

# SCIENTIFIC REPORTS



OPEN

## Development of Quantitative Real-time PCR Assays for Different Clades of “*Candidatus Accumulibacter*”

An Ni Zhang, Yanping Mao & Tong Zhang

Received: 13 October 2015

Accepted: 18 March 2016

Published: 04 May 2016

We designed novel quantitative real-time polymerase chain reaction (qPCR) primers for the polyphosphate kinase 1 (*ppk1*) gene, targeting eight individual “*Candidatus Accumulibacter*” (referred to as *Accumulibacter*) clades. An evaluation of primer sets was conducted regarding the coverage, specificity, and PCR efficiency. (i) All primer sets were designed to cover all available sequences of the target clade. (ii) The phylogenetic analysis of the sequences retrieved from the qPCR products by each primer set demonstrated a high level of specificity. (iii) All calibration curves presented high PCR efficiencies in the range of 85–112% ( $R^2 = 0.962–0.998$ ). In addition, the possible interference of non-target amplicons was individually examined using the qPCR assay for 13 *Accumulibacter* clades, which were either undetected or showed negligible detection. With the primers designed by other research groups, a highly selective and sensitive qPCR-based method was developed to quantify all *Accumulibacter* clades, with the exception of Clade IE, in one assay, which enables more comprehensive insights into the community dynamics. The applicability to environmental samples was demonstrated by profiling the *Accumulibacter* clades in activated sludge samples of nine full-scale wastewater treatment plants.

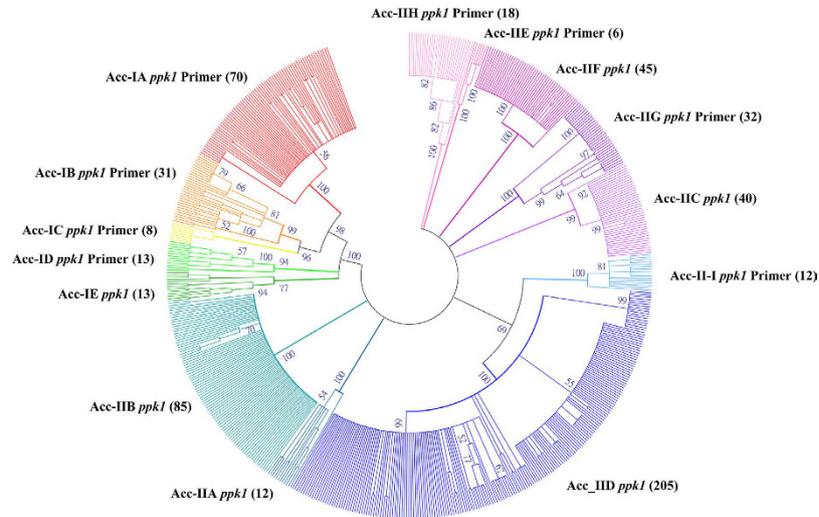
Enhanced biological phosphorus removal (EBPR) has been highly utilized to remove phosphorous in wastewater treatment plants (WWTPs) around the world for nearly four decades. Polyphosphate accumulating organisms (PAOs) serve a major role in phosphorus removal by accumulating excess phosphate from wastewater inside their cell bodies. “*Candidatus Accumulibacter*” (referred to as *Accumulibacter*) has been identified as one of the primary unculturable PAOs in both full-scale WWTPs<sup>1–3</sup> and lab-scale reactors<sup>4</sup> that conduct EBPR.

The 16S ribosomal RNA (rRNA) method was used as the primary approach to identify the occurrence and phylogeny of *Accumulibacter* in EBPR systems, until their inability to partition *Accumulibacter* into ecotypes, owing to their high similarities, was proven. Then, polyphosphate kinase 1 (*ppk1*) gene encoding the major enzyme for catalysing polyphosphate synthesis was applied as a genetic marker to partition the *Accumulibacter* lineage into Type I and Type II (consists of four clades)<sup>5,6</sup>. Subsequently, quantitative real-time PCR (qPCR) primers that target *Accumulibacter* Type I and Clades IIA, IIB, IIC and IID, were designed and verified to differentiate divergence among *Accumulibacter* ecotypes<sup>6</sup>.

Knowledge of the phylogenetic and genetic diversity within the *Accumulibacter* lineage was then substantially expanded. In 2008, Peterson *et al.*<sup>7</sup> subdivided the *Accumulibacter* Type I into five clades and discovered Clades IIE, IIF and IIG of Type II, based on the phylogenetic distance of the *ppk1* gene. The *Accumulibacter* lineage comprises five clades in Type I and nine clades in Type II, including the novel clades proposed by Mao *et al.*<sup>8</sup>.

In contrast with the rapid exploration inside the *Accumulibacter* lineage, the primer design work for recent discovered clades proceeded slowly. Only one primer set has been designed for Clade IIF<sup>9</sup>; no sets have been designed for the other identified clades (IA, IB, IC, ID, IE, IIE, IIG, IIH, and II-I). Retaining the coverage and sensitivity to the target clades has become a challenge for the previously designed qPCR assay. A recent study has revealed that the total *Accumulibacter* lineage in 10 of 18 WWTPs, as determined by former primer sets for the *ppk1* gene<sup>6</sup>, accounted for less than 50% of the total *Accumulibacter* determined by the *Accumulibacter* 16S rRNA gene<sup>8</sup>. The first reason for this result was that Clade IB was not covered by the primer set Acc-ppk1-763f

Environmental Biotechnology Laboratory, Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong. Correspondence and requests for materials should be addressed to T.Z. (email: zhangt@hku.hk)



**Figure 1.** Condensed phylogenetic tree for 605 available *Accumilibacter ppk1* gene sequences, which were constructed by the maximum likelihood method using the Tamura-Nei model. Each clade was labelled by the previously assigned Acc (*Accumilibacter*) clade name<sup>6,7</sup> and the primer name that was designed in this study. Bootstrap values are displayed as a percentage of 1000 replicates.

and Acc-ppk1-1170r, which targets *Accumilibacter* Type I, and has therefore been described as unclassified<sup>8,10</sup>. Clade IB has been found to be the dominant *Accumilibacter* clade in an EBPR reactor, and a draft genome has been reconstructed<sup>10</sup>. The second reason was that the other dominant clades were not covered by the previous qPCR assay, which required new primer sets for identification and quantification. Thus, the primer design for these clades is required to study the entire profile of *Accumilibacter* clades in different environments.

The *Accumilibacter ppk1* gene sequences in public databases have been expanded and rapidly updated in recent years, providing sufficient reference to design primer sets with good coverage and specificity. Therefore, this study sought to develop a qPCR assay by primer design sets to comprehensively and specifically profile the distribution of *Accumilibacter* clades.

## Results

**Phylogenetic analysis of *ppk1* gene.** A database of 605 *ppk1* gene sequences of 14 known *Accumilibacter* clades (Supplementary Table S7) was retrieved from the NCBI database<sup>7–9,11</sup>. The maximum likelihood method with 1000 bootstrap replicates was used to construct a phylogenetic tree by MEGA (v. 6.06)<sup>12</sup>. The condensed tree with a bootstrap support value cutoff of 50%, is displayed to indicate the relationships among *ppk1* gene homologues (Fig. 1). The *ppk1* gene sequences were classified into five clades in Type I and nine clades in Type II, in accordance with the classification in previous studies<sup>5,7,8,11</sup>.

**Design of qPCR primer sets.** Primer design was conducted using PRISE2<sup>13</sup>; it focused on nine clades without available primers, including Clades IA, IB, IC, ID, and IE in Type I and Clades IIE, IIG, IIH and II-I in Type II. The amplicon length, annealing temperature and inter/intra complimentary between forward primers and reverse primers were considered.

During primer design, sequences from the database were successively selected to represent different target clades, whereas all other sequences were selected to represent the non-target clades. The final primer sets were selected after considering the following criteria: (i) Both forward and reverse primers exactly match the majority of the target sequences and none of the non-target sequences, under very stringent settings (indicates high coverage and high specificity assuming no mismatch during the qPCR). (ii) Both forward and reverse primers match as few non-target sequences as possible, under very flexible settings (indicates high specificity considering potential mismatches during qPCR). (iii) Under flexible settings, the mismatches between the candidate primers and non-target sequences should be as far away from the 5' end of the candidate primers as possible. The percentage of target sequences for both primers was calculated using the most stringent setting—OMAM (mismatch allowance mechanism<sup>13</sup>) and the percentage of non-target sequences was calculated using all MAM settings (Supplementary Materials and Methods, Supplementary Table S4). As an example, the comparison between 3MAM and 1MAM, which showed 4.28% non-target sequences and 1.43% non-target sequences, respectively, aligned to Primer IB, indicated that the unspecific amplification of the 16 sequences of Clades ID and IE were more likely to be inhibited by distinctive 3' end dissimilarity (Supplementary Table S1). However, Primer IB could not discriminate the other eight sequences of Clade IC at the terminal 3' position, which were reflected by the comparison between 1MAM and 1MAM\*. In comparison with other *in silico*-based primer design strategies, in our strategy, positional mismatches towards non-target sequences were considered to enhance the discriminatory power of the primer 3' end. Thus, the specificity of the candidate primers was enhanced by introducing multiple MAM patterns to *in silico* evaluation, especially when many candidates satisfied the first two criteria. Criterion (iii) excluded candidate primers with mismatches at less discriminatory positions (central region or the primer 5' end), which would certainly spare many efforts in *in vitro* examination. After *in silico* screening, two candidate primer sets were

selected for each clade; the percentage of target sequences and the percentage of non-target sequences were 100% and 0%, respectively, when a perfect match was required during qPCR. According to these criteria, no primer set that targets the entire Type I or the entire Type II could be designed. In addition, the primer set Acc-ppk1-763f and Acc-ppk1-1170r, which targets *Accumulibacter* Type I<sup>6</sup> partially covered the *ppk1* gene sequences that belong to Clade IA and one sequence of Clade IB (accession number EU432881) (Supplementary Table S6).

**Evaluation of the novel qPCR primer sets.** Three mixtures of DNA samples from diverse sludge sources (18 WWTPs and an EBPR lab-scale reactor) were cocktailled as a DNA template to conduct PCR and qPCR using the candidate primer sets. Due to the rare appearance of Clade IE in the DNA template, primer sets that target Clade IE were not evaluated in this study. Amplicons of each primer set were cloned and sequenced for phylogenetic analysis. According to the assignment results, primer sets with amplicons that were affiliated only inside their target clades were selected for the efficiency evaluation. Two candidate primer sets for Clades IA, ID and IIE and one candidate primer set for Clades IC, IIF, IIG, IIH and II-I matched their theoretical specificities. Although the application of MAM patterns can increase the possibility of achieving specificity-qualified candidate primers, i.e., 11 of the 18 designed primers in this study, sequencing-based cross-amplification evaluation is necessary to examine the cross-amplification that is not completely predicted by *in silico* checking.

For Clade IB, the phylogenetic tree indicated that the qPCR products from two candidate primer sets were clustered within Clades IB, IC and ID. One primer set with relatively higher specificity was selected, and the qPCR condition was optimized by increasing annealing temperature. The qPCR products of two elevated annealing temperatures (57.0 °C and 58.0 °C) exhibited bright bands on agarose gel (Supplementary Fig. S9) and were cloned and sequenced. The consensus trees (Supplementary Fig. S2) indicated that the slight increase in annealing temperature significantly improved the specificity without a loss of coverage, on the basis of the diversity of two achieved clone libraries. To test the potential cross-amplification, the annealing temperature of qPCR using Primer-IB was set at 57.0 °C during future experiments.

The representative sequences obtained using the designed primer sets displayed high intra-clade dissimilarity, with identities to reference sequences from 93% to 100%. This result demonstrated the diversity and divergence within each clade in the DNA template. Although the true coverage of the *Accumulibacter* clades was not determined, the diversity of the sequences rendered confidence in the application of these primers to an extensive range of target sequences in the environment. This diversity also supported the specificity of the primers because numerous and diverse non-target sequences from all known clades, with the exception of Clade IE, were examined (Supplementary Table S2).

**PCR efficiency and interference of the designed primer sets.** To evaluate the PCR efficiency, plasmids that contain amplicons of a specific clade were obtained using the candidate primers and previously designed primer sets<sup>6</sup>. The amplification effectiveness was evaluated by melting curves and six-point calibration curves in a ten-fold series. Six replicates for each plasmid copy number from 10<sup>3</sup> to 10<sup>8</sup> were applied within each assay. Only one primer set was selected according to its PCR efficiency. For all primer sets, a high PCR efficiency (85–112%) with a correlation coefficient in the range of 0.962–0.998 was achieved at the optimized annealing temperature and primer concentration (Table 1) to ensure the accuracy during quantification.

To explore the potential interference among different amplicons, we individually conducted qPCR with each primer set, using plasmids that contain amplicons of 13 *Accumulibacter* clades generated with primers listed in Supplementary Table S1, as DNA templates with the copy number of 10<sup>7</sup>. In this paper, a concept of “relative efficiency” was introduced to quantify the interference effect, which was calculated using the Ct value (referred to as the threshold cycle) with the following formula assuming a 100% amplification efficiency.

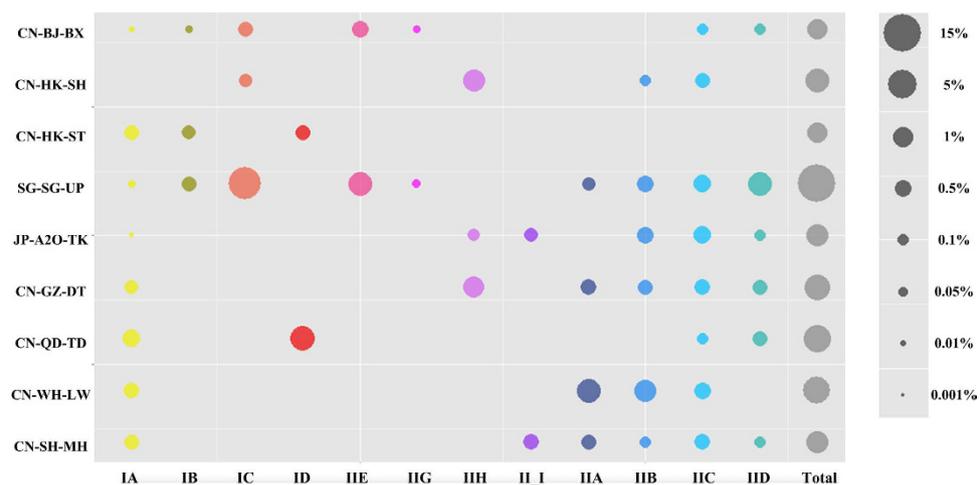
$$\text{Relative efficiency} = 2^{(\text{Ct value of a non-target amplicon} - \text{Ct value of the target amplicon})} \quad (1)$$

To define whether the existence of non-target amplicons might influence accurate quantification, we introduced an observed delay of less than ten cycles (referred to as the threshold cycle) as the criterion, which was equivalent to a relative efficiency of more than 0.1%<sup>14</sup>. As shown in Supplementary Table S3, 41.3% combinations had no interference (interference was not detected by qPCR, green nodes) and 57.7% had negligible interference (relative efficiency less than 2<sup>-10</sup>, yellow nodes), whereas only the combination of Primer-IC and Clade IIF exhibited slight interference (2<sup>-9.95</sup>, red node). The sum of nonspecific relative efficiencies for each primer set approximately indicated the total possibility of interference from non-target amplicons, which was significantly low—from 2<sup>-17.0</sup> to 2<sup>-8.5</sup>.

**Application of the newly designed qPCR primer sets to quantify *Accumulibacter* clades in WWTPs.** The occurrence of *Accumulibacter* clades in 18 full-scale WWTPs worldwide has previously been revealed<sup>8</sup> by the qPCR-based method, with primer sets designed by He *et al.*<sup>6</sup>. More than 50% of the total *Accumulibacter* in ten of the 18 WWTPs may belong to some clades not targeted by previously designed primer sets. In this study, the newly designed primer sets were used to formulate an assay investigating the unclassified *Accumulibacter* in nine WWTPs, including the WWTPs of mainland China (CN-BJ-BX: Bei-Xiao-He, Beijing; CN-GZ-DT: Da-Tan-Sha, Guangzhou; CN-QD-TD: Tuan-Dao, Qingdao; CN-WH-LW: Long-Wang-Zui, Wuhan and CN-SH-MH: Min-Hang, Shanghai), Hong Kong (CN-HK-ST: Sha-Tin and CN-HK-SH: Shek-Wu-Hui), Singapore (SG-SG-UP: Ulu Pandan), and Japan (JP-A2O-TK: Tsukuba, WWTP using anaerobic/anoxic/aerobic (A/A/O) process). The abundance of one clade in the total bacterial population was calculated using the ratio of the cell number of *Accumulibacter* (assuming one *ppk1* gene per cell) to the total bacteria cell number (based on 16S rRNA gene copy number). Both the *ppk1* and 16S rRNA gene copy numbers were obtained by qPCR<sup>6,8</sup>. The average *rrn* (rRNA) operon copy number per cell for each sample, except JP-A2O-TK, was calculated based on a

Primer <sup>a</sup>	Sequence (5'-3')	Target <sup>b</sup>	Amplicon size (bp)	Coverage <sup>c</sup>		Specificity ( <i>in silico</i> ) <sup>c</sup>		qPCR conditions		qPCR performance (n = 6)	
				Percentage of target sequences (0MAM <sup>d</sup> )	Percentage of non-target sequences (0MAM)	Primer concentration (nM)	T <sup>a</sup> (°C) <sup>e</sup>	Standard curve correlation coefficient	PCR efficiency <sup>f</sup> (%)		
Primer-IA	Acc-ppk1-974f	Clade IA	140	100%	0.0%	400	62.1	0.994 ± 0.002	108.1 ± 4.2		
	Acc-ppk1-1113r									AATGATCGGATTGAAGCTCTGGTAG	
Primer-IB*	Acc-ppk1-372f	Clade IB	282	100%	0.0%	400	58.0	0.982 ± 0.011	112.2 ± 8.7		
	Acc-ppk1-653r									AAGCAGTATTCGCTGTC	
Primer-IC	Acc-ppk1-362f	Clade IC	397	100%	0.0%	350	65.5	0.986 ± 0.013	100.9 ± 4.3		
	Acc-ppk1-758r									AACAGGTTGCTGTTGCGCGTGA	
Primer-ID	Acc-ppk1-634f	Clade ID	215	100%	0.0%	400	58.1	0.984 ± 0.012	84.7 ± 7.9		
	Acc-ppk1-848r									ACTTCGAGGCGGACG	
Primer-IIE	Acc-ppk1-757f	Clade IIE	373	100%	0.0%	400	57.8	0.967 ± 0.016	100.0 ± 8.8		
	Acc-ppk1-1129r									ATTGTTTCGAGCAACTCGATG	
Primer-IIG	Acc-ppk1-410f	Clade IIG	105	100%	0.0%	400	62.1	0.979 ± 0.026	94.9 ± 8.8		
	Acc-ppk1-514r									TGTTGAGTACGCGCGGGA	
Primer-IIH	Acc-ppk1-701f	Clade IIH	228	100%	0.0%	400	57.5	0.980 ± 0.007	93.2 ± 8.1		
	Acc-ppk1-928r									TCATCGCTTCGGAGCA	
Primer-II-I	Acc-ppk1-688f	Clade II-I	259	100%	0.0%	400	58.5	0.993 ± 0.004	102.2 ± 4.5		
	Acc-ppk1-946r									TGAAGTGTCCGAGCAGGA	

**Table 1. Information, qPCR conditions and performance of primers designed in this study.** <sup>a</sup>Acc, Accumulibacter. <sup>b</sup>Clade, Accumulibacter Clade. <sup>c</sup>The percentage of target and non-target sequences was calculated against 605 available *ppk1* gene sequences of 14 known Accumulibacter clades. <sup>d</sup>0MAM, no mismatch allowed, which indicated that a perfect match was required during the qPCR. <sup>e</sup>T<sup>a</sup>, annealing temperature. <sup>f</sup>PCR efficiency was estimated from the slope of the standard curve by equation (2). \*During the evaluation, the annealing temperature of Primer-IB was set to 57°C for potential cross-amplification detection. The standard curve of each primer set was examined by six-replicate tests.



**Figure 2. Estimated abundance of Accumulibacter clades relative to the total bacterial community in nine WWTPs.** The size of the nodes represents the abundance of the indicated clade, as estimated by the Accumulibacter *ppk1* gene copy number to the bacterial 16S rRNA gene copy number<sup>6,7</sup>. The scale bar at the right side provides the reference value for the abundance.

reanalysis of their bacterial community structure by the QIIME pipeline<sup>15</sup> (1.70v), using our previous data<sup>16</sup> via Copyrighter-0.46<sup>17</sup>. For JP-A2O-TK, the *rrn* operon copy number per cell was represented by the average (2.04 copies per cell) of all samples.

The profile of different Accumulibacter clades in nine samples revealed that Accumulibacter existed in the SG-SG-UP with the highest abundance of 14.6% in the total bacterial community (Fig. 2). Among the 12 Accumulibacter clades, Clades IA, IIB and IIC appeared to be universally present and well adaptable, because they were detected in eight or nine geographically and operationally distinct WWTPs. The Accumulibacter lineage in SG-SG-UP and CN-BJ-BX exhibited a similar pattern and was dominated by Clades IC and IIE. Two

novel clades—IIH and II-I—were discovered from CN-GZ-DT and JP-A2O-TK<sup>8</sup>, accounting for 1.3% and 0.3%, respectively, of the total bacterial community.

## Discussion

In this study, ten sets of qPCR primers were designed to target the Accumulibacter Clades IA, IB, IC, ID, IE, IIE, IIG, IIH, and II-I, which had no clade-specific primers available. The primer design method in this study introduced positional MAM<sup>13</sup> settings to overcome the lack of valuing the position of mismatches in previous design strategies. The specificity and PCR efficiency of each candidate primer set were assessed and verified via PCR-cloning-sequencing, phylogenetic analysis, standard curves, and interference evaluation. Primer IE could not be evaluated due to limited sludge samples. Eight sets of primers (Table 1) were verified as highly selective and efficient. These primer sets exhibited 100% coverage of their target reference clades, 0% cross-amplification and high PCR efficiency (85–112%). The DNA template in this study consisted of spatially and temporally diverse sludge samples around the world, covering 13 known Accumulibacter clades and provided high confidence for the assessment results.

However, some limitations of this study remain. First, the primer set for Clade IE could not be assessed, owing to the rare occurrence of Clade IE in the samples. In addition, the coverage of the target clade and the diversity of the retrieved sequences were demonstrated only by *in silico* analysis as this diversity could not be validated via experiments. Only 20 clones were selected from the PCR products of each primer set for specificity evaluation. Finally, only three amplicons were selected from each clade for PCR efficiency evaluation; therefore, it may not have represented the diversity of samples in the environment.

With previously designed primer pairs<sup>6,9</sup>, we expanded the limited qPCR-based assay of the *ppk1* gene to a comprehensive assay to identify and quantify known Accumulibacter clades. The resolution of primers that target Type I was substantially improved by the individual quantification of Clades IA, IB, IC and ID. In addition, the profile of Accumulibacter Type II was comprehensively assessed, because all known clades were covered in the expanded assay. Therefore, this assay significantly resolves the problem in which a large portion of Accumulibacter clades could not be previously identified (Supplementary Table S5). However, the unclassified Accumulibacter abundance from 33.1% to 64.4% remains in 5 of 9 samples. The percentage of unidentified clades may have been from Clades IE and IIF, which were not quantified in this study, and Clade IIC, which was only partially covered by Primer IIC with 36.4% coverage (Supplementary Table S1) and the unknown Accumulibacter clades.

With more data provided by this new assay, potential directions for subsequent studies can be proposed: draft genome retrieving, gene expression dynamics and metabolic pathway analysis of dominant clades, especially for samples that were not previously reported<sup>18</sup>. For example, the results in this study demonstrated the high abundance of Clade IIE, i.e., 2.2% and 0.5% of the total bacterial community in SG-SG-UP samples and CN-BJ-BX samples, respectively, which indicates the significance and possibility of metagenomic analysis and genome binning of Clade IIE, which remains unknown.

## Materials and Methods

**Sludge samples collection.** Three groups of sludge samples were collected as previously described<sup>8,10,18</sup>.

- (i) Temporal activated sludge (AS) samples were collected monthly from the middle of an aerobic (oxic) tank at the Shatin WWTP in Hong Kong from 2013 to 2014 (22°23'N 114°11'E). The tank, which was designed with a conventional anoxic/oxic (A/O) treatment process<sup>18</sup>, contained Accumulibacter Clades IB, IC and ID (Supplementary Table S2). The samples with a high abundance of Accumulibacter (>0.2% of total bacterial community) were selected based on our previous 16S rRNA gene analysis<sup>18</sup>.
- (ii) AS samples were collected from 18 globally distributed and full-scale WWTPs in Asia (Hong Kong, mainland China, Japan and Singapore), North America (Canada and the United States) and Europe (the United Kingdom)<sup>8,16</sup>, which contained Accumulibacter Clades IA, IB, IC, ID, IIE, IIG, IIH and II-I (Supplementary Table S2).
- (iii) PAO-enriched AS samples were collected from a well-performing EBPR batch reactor that was fed with acetate as the carbon source<sup>10</sup> and contained Accumulibacter Clades IA, IB, IC, ID, IIE, IIG and IIH (Supplementary Table S2).

All samples were fixed by 50% ethanol at sampling and stored under  $-20^{\circ}\text{C}$  prior to DNA extraction.

**DNA extraction.** Total DNA was extracted from 2.0 mL of the sludge samples using a FastDNA<sup>®</sup> SPIN Kit for Soil (MP Biomedicals, Solon, OH). DNA purity at 260 nm and 280 nm was verified by a NanoDrop<sup>®</sup> Spectrophotometer ND-100 (Thermo Fisher Scientific, USA), and the concentrations of the DNA samples were measured by a Qubit<sup>®</sup> 2.0 Fluorometer (Invitrogen, Life Technology, USA).

The DNA samples were diluted to the same concentration and were completely mixed for each group. Three groups of mixed samples represented the AS of one full-scale WWTP in different operational conditions across a broad range of time, the AS of full-scale WWTPs from an extensive range of geographical locations, and the AS of the EBPR laboratory reactor. These samples provided high diversity of Accumulibacter lineage for 13 target clades.

**The *ppk1* gene database construction and phylogenetic analysis.** A *ppk1* gene database was constructed with the reference sequences<sup>7–9,11</sup>, which were downloaded from the NCBI database. A total of 605 *ppk1* gene sequences from 14 known Accumulibacter clades were included. All sequences were aligned, and a phylogenetic tree was constructed using the maximum likelihood method by MEGA (v. 6.06)<sup>12</sup>. Bootstraps were

calculated for 1000 replicates. The condensed tree was generated by establishing a cutoff bootstrap support value of 50% (Fig. 1).

**Primer design.** Primer sets for qPCR were designed by PRISE2<sup>13</sup> for Clades IA, IB, IC, ID and IE of Type I and for Clades IIE, IIG, IIH and II-I of Type II. Amplicons larger than 450 bp and primers with degenerate bases were avoided. The annealing temperature difference and inter/intra complimentary between forward primers and reverse primers were strictly controlled to ensure efficient amplification. Four different patterns of MAM were used to select highly specific primers, which are specified in the Supplementary Materials and Methods. Two candidate primer sets with 100% coverage of target clades and 0% coverage of non-target clades were selected for evaluation. A PCR-cloning-sequencing test was performed to ensure the specificity of each primer set, with the exception of Clade IE.

**qPCR conditions and *ppk1* gene clone library construction.** Amplification of *Accumulibacter ppk1* gene fragments was conducted on a BioRad iCycler (v. 5.0, BioRad, Hercules, CA) in a 25  $\mu$ L reaction volume. The qPCR was individually conducted on 5 ng mixed DNA samples as a template by primer sets for the *ppk1* gene (Table 1) and the primer set 341f/534r<sup>6</sup> for 16S rRNA gene using SYBR Premix Dimer Eraser<sup>TM</sup> (Takara, Japan). Thermal cycling and fluorescence detection were performed using the programme of initial denaturation at 95 °C for 30 s, followed by 40 cycles of denaturation at 94 °C for 5 s, annealing for 45 s, and extension at 72 °C for 30 s. The primer concentration and annealing temperature were adjusted to optimize PCR efficiency and specificity.

The amplicons of each group were purified by a quick-spin PCR Product Purification Kit (iNtRON, Korea) and equally mixed according to their concentrations. The mixed amplicons were verified by agarose gel and Bioanalyzer (2100, Agilent Technologies, Santa Clara, CA, USA). The amplicons with a single bright DNA band (or a single peak) at the expected size were selected for cloning using the pMD18-T vector (TaKaRa, Japan). Twenty clones were randomly selected from each library and sequenced to confirm the specificity of each primer set. The 20 sequences were searched against the *ppk1* gene database using BLASTN (v. 2.2.28+) and were aligned with the reference sequences to construct the phylogenetic tree (Supplementary Figs S1–S8). One primer set was selected for each clade considering the specificity and diversity within the target clade based on the phylogenetic results.

**Primer evaluation.** Six-point calibration curves and melting curves were used to evaluate the efficiency by ten-fold series dilution. Regarding the reproducibility, the intra-assay variability was measured by conducting each single run using six replicates of amplicons with copy numbers from 10<sup>3</sup> to 10<sup>8</sup>. The copy number of amplicons was calculated from the average molecular weight and the mass concentration<sup>20,21</sup>. A negative control without a DNA template was included to verify the possible contamination or primer-dimer formation in the PCR reactions.

The amplicons selected from each clade, with a copy number of 10<sup>7</sup>, were employed to evaluate the interference in the triplicate tests. For Clade IIC, two types of amplicons (IIC-1 and IIC-2) were obtained by the primer set that targets Acc-IIC *ppk1* and Acc-IIC *ppk1*, excluding OTU NS D3<sup>6</sup>.

**Application of the designed qPCR assay.** DNA samples from nine globally distributed WWTPs were selected to profile different clades of *Accumulibacter*. These samples have been previously published and contain more than 50% unidentified *Accumulibacter* due to the limited availability of qPCR primers. In the qPCR reaction, 5 ng of genomic DNA was used as a template. To verify the intra-assay variability, the interference evaluation of the amplicons and the DNA samples were conducted in triplicate tests. The averages and standard deviations were calculated.

**Data analysis.** The qPCR amplification efficiency was calculated based on the slope of the standard curve.

$$\text{PCR efficiency} = 10^{-1/\text{slope}} - 1 \quad (2)$$

The qPCR amplification was considered valid only when the standard curves had a correlation coefficient greater than 0.96 and a PCR efficiency within the range 85–115%.

The abundance of *Accumulibacter* among the total bacterial community was normalized against the average 16S rRNA gene copy number per cell of each sample, which was calculated based on the bacterial composition by Copyrighter-0.46<sup>17</sup>. The bacterial composition was extracted from our previous study<sup>16</sup>. For JP-A2O-TK, we used the average of other AS samples.

**Nucleotide sequence accession numbers.** The *ppk1* gene sequences obtained by Sanger sequencing have been deposited at GenBank under accession numbers of KT752523–KT752702. The accession numbers of amplicons applied in the PCR efficiency evaluation are as follows: KT752544–KT752546 (Clade IA), KT752630–KT752632 (Clade IB), KT752693–KT752695 (Clade IC), KT752606–KT752608 (Clade ID), KT752533–KT752535 (Clade IIE), KT752565–KT752567 (Clade IIG), KT752667–KT752669 (Clade IIH) and KT752590–KT752592 (Clade II-I). The accession numbers of amplicons applied in the interference assay are as follows: KT752544 (Clade IA), KT752630 (Clade IB), KT752693 (Clade IC), KT752606 (Clade ID), KP737891 (Clade IIA), KP738049 (IIB), KP737878 and KP738080 (IIC), KP738094 (IID), KT752533 (Clade IIE), KT752565 (Clade IIG), KT752667 (Clade IIH) and KT752590 (Clade II-I).

## References

- Oehmen, A. *et al.* Advances in enhanced biological phosphorus removal: from micro to macro scale. *Water Res.* **41**, 2271–300 (2007).
- Kong, Y., Nielsen, J. L. & Nielsen, P. H. Microautoradiographic study of *Rhodocyclus*-related polyphosphate-accumulating bacteria in full-scale enhanced biological phosphorus removal plants. *Appl. Environ. Microbiol.* **70**, 5383–90 (2004).
- Mielczarek, A. T., Nguyen, H. T., Nielsen, J. L. & Nielsen, P. H. Population dynamics of bacteria involved in enhanced biological phosphorus removal in Danish wastewater treatment plants. *Water Res.* **47**, 1529–44 (2013).
- Hesselmann, R. P. X., Werlen, C., Hahn, D., van der Meer, J. R. & Zehnder, A. J. B. Enrichment, Phylogenetic Analysis and Detection of a Bacterium That Performs Enhanced Biological Phosphate Removal in Activated Sludge. *Syst. and Appl. Microbiol.* **22**, 454–465 (1999).
- McMahon, K. D. *et al.* Polyphosphate kinase genes from full-scale activated sludge plants. *Appl. Microbiol. & Biotechnol.* **77**, 167–73 (2007).
- He, S., Gall, D. L. & McMahon, K. D. “*Candidatus Accumulibacter*” population structure in enhanced biological phosphorus removal sludges as revealed by polyphosphate kinase genes. *Appl. & Environ. Microbiol.* **73**, 5865–74 (2007).
- Peterson, S. B., Warnecke, F., Madejska, J., McMahon, K. D. & Hugenholtz, P. Environmental distribution and population biology of *Candidatus Accumulibacter*, a primary agent of biological phosphorus removal. *Environ. Microbiol.* **10**, 2692–703 (2008).
- Mao, Y., Graham, D. W., Tamaki, H. & Zhang, T. Dominant and novel clades of *Candidatus Accumulibacter phosphatis* in 18 globally distributed full-scale wastewater treatment plants. *Sci. Rep.* **5**, 11857 (2015).
- Ong, Y. H. *et al.* High-temperature EBPR process: the performance, analysis of PAOs and GAOs and the fine-scale population study of *Candidatus “Accumulibacter phosphatis”*. *Water Res.* **64**, 102–12 (2014).
- Mao, Y., Yu, K., Xia, Y., Chao, Y. & Zhang, T. Genome reconstruction and gene expression of “*Candidatus Accumulibacter phosphatis*” Clade IB performing biological phosphorus removal. *Environ. Sci. Technol.* **48**, 10363–71 (2014).
- Huang, P., Li, L., Kotay, S. M. & Goel, R. Carbon mass balance and microbial ecology in a laboratory scale reactor achieving simultaneous sludge reduction and nutrient removal. *Water Res.* **53**, 153–67 (2014).
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–9 (2013).
- Huang, Y. T., Yang, J.-i., Chrobak, M. & Borneman, J. PRISE2: Software for designing sequence-selective PCR primers and probes. *BMC bioinformatics.* **15**, 317 (2014).
- Wright, E. S. *et al.* Exploiting extension bias in polymerase chain reaction to improve primer specificity in ensembles of nearly identical DNA templates. *Environ. Microbiol.* **16**, 1354–65 (2014).
- Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods.* **7**, 335–6 (2010).
- Zhang, T., Shao, M. F. & Ye, L. 454 pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants. *ISME J.* **6**, 1137–47 (2012).
- Angly, F. E. *et al.* CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. *Microbiome.* **2**, 1–13 (2014).
- Ju, F. & Zhang, T. Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. *ISME J.* **9**, 683–95 (2015).
- Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics.* **10**, 421 (2009).
- Whelan, J. A., Russell, N. B. & Whelan, M. A. A method for the absolute quantification of cDNA using real-time PCR. *J. of Immunol. Methods.* **278**, 261–269 (2003).
- Kim, J., Lim, J. & Lee, C. Quantitative real-time PCR approaches for microbial community studies in wastewater treatment systems: applications and considerations. *Biotechnol. Adv.* **31**, 1358–73 (2013).

## Acknowledgements

Zhang A. N. acknowledges the University of Hong Kong for the postgraduate studentship. Dr. Mao Y. acknowledges the University of Hong Kong for the postdoctoral fellowship. The authors are thankful for the financial support received by the Hong Kong General Research Fund (HKU 172057/15E).

## Author Contributions

A.N.Z. conducted the experiments, analysed the data, and wrote the manuscript. Y.M. assisted in sludge sample collection and provided important suggestions in data analysis and manuscript preparation. T.Z. designed the experiments and edited this manuscript. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Zhang, A. N. *et al.* Development of Quantitative Real-time PCR Assays for Different Clades of “*Candidatus Accumulibacter*”. *Sci. Rep.* **6**, 23993; doi: 10.1038/srep23993 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>