run on 96 well plates using a CFX86 thermocycler (Bio-Rad). All qPCRs were run in 20 μl reactions comprising 10 μl of 2X SsoFast™ EvaGreen® Supermix (Bio-Rad), 0.4 μM of each primer, 4 μl of diluted metagenomic DNA or cDNA, and nuclease free sterile water. The two-step qPCR protocol for NSAP genes was as follows: 3 min at 98 °C, followed by 40 cycles of 15 s at 98 °C and 30 s at 58 °C (CAAP) or 55 °C (CBAP and CCAP). A dissociation step (melt curve) followed the amplification to ensure specificity of the reaction, and

the PCR products were checked on an agarose gel as well. Baseline and threshold fluorescent values were automatically calculated by the Bio-Rad CFX96 Touch™ software (Bio-Rad, Hercules, CA, USA). Quantification cycle (C_q) threshold values were ensured to be above baseline and within the linear range of the exponential amplification curve. No template controls were 4 C_q or greater than sample wells. In addition, the positive result cut-off used was 1 cycle beyond the average C_q for the lowest standard. Samples were compared to a plasmid vector standard curve which conformed to MIQE guidelines (Bustin et al. 2009; Taylor et al. 2010).

High-throughput sequencing

To further demonstrate the specificity of the designed primer sets and their suitability for targeting NSAP diversity, NSAP gene targets were PCR amplified for high-throughput sequencing. Duplicate soil metagenomic DNA was amplified and pooled within sites for a remaining total of 8 samples per site. A previously established amplicon sequencing protocol was used (Fraser et al. 2015b). Amplicon libraries were prepared in three serial PCR stages in a total of 60–65 PCR cycles: 35–40 cycles to generate amplicons in the first stage, 10 cycles to add adapter tags in the second stage, and 15 cycles to add index tags in the third stage. Community composition is not expected to be altered by the last 25 cycles. In Stage 1, PCR was conducted in a 25 μl reaction; each contained 5 μl of 5X Phusion HF buffer, 0.25 μl of Phusion™ Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific), and 200 μM dNTPs, 0.4 μM of each primer, and 4 μl of diluted soil DNA or cDNA. The PCR conditions for Stage 1 were as follows: initial denaturation 98 °C for 1 min, followed by 35 cycles (CAAP and CBAP gDNA) or 40 cycles (CAAP and CBAP cDNA, and CCAP gDNA and cDNA) incorporating 98 °C for 10 s, 55 °C (CBAP and CCAP) or 58 °C (CAAP) for 15 s, and 72 °C for 10 s. Final step at 72 °C for 5 min. The PCR reactions were purified using silica spin columns (Wizard® SV Gel and PCR Clean-Up System; Promega) and eluted in 30 μl nuclease free water. PCR products were stored up to three days at 4 °C until use in Stage 2. In Stage 2, PCR was conducted to add proprietary Illumina adapter sequences (Illumina Inc., San Diego, CA). The Illumina overhang adapter sequences added to locus-specific sequences were: Forward overhang: 5′ TCGTCGGCAGCGTCAGATGT

GTATAAGAGACAG-[locus specific sequence] and Reverse overhang: 5′ GTCTCGTGGGCTCGGAGATG TGTATAAGAGACAG-[locus specific sequence]. Each reaction (25 μl) was the same as Stage 1 PCR except with 0.5 μl of Phusion™ Hot Start II High-Fidelity DNA Polymerase, 0.2 μM of each primer with Illumina adapter attached to 5′ ends, and 4 μl of purified Stage 1 amplicon as template, and only 10 cycles at 55 °C (CAAP) or 58 °C (CBAP and CCAP) annealing temperature were used. The PCR reactions were purified using spin columns as described above. In Stage 3, index tags from the Nextera XT Index Kit v2 were added to the ends of the amplicons for each sample. Primers (10 μM) consisting of Illumina's proprietary index tags and adapters were purchased from the University of Guelph Advanced Analysis Centre-Genomics (Guelph, ON, Canada). Single 25 μl PCR reactions were used containing the same reagents as in Stage 2 save that 0.375 μl of Phusion™ Hot Start II High-Fidelity DNA Polymerase was used and 1 μl of purified Stage 2 amplicon as the DNA template. The PCR cycling conditions were as for Stage 2 except 15 total cycles were used. The PCR reactions were purified by magnetic beads (Agencourt AmPure XP; Beckman Coulter, Brea, CA) and resuspended in 25 μl and frozen at −20 °C until submission to the University of Guelph Advanced Analysis Centre Genomics (Guelph, ON, Canada) for sequencing. Prior to sequencing, libraries were normalized by Sequalprep (Thermo Fisher Scientific), and library quality assessed from a random sample of 12 samples using a Bioanalyzer DNA1000 chip (Agilent, Santa Clara, California). Multiplexed sample sequencing was conducted using an Illumina MiSeq Reagent Kit v2 (500-cycle) (Illumina Inc., San Diego, CA) producing paired end reads 250 bp in length. Unprocessed FASTQ files were obtained for subsequent analysis.

Bioinformatics pipeline

Overlapping paired end reads were assembled using PEAR software (Zhang et al. 2014). Primers and poor quality sequences were removed using cutadapt software (Martin 2011). After quality filtering, a range of 21,834 to 97,649 reads in between 115 and 120 bp in length was recovered for CAAP amplicons; 32,921 to 178,375 reads 120 to 126 bp in length were recovered for CBAP, and 25,901 to 188,412 reads 108 to 115 bp in length were recovered for CCAP (Supplemental Table S1). The UPARSE pipeline was used to cluster

sequences into centroid operational taxonomic units (OTUs) (Edgar 2013). An 88% sequence similarity threshold was used for OTU clustering for CAAP and CCAP, and an 85% threshold for CBAP, based on sequence similarity found between NSAP amino acid sequences. Reads were mapped back to OTUs into an OTU table using USEARCH software (Edgar 2010). Reads per OTU were standardized by the number of sample reads to allow comparison among samples. To calculate primer specificity, the percent target reads was calculated as mapped target reads out of the sum of target and non-target read abundances per sample. Non-target reads were considered reads removed during the cutadapt step based on length or those not containing the primers, along with mapped reads that were identified by NCBI blast as non-target. Average primer specificity was calculated for each primer set from all available samples. Samples with read counts fewer than 1000 were not included in analyses (one sample from the Winchmore site).

Rarefaction plots

Rarefaction curves were made using OTU tables generated by USEARCH, formatted for use by mothur v.1.38.1 (Schloss et al. 2009). The sum of reads per OTU for each category (DNA, cDNA) were pooled.

Phylogeny of OTUs

Centroid sequences for each OTU were screened for frameshift mutations using FrameBot (Fish et al. 2013) and for non-target sequences (non-acid phosphatase sequences) using BLASTx or BLASTn searches against NCBI's non-redundant 'nr' (protein) or nt (nucleotide) databases. This resulted in 158 'target' (acid phosphatase) OTUs for CAAP (this number includes acid phosphatase frameshift mutant centroid OTUs, which made up a tiny fraction of sequences), and 9 'non-target' (not acid phosphatase) OTUs. For CBAP, there were 12 target OTUs and one non-target OTU, and for CCAP, 7 target and four non-target OTUs. Frameshift mutants were removed from analysis. The target OTU centroid sequences were submitted to NCBI GenBank with accession numbers KY368398 to KY368571. The highest scoring genus match from NCBI Genbank using protein-protein blast was included in a table for each OTU (Supplemental Tables S2-S4). The E-values were at minimum E-10, except for CAAP OTU 62 (OTU62a)

which was E-07. The average relative abundance of bacterial orders (CAAP) or genera (CBAP and CCAP) was calculated from the DNA OTU libraries for each site.

Phylogenetic tree

Translated CAAP, CBAP, and CCAP NSAP sequenced amplicons and representative sequences from Genbank were aligned using Clustal W (Larkin et al. 2007) and a neighbor joining tree was generated. The resultant Newick file was imported to the Interactive tree of life v3 (iTOL) (Letunic et al. 2016) in order to present the phylogenetic tree.

Results

Primer specificity

The three oligonucleotide primer sets amplified targets of approximately 150 bp in size as compared to the molecular ladder, and matched their expected sizes. From the initial screening of primers against pure culture DNA (Fig. S1) and clone libraries from soil metagenomic DNA (data not shown), the designed NSAP class primers were specific to the intended targets. The dissociation (melt) curves associated with the three NSAP primer sets indicated a single major product was found following qPCR (Figs. S2-S4). Average melt peak temperatures of NSAP amplicons varied <1 °C between samples and sites. This also corresponded with the single band observed in agarose gels following gel electrophoresis of end-point PCR products. High-throughput sequencing further confirmed the amplified targets to be highly specific to the intended classes of NSAP genes with a low rate of non-specific amplification (Table 2).

NSAP phylogeny

The OTUs were categorized at levels of sequence similarity appropriate based on known sequence data for the classes; 88% for CAAP and CCAP, and 85% for CBAP. At these levels an apparent difference in OTU richness was observed; CAAP had almost 100 fold more OTUs than CCAP and 14 fold more than CBAP (Table 2).

A phylogenetic tree was generated to indicate the relationship of the three sequenced classes of NSAP genes as compared to known sequences available on Genbank (Fig. 1b). CBAP and CCAP are more closely related than either is to CAAP, as shown previously (Gandhi and Chandra 2012). Upon inspection of the generated tree, all OTUs grouped within their expected clusters. Of note was a group of three CAAP sequences taken from NCBI (*Granulibacter bethesdensis* CGDNIH1, ABI62719.1; *Pseudomonas aeruginosa*, KGB84782.1; *Corynebacterium jeikeium*, WP_005293746.1) that clustered more closely with CBAP and CCAP than to the other CAAP sequences. The CAAP OTU 62 (OTU62a), identified as the genus *Pandoraea* with 64% similarity also clustered closely to this group, indicating that the CAAP primers can capture the diversity of published sequences. However, many OTUs remain unidentified based on the available NSAP gene sequences on NCBI. This was most evident for CAAP, which had the greatest number of unique sequences. However, for all three classes there were clusters of unique OTUs which highlights the need to further explore the diversity of bacterial NSAP genes.

Taxonomic identification of generated OTUs was attempted using protein blast on the nr NCBI database (Table 3); however, for many sequences, several results were plausible, and the closest match was recorded (Tables S2-S4). Taxonomic identification of CAAP was done at the level of bacterial order and a relative abundance graph was generated (Fig. 2a). CBAP and CCAP OTUs

Table 2 Abundance of OTUs and target specificity using NSAP oligonucleotide primer sets

Class primer set	Number of OTUs	Specificity (%)[a]
CAAP	158	97.2
CBAP	12	99.5
CCAP	7	94.8

[a] Target specificity was calculated by subtracting the non-targets reads from the total mappable reads, divided by the sum of mappable and length filtered reads

Table 3 Richness of detected sequences from each NSAP class primer set. OTUs were compared to the nr database on Genbank via protein-protein Blast. See full table of results in Supplementary (S2-S4)

NSAP	Phyla	Class	Order	Family
CAAP	2	5	13	19
CBAP	1	1	1	1
CCAP	2	2	2	2

Fig. 2 Relative abundance of taxonomic groups in descending order of abundance, at the two sites Christchurch and Winchmore; **a** CAAP – bacterial orders; **b** CBAP – bacterial genera; **c** CCAP – bacterial genera
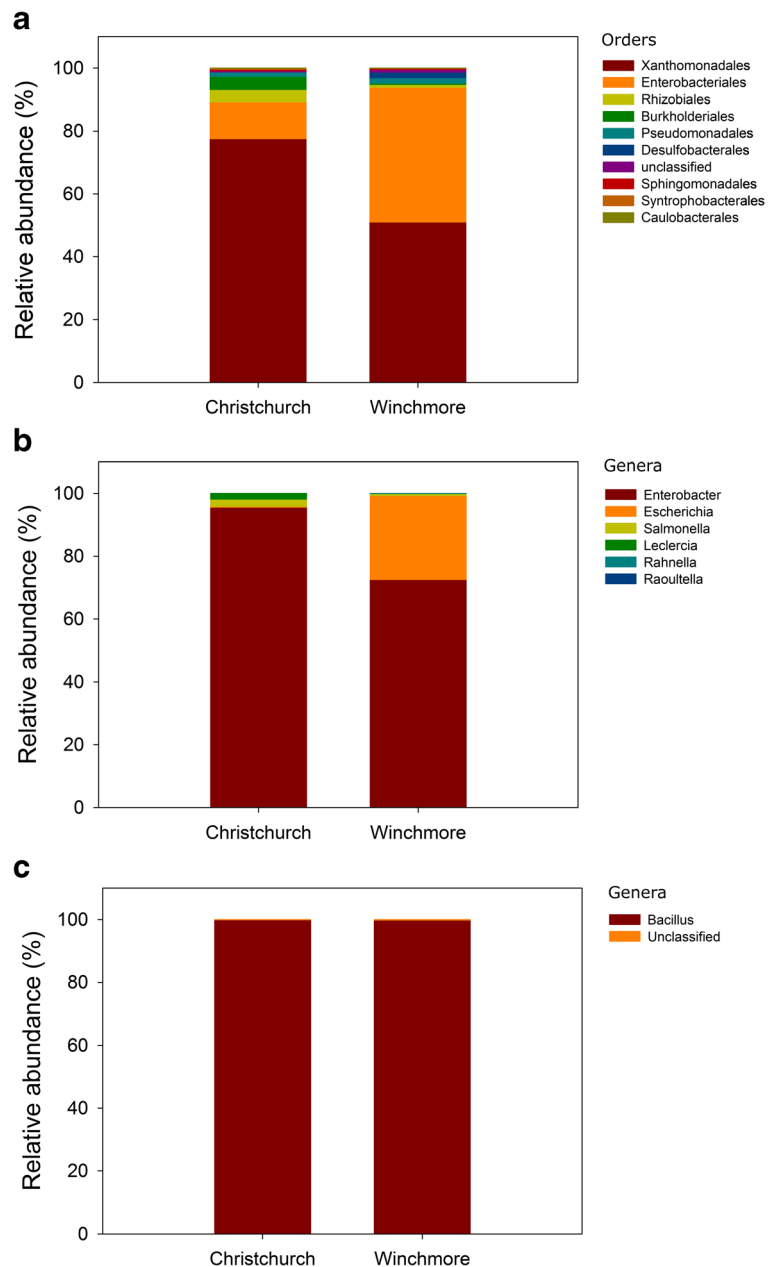
were more easily identified at the genus level (Fig. 2b and c). The orders Xanthomonadales and Enterobacteriales dominated the pool of CAAP OTUs in both grassland soils tested (Fig. 2a). An elevated abundance of Enterobacteriales was observed at the Winchmore site, whereas Rhizobiales and Burkholderiales were more abundant at the Christchurch site. The only taxon consistently found in the cDNA pool but not DNA was OTU21a. The closest match of this OTU was to *Stenotrophomonas* spp. at 85% identity, which is an

organism that was commonly found among the OTUs. The CBAP primer set also successfully targeted the group of Gammaproteobacteria sequences from which the primers were designed, dominated by the family Enterobacteriaceae. The genus *Enterobacter* was dominant at both sites (Fig. 2b). *Escherichia* comprised on average 26% of the community at Winchmore, and other genera made up a small proportion at both sites. The CCAP primer set almost solely targeted the genus *Bacillus*, and with a minor proportion of unclassified OTUs potentially

belonging to the order Sphingobacteriales (Fig. 2c). The blastp queries matching the genus *Bacillus* were shown to have a sequence identity up to 100%, whereas the closest match to the remaining OTUs was at most 65% sequence identity (Table S3). While certain *Xanthomonas* spp. are known to harbour both the CAAP and CCAP NSAP classes of genes, a preliminary experiment with pure culture DNA indicated that the CCAP primer set was not able to target this gene from type strain *Xanthomonas campestris* ATCC 33913 (Fig. S1). This genus was still identified using CAAP primers in these samples. The CCAP primer set thus likely errs as overly specific, potentially limiting non-target amplification.

Rarefaction curves were generated for the metagenomic DNA and cDNA at each site sampled and each NSAP class targeted that indicated a sufficient sequencing depth was reached for the number of OTUs identified (Fig. 3a, b and c). The differences observed in number of OTUs generated from each pool of DNA and cDNA are indicated by separated curves; however, the rate at which sequencing depth was sufficient to cover each group of sequences remained the same between sites and between pools of nucleic acids.

Quantification from soil

In the current study, all three NSAP primers were used successfully to quantify target genes from both soil sites tested (Table 4). The standard curves generated for each NSAP class primer set were within the MIQE guidelines (Bustin et al. 2009; Taylor et al. 2010). The standard curves had efficiencies and R^2 of 95% and 0.995, 100.5% and 0.998, and 97.7% and 0.997, for CAAP, CBAP, and CCAP, respectively.

NSAP gene targets were generally highest in abundance at the Christchurch site, despite similar soil characteristics with Winchmore. Reverse-transcribed RNA transcriptomes (cDNA) were also profiled and successfully amplified CAAP and CBAP gene targets at both sites, up to 4 log copies per gram dry soil (Table 4). CCAP transcripts were present but below detection at both sites; however, it was still clear that this group of bacteria was active in several of the soil samples tested since CCAP transcripts were analyzed by HTS. The distribution of NSAP classes is currently unknown in the soil environment, but measurements from the current study indicate that CAAP and CBAP are more abundant than CCAP in the New Zealand grassland soils tested.
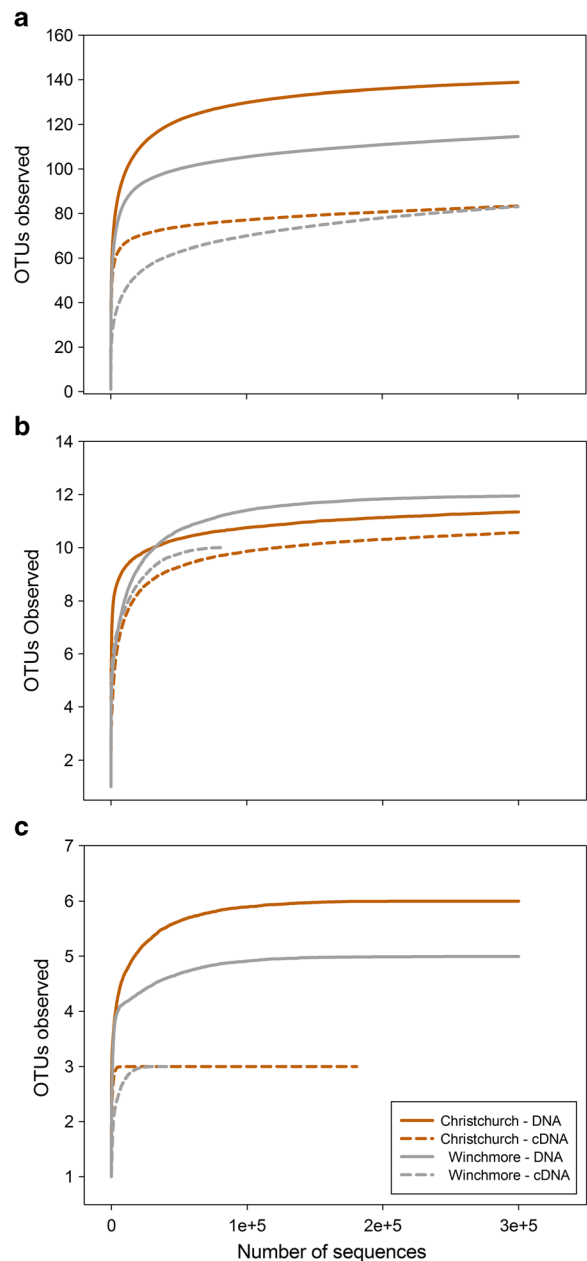
Fig. 3 Rarefaction curves for the two sites (Christchurch and Winchmore); **a** CAAP, **b** CBAP, and (**c**) CCAP. DNA and cDNA results were separately pooled by site. OTUs were clustered at a similarity cutoff of 88% (CAAP), 85% (CBAP), or 88% (CCAP)

Discussion

In the current study, we have aimed to increase the number of tools available for researchers to examine microbes influencing the important turnover of P in soils and other environments. The bacterial NSAPs may

Table 4 Summary of qPCR results targeting NSAP genes (log gene copies g^{-1} dry soil) at the two grassland sites, along with soil characteristics; Olsen P (mg kg^{-1} soil), acid phosphatase activity (ACP; μmol g^{-1} dry soil h^{-1})

Site	Soil type	pH	Olsen P	ACP	CAAP[a] (RNA)	CBAP[a] (RNA)	CCAP[a] (RNA)
Christchurch	Wakanui silt loam	5.6–5.8	15–36	10–18	6–7.5 (0–3)	7–8 (0–4)	4–5 (BD)[b]
Winchmore	Lismore stony silt loam	5.7	8–45	9–12	5–6 (0–2)	6–7 (0–2)	4–5 (BD)

[a] Standard curve efficiencies and R^2 values were within MIQE guidelines [34,35]

[b] BD – Below detection

comprise a significant proportion of the genetic potential for organic P mineralization in soil, but to date have not been characterized apart from certain bacterial isolates (Rossolini et al. 1998). Total phosphatase activity is due to the combined activity of many phosphatase enzymes, that are not easily distinguished.

It was clear that a molecular approach targeting the specific NSAP classes would be required for analyzing the bacterial functional communities. Three sets of non-degenerate oligonucleotide primers were validated in this study to specifically target the individual classes of bacterial NSAPs, tested in two acidic grassland soils. These primers were first validated using pure culture DNA and soil metagenomic DNA and reverse-transcribed RNA transcripts. Community analyses using qPCR and HTS indicated that the primer sets were specific and suitable to targeting the three NSAP classes.

qPCR is a remarkably useful tool for specifically identifying the abundance of target genes in a sample. The NSAP class primers in this study were in part designed to be used in qPCR for future research in phosphorus cycling in environmental systems. The CAAP primers have been shown to successfully amplify targets in soils from P.E.I., Canada (Fraser et al. 2017), as well as a podzolic soil from the U.K, and from a heavy clay humic vertisol from Manitoba, Canada (data not shown). Quantification of CBAP and CCAP genes and transcripts has not previously been performed. Differences in gene and transcript abundance were found when comparing the three NSAP classes in the grassland soils tested, which may or may not hold for other soils tested. Furthermore, genetic potential for phosphatase production may vary considerably even between closely related species (Rossolini et al. 1998). Different NSAP classes have been identified in the same organism, and NSAP genes may be found in multiple copies within the genome, which indicates varying roles may

be played by each class. This may have implications for quantitative and HTS analyses. An assessment of the few available genomes on NCBI revealed an approximate number of NSAP alleles. CAAP (Genbank: cd03397) was found between one and five copies per genome, but in general, one to two copies were most common. For instance, *Stenotrophomonas maltophilia* was commonly found in the CAAP OTU library and the genome of strain K279a was shown to contain a single CAAP gene copy identified by protein homology. However, at the upper range, *Shigella* was found to contain five plasmid-based genes. NSAP genes found on plasmids may therefore be horizontally transferred. Horizontal gene transfer (HGT) of CAAP genes was previously reported in *Salmonella* (Groisman et al. 1992). In addition, HGT of CBAP genes may have been common in Enterobacteria (Rossolini et al. 1998). CBAP (Genbank: class B acid phosphatase, *aphA*) alleles were found up to twice per genome. Two copies of CBAP identified by homology in the genome of *Enterobacter aerogenes* KCTC 2190 were calculated to share 46% protein sequence similarity. *Enterobacter* was a dominant genus in the CBAP OTU library, and thus CBAP primers may target both copies. CCAP (Genbank: 5′-nucleotidase, lipoprotein e(P4) family) appeared to be found most commonly as one copy per genome. Two copies were found in *Clostridium botulinum* BKT015925, but this organism was not found in the current samples.

To our knowledge there is no prior environmental data comparing the prevalence or activity of the three classes of NSAPs. Another research group successfully targeted the CAAPs in beech forest soil using a different oligonucleotide primer set and were able to target a total of 7 phyla, 20 orders, and 36 families based on HTS results (Bergkemper et al. 2016). Based on profile Hidden Markov Models (HMM), the dominant taxonomic

order they found was Rhizobiales, comprising 57% of sequences, followed by Xanthomonadales (29%), Enterobacteriales (8%), and Burkholderiales (6%). While the nucleotide clustering method used in this study differs from the profile HMMs used to analyze the beech forest soils, similar bacterial orders were found. Differences in relative abundance of the bacterial orders can be due to the primers or soil samples tested. Members of the family Xanthomonadaceae (order Xanthomonadales) are known to contain several organic P mineralizing genes (Jorquera et al. 2008). Within this family, *Stenotrophomonas* spp. have been shown to be known colonizers of plants and are considered plant growth promoting (Majeed et al. 2015), and important contributors of organic P mineralization from P pools including phosphonates (Fox et al. 2014). The dominant *Enterobacter* CBAP OTUs and *Bacillus* CCAP OTUs, indicated a potential lack of diversity found in these soils. The Enterobacteriaceae are well known phosphatase producers (Pompei et al. 1993), and *Enterobacter* spp. have previously been studied for phosphate solubilization, P mineralization, and in plant growth promotion studies (Gyaneshwar et al. 1999; Jorquera et al. 2008; Majeed et al. 2015). *Bacillus* spp. are also well known for phosphorus solubilization and have been studied for biotechnological applications including using NSAP for bioremediation (Rodríguez et al. 2006).

Overall, Proteobacteria were dominant in the targeted CAAPs found in the current and previous studies. There is a strong presence of Alphaproteobacteria and Gammaproteobacteria sequences available in online databases. In fact, Proteobacteria currently represent approximately 80% of the known NSAP CAAP sequences found in NCBI Genbank (Genbank; cd03397). The three NSAP class primer sets did target the dominant taxonomic groups found within each class, but may not necessarily cover the full range of potential NSAP gene sequences in these soils. Future studies should use this validated methodology both to expand upon the diversity of sequences in different environmental samples, and to incorporate useable NSAP gene abundance and gene expression data with biogeochemical and ecological data.

References

Adair KL, Wratten S, Lear G (2013) Soil phosphorus depletion and shifts in plant communities change bacterial community structure in a long-term grassland management trial. Environ Microbiol Rep 5:404–413. doi:10.1111/1758-2229.12049

Altschul SFS, Gish W, Miller W et al (1990) Basic local alignment search tool. J Mol Biol 215:403–410. doi:10.1016/S0022-2836(05)80360-2

Bergkemper F, Kublik S, Lang F et al (2016) Novel oligonucleotide primers reveal a high diversity of microbes which drive phosphorous turnover in soil. J Microbiol Methods 125:91–97. doi:10.1016/j.mimet.2016.04.011

Bustin SA, Benes V, Garson JA et al (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 55:611–622. doi:10.1373/clinchem.2008.112797

Condron LM, Goh KM (1989) Effects of long-term phosphatic fertiliser applications on amounts and forms of phosphorus applications on amounts and forms of phosphorus in soils under irrigated pasture in New Zealand. J Soil Sci 40:383–395. doi:10.1111/j.1365-2389.1989.tb01282.x

Condron LM, Turner BL, Cade-Menun BJ (2005) Chemistry and dynamics of soil organic phosphorus. Phosphorus Agric Environ 46:87–121. doi:10.2134/agronmonogr46.c4

Condron L, Black A, Wakelin S (2012) Effects of long-term fertiliser inputs on the quantities of organic carbon in a soil profile under irrigated grazed pasture. New Zeal J Agric Res 55:161–164. doi:10.1080/00288233.2012.662898

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461. doi:10.1093/bioinformatics/btq461

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10:996–998. doi:10.1038/nmeth.2604

Farrell M, Prendergast-Miller M, Jones DL et al (2014) Soil microbial organic nitrogen uptake is regulated by carbon availability. Soil Biol Biochem 77:261–267. doi:10.1016/j.soilbio.2014.07.003

Fish JA, Chai B, Wang Q et al (2013) FunGene: the functional gene pipeline and repository. Front Microbiol 4:1–14. doi:10.3389/fmicb.2013.00291

Fox A, Kwapinski W, Griffiths BS, Schmalenberger A (2014) The role of sulfur- and phosphorus-mobilizing bacteria in biochar-induced growth promotion of *Lolium perenne*. FEMS Microbiol Ecol 90:78–91. doi:10.1111/1574-6941.12374

Fraser T, Lynch DH, Entz MH, Dunfield KE (2015a) Linking alkaline phosphatase activity with bacterial *phoD* gene abundance in soil from a long-term management trial. Geoderma 257–258:115–122. doi:10.1016/j.geoderma.2014.10.016

Fraser TD, Lynch DH, Bent E et al (2015b) Soil bacterial *phoD* gene abundance and expression in response to applied phosphorus and long-term management. Soil Biol Biochem 88:137–147. doi:10.1016/j.soilbio.2015.04.014

Fraser TD, Lynch DH, Gaiero J et al (2017) Quantification of bacterial non-specific acid (*phoC*) and alkaline (*phoD*) phosphatase genes in bulk and rhizosphere soil from organically managed soybean fields. Appl Soil Ecol 111:48–56. doi:10.1016/j.apsoil.2016.11.013

Gandhi NU, Chandra SB (2012) A comparative analysis of three classes of bacterial non-specific acid phosphatases and archaeal phosphoesterases: evolutionary perspective. Acta Inform Medica 20:167–173. doi:10.5455/aim.2012.20.167-173

Groisman EA, Saier MH, Ochman H (1992) Horizontal transfer of a phosphatase gene as evidence for mosaic structure of the *Salmonella* genome. EMBO J 11:1309–1316

Gyaneshwar P, Parekh LJ, Archana G et al (1999) Involvement of a phosphate starvation inducible glucose dehydrogenase in soil phosphate solubilization by *Enterobacter asburiae*. FEMS Microbiol Lett 171:223–229. doi:10.1016/S0378-1097(99)00003-8

Harrison AF (1987) Soil organic phosphorus : a review of world literature. CAB International

Huang H, Shi P, Wang Y et al (2009) Diversity of beta-propeller phytase genes in the intestinal contents of grass carp provides insight into the release of major phosphorus from phytate in nature. Appl Environ Microbiol 75:1508–1516. doi:10.1128/AEM.02188-08

Hulo N, Bairoch A, Bulliard V et al (2006) The PROSITE database. Nucleic Acids Res 34:D227–D230. doi:10.1093/nar/gkj063

Jorquera M, Martínez O, Maruyama F et al (2008) Current and future biotechnological applications of bacterial phytases and phytase-producing bacteria. Microbes Environ 23:182–191. doi:10.1264/jsme2.23.182

Jorquera M, Inostroza N, Lagos L et al (2014) Bacterial community structure and detection of putative plant growth-promoting rhizobacteria associated with plants grown in Chilean agro-ecosystems and undisturbed ecosystems. Biol Fertil Soils 50:1141–1153. doi:10.1007/s00374-014-0935-6

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. Bioinformatics 23:1289–1291. doi:10.1093/bioinformatics/btm091

Lane CE, Hulgan D, O'Quinn K, Benton MG (2015) CEMAsuite: Open source degenerate PCR primer design. Bioinformatics 31:3688–3690. doi:10.1093/bioinformatics/btv420

Larkin MA, Blackshields G, Brown NP et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948. doi:10.1093/bioinformatics/btm404

Letunic I, Bork P, D.H. H et al (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44:W242–W245. doi:10.1093/nar/gkw290

Lim BL, Yeung P, Cheng C, Hill JE (2007) Distribution and diversity of phytate-mineralizing bacteria. ISME J 140:321–330. doi:10.1038/ismej.2007.40

Majeed A, Kaleem Abbasi M, Hameed S et al (2015) Isolation and characterization of plant growth-promoting rhizobacteria from wheat rhizosphere and their effect on plant growth promotion. Front Microbiol 6:1–10. doi:10.3389/fmicb.2015.00198

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetJournal 17:10. doi:10.14806/ej.17.1.200

Martinez RJ, Beazley MJ, Taillefert M et al (2007) Aerobic uranium (VI) bioprecipitation by metal-resistant bacteria isolated from radionuclide- and metal-contaminated subsurface soils. Environ Microbiol 9:3122–3133. doi:10.1111/j.1462-2920.2007.01422.x

Morrison E, Newman S, Bae HS et al (2016) Microbial genetic and enzymatic responses to an anthropogenic phosphorus gradient within a subtropical peatland. Geoderma 268:119–127. doi:10.1016/j.geoderma.2016.01.008

Nakas JP, Gould WD, Klein DA (1987) Origin and expression of phosphatase activity in a semi-arid grassland soil. Soil Biol Biochem 19:13–18. doi:10.1016/0038-0717(87)90118-0

Nakashima BA, McAllister TA, Sharma R, Selinger LB (2007) Diversity of phytases in the rumen. Microb Ecol 53:82–88. doi:10.1007/s00248-006-9147-4

Nannipieri P, Giagnoni L, Landi L, Renella G (2011) Phosphorus in action: biological processes in soil phosphorus cycling. In: Soil biology. Springer, Heidelberg Dordrecht London New York, pp 215–244

Pompei R, Ingianni A, Foddis G et al (1993) Patterns of phosphatase activity among enterobacterial species. Int J Syst Bacteriol 43:174–178. doi:10.1099/00207713-43-1-174

Ragot SA, Kertesz MA, Bünemann EK (2015) *phoD* alkaline phosphatase gene diversity in soil. Appl Environ Microbiol 81:7281–7289. doi:10.1128/AEM.01823-15

Ragot SA, Huguenin-Elie O, Kertesz MA et al (2016) Total and active microbial communities and *phoD* as affected by phosphate depletion and pH in soil. Plant Soil 408:15–30. doi:10.1007/s11104-016-2902-5

Rodríguez H, Fraga R (1999) Phosphate solubilizing bacteria and their role in plant growth promotion. Biotechnol Adv 17:319–339. doi:10.1016/S0734-9750(99)00014-2

Rodríguez H, Fraga R, Gonzalez T, Bashan Y (2006) Genetics of phosphate solubilization and its potential applications for improving plant growth-promoting bacteria. Plant Soil 287:15–21. doi:10.1007/s11104-006-9056-9

Rossolini GM, Schippa S, Riccio ML et al (1998) Bacterial nonspecific acid phosphohydrolases: physiology, evolution and use as tools in microbial biotechnology. Cell Mol Life Sci 54:833–850. doi:10.1007/s000180050212

Sakurai M, Wasaki J, Tomizawa Y et al (2008) Analysis of bacterial communities on alkaline phosphatase genes in soil supplied with organic matter. Soil Sci Plant Nutr 54:62–71. doi:10.1111/j.1747-0765.2007.00210.x

Sanguin H, Wilson NL, Kertesz MA (2016) Assessment of functional diversity and structure of phytate-hydrolysing bacterial community in *Lolium perenne* rhizosphere. Plant Soil 401:151–167. doi:10.1007/s11104-015-2512-7

Schloss PD, Westcott SL, Ryabin T et al (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541. doi:10.1128/AEM.01541-09

Sievers F, Wilm A, Dineen D et al (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. doi:10.1038/msb.2011.75

Simpson M, McLenaghen RD, Chirino-Valle I, Condron LM (2012) Effects of long-term grassland management on the chemical nature and bioavailability of soil phosphorus. Biol Fertil Soils 48:607–611. doi:10.1007/s00374-011-0661-2

Singh H, Reilly TJ, Tanner JJ (2011) Structural basis of the inhibition of class C acid phosphatases by adenosine 5′-phosphorothioate. FEBS J 278:4374–4381. doi:10.1111/j.1742-4658.2011.08360.x

51

Tamburini F, Pfahler V, Bünemann EK et al (2012) Oxygen isotopes unravel the role of microorganisms in phosphate cycling in soils. Environ Sci Technol 46:5956–5962. doi:10.1021/es300311h

Tarafdar JC, Jungk A (1987) Phosphatase activity in the rhizosphere and its relation to the depletion of soil organic phosphorus. Biol Fertil Soils 3:199–204. doi:10.1007/BF00640630

Taylor S, Wakem M, Dijkman G et al (2010) A practical approach to RT-qPCR-Publishing data that conform to the MIQE guidelines. Methods 50:S1–S5. doi:10.1016/j.ymeth.2010.01.005

Thaller MC, Berlutti F, Schippa S et al (1994) Characterization and sequence of PhoC, the principal phosphate-irrepressible acid phosphatase of *Morganella morganii*. Microbiology 140:1341–1350. doi:10.1099/00221287-140-6-1341

Thaller MC, Schippa S, Rossolini GM (1998) Conserved sequence motifs among bacterial, eukaryotic, and archaeal phosphatases that define a new phosphohydrolase superfamily. Protein Sci 7:1647–1652. doi:10.1002/pro.5560070722

Turner BL, Papházy MJ, Haygarth PM, McKelvie ID (2002) Inositol phosphates in the environment. Philos Trans R Soc Lond Ser B Biol Sci 357:449–469. doi:10.1098/rstb.2001.0837

Untergasser A, Cutcutache I, Koressaar T et al (2012) Primer3–new capabilities and interfaces. Nucleic Acids Res 40:e115–e115. doi:10.1093/nar/gks596

Yarwood S, Brewer E, Yarwood R et al (2013) Soil microbe active community composition and capability of responding to litter addition after 12 years of no inputs. Appl Environ Microbiol 79:1385–1392. doi:10.1128/AEM.03181-12

Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina paired-end reAd mergeR. Bioinformatics 30:614–620. doi:10.1093/bioinformatics/btt593