Inland Norway University of Applied Sciences

Faculty of Applied Ecology, Agricultural Sciences, and Biotechnology

Nida Liaqat

Master's Thesis

Gene expression profiles of phosphorus-accumulating ppk1 gene in an EBPR system

Master's in applied and Commercial Biotechnology

2024

Consent to lending by University College Library	**YES ☐**	**NO ☒**
Consent to accessibility in digital archive Brage	**YES ☐**	**NO ☒**

Acknowledgment

I would like to express my deepest gratitude to my supervisor, Professor Wenche jonhansen, for giving me the opportunity to work on a project that genuinely interests me. Her expert guidance and valuable suggestions have been instrumental in my learning journey, and working with her has been an immense pleasure.

I am also thankful to Wenche kristiansen for their exceptional assistance during my research work. Her technical expertise significantly helped me refine the objectives of this project, and she provided a comfortable and encouraging environment throughout my work.

Special thanks go to my all sisters, especially Ayesha, and my brother, Shoaib, Saqib, and my husband, Usama Aamir for their unwavering support and appreciation. Their help has been crucial in completing this project.

I also want to acknowledge my loving parents. Their selfless sacrifices, tremendous efforts, and ceaseless prayers have enabled me to pursue and complete my master's thesis.

Lastly, I am grateful to all my friends and supportive teachers who have helped me at every step of this journey. Their encouragement and assistance have been deeply appreciated.

Nida Liaqat

Hamar, 2024

Abbreviations

EBPR	Enhanced biological phosphorous removal
BLAST	Basic local alignment search tool
CD	Coding sequence
cDNA	Complementary Deoxy ribonucleic acid
ATP, ADP, or AMP	Adenosine tri-, di-, or mono-phosphate
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	*Escherichia coli*
gDNA	Genomic DNA
LB	kanamycin-– Luria-Bertani medium comprising Kanamycin antibiotic
mRNA	Messenger RNA
NCBI	National Centre for Biotechnology Information
PCR	Polymerase chain reaction
SBR	Sequencing Batch Reactor
SP	Soluble phosphorus
PP	Particulate phosphorus
USEPA	The United States environmental protection agency
BNR	Biological Nutrient Removal
AS	Activated sludge
MBBR	Moving bed biofilm-reactor
PPGK	Polyphosphate glucokinase

PKK	Polyphosphate Kinase
EDTA	Ethylene diaminetetra acetic acid
PAOs	Polyphosphate accumulating organisms
PAH	Poly-β-hydroxy-alkanoats
WWTP	Wastewater Treatment Plant
Tg	Tera-gram
EC	Electrocoagulation
RT	Reverse transcriptase
VFAs	Volatile fatty acids

Table of Content

List of Tables

List of Figures

Abstract

The effluent from industries and urbanization contains phosphorus, which pollutes water bodies and depletes phosphorus sources. To address this problem, the Hias method is considered an advanced and effective technique for phosphorus recovery and removal, combining Enhanced Biological Phosphorus Removal (EBPR) and Moving Bed Biofilm Reactor (MBBR) systems. This study aims to determine the diversity of polyphosphate-accumulating organisms (PAO) *Candidatus Accumulibacter* in the Hias biofilms using phosphate kinase (*ppk1*) gene as a genetic marker. Eight primer sets targeting the *ppk1* gene were designed for the PCR amplification of samples taken from Hias biofilm, with the optimum temperature determined using gradient PCR. Pfam and phylogenetic analysis were performed to examine the conserved domains and evolutionary relationships of eight *Ca.* Accumulibacter sequences obtained through metagenomic studies. TOPO-TA cloning and sanger sequencing were conducted for the identification and validation amplified products. Finally, RT-qPCR was performed to determine the *ppk1* expression of *Accumulibacter* species involved in EBPR using the expression level of the *ppk1* gene across different zones of the bioreactor. The results indicate that all analyzed *Ca.* Accumulibacter sequences contain similar domain architecture of polyphosphate kinase domains, with their phylogeny revealing the existence of four distinct clade with high confidence. Sanger sequencing of *ppk1* gene fragments from two cloning reactions resulted in high-quality sequences, showing over 85% identity with the Ca. Accumulibacter Phosphatis sequence validating the presence of species abundantly in zone 4. RT-qPCR analysis revealed fluctuating Ct values across reactor zones, indicating upregulation of the *ppk1* gene in anaerobic and downregulation in aerobic zone. This study contributes to the effective phosphorus removal by analysing the complex diversity of PAO Ca. Accumulibacter Phosphatis and phosphorus removal efficiency by *ppk1* gene. It is concluded that, Ca. Accumulibacter Phosphatis was the largely abundant P accumulation bacterial group in Hias process. This explain the intricate diversity of Hias samples and necessitates further analysis of *ppk1* transcript dynamic across other PAO strains and biofilm samples.

Key words: Moving-bed biofilm-reactor MBBR, Enhanced biological phosphorous removal (EBPR), Hias process, polyphosphate-accumulating organisms, *Candidatus Accumulibacter*, *ppk1* gene and RT-qPCR

1. Introduction

1.1. Phosphorous Sources and Eutrophication

Freshwater is a critical reserve for the survival of life on Earth. However, unprecedented threats caused by anthropogenic activities are threatening global freshwater ecosystems. A grave risk to aquatic environments has arisen because of the growing industry, development of agriculture, and augmented production of wastewater in recent decades. Plants growth depend on important nutrients like phosphorus (P) and nitrogen (N), and their concentrations in water influence the trophic status of marine biomes (Malmqvist & Rundle, 2002; Strayer & Dudgeon, 2010).

In terms of nourishment, P is considered the most essential nutrient. In soil, the quantity of dissolved P can be controlled to influence the growth of plants (Malmqvist & Rundle, 2002; Strayer & Dudgeon, 2010). The P contamination is caused by human industrial wastes, domestic sewage, predominantly food-handling sewages (Barnes et al., 1985), and phosphorus-rich fertilized terrestrial overflow (Bhagowati & Ahamad, 2019; Kauppi, 1985).

Wastewater contains higher concentration of P more than 0.6 Tg (Teragram) causes ecological and agricultural issues, signifying the need of P conversion into crop-friendly manures (Schoumans et al., 2015). Consequently, advancements in wastewater treatment developments are gradually promoting resource recovery. eutrophication ensues in lakes due to high concentrations of phosphorus, (Schindler et al., 2013). This is the main environmental issue today, leading to the failure to maintain a 0.003 and 0.8 µg/L range of phosphorus concentration (Smith & Schindler, 2009). In eutrophication, ample algal blooms, typically consist of purple-green algae (*Cyanophycae*) grows (Tiwari et al., 2017). This results in dissolved oxygen being depleted from the hypolimnion, while organic compounds and nutrients become more concentrated in the water, further exacerbating the process of eutrophication (Vollenweider, 1975). Such toxic water can also poison cattle after drinking it (Badger & Thomason, 2020). A study presents that, main factors responsible for accelerating eutrophication processes in waterbodies is municipal wastewater, which contains huge amount of N and P compounds (Bhagowati & Ahamad, 2019). Numerous countries have presented data related to animal health and Human health that are being affected by eutrophication caused by cyanobacteria and toxic microalgae (Granéli et al., 2008). The United States Environmental Protection Agency (USEPA) advises limiting phosphorus amounts to control eutrophication. It suggests that a stream that enters a water reservoir must not exceed 0.05 mg P/L to control P levels. USEPA

also recommended that water released into reservoirs must not exceed 0.1 mg/L for P eutrophication control measures (Wei et al., 2022).

1.2. Significance of Phosphate

Phosphorous is a substantial element of life as it stores genetic information and transfers energy intracellularly during metabolic processes. P is a constituent of cell membrane phospholipid molecules, as well as nucleic acids and adenosine mono- (ADP, ATP, or AMP), di- or triphosphate (Elser, 2012). Phosphorous has no substitute and holds great significance (Desmidt et al., 2015). Albino discovered P in 1688 in plants (Corbridge, 2016). The most vital source of P is mined phosphate rock, but the limited supply of P poses a great threat for the future. Archaeologically, phosphate rock has been over-used for the fertilizer business, leading to a surge in pressure on current reserves that may no longer be accessible within few years (Cordell et al., 2009). Production of phosphorus is only possible through mining, and it takes several million years to complete the natural inorganic cycle including tectonic alteration, erosion, and sedimentation. As early as 2033, phosphorous resources are expected to be depleted from mined phosphate rock, according to UK Soil Association statement (Soil Association, 2010) (Cornel & Schaum, 2009). It is anticipated that phosphorus will become more expensive and rare in the coming few years, potentially threating food security. Consequently, conservation of phosphorous resources is extremely essential for the benefits of human beings through the anthropogenic organic phosphorous cycle (Cordell & White, 2013).

1.3. Methods of removing phosphorous from wastewater

1.3.1. Chemical Methods

Wastewater treatment approaches categorize P into two forms: soluble phosphorus (SP) and particulate phosphorus (PP). SP is further categorized into two forms: soluble polyphosphate (SOPP) and soluble organic phosphate (SOP). Understanding the difference between these two forms is critical for developing wastewater treatment methodologies for P removal (Li & Brett, 2015). Generally, criteria for P removal are developed based on the total phosphorus (TP) amount. The discharge limit of municipal WWTP is recommended to be 1.0 mg/L of TP (Ge et al., 2018). Norwegian local codes and municipalities allow a waste P release limit of 1.0 mg/L

(Heistad et al., 2006). Furthermore, Norway has implemented strict local strategies and regulations regarding wastewater quality (Paruch et al., 2016).

Wastewater treatment employs various methods including biological, physical, and chemical methods (Zhu et al., 1996). P removal in wastewater treatment via chemical precipitation consists of four steps: precipitation, coagulation, flocculation and parting. This process involves the introduction of metal salts like iron or aluminum, which act as coagulants capable of inducing P precipitation. During coagulation, particles of the precipitated metal mix with P, leading to flocculation. Large heavy flocs form after the addition of coagulants, which can be easily separated from the liquid phase (Ødegaard, 2016). In chemical methods, various coagulants are used e.g., aluminum polychloride (Aln(OH)mCl(3n-m), iron chloride ($FeCl_2$), iron chloride ($FeCl_3$), iron sulphate ($Fe_2(SO_4)_3$), sodium aluminate ($NaAl_2O$), iron sulphate ($FeSO_4$) and aluminum sulphate ($Al_2(SO4)_3$) (Aboulhassan et al., 2006). The selection of P removal technique must be the most cost-effective method. There are two key approaches use for P retrieval in WWTPs. First method is created based on incinerated sludge sewage, while the second procedure is based on crystallization (e.g. struvite creation). Crystallization-based approaches are preferred for the elimination of P solely through biological processes. Incinerated sludge-based methodologies are used when employing chemical-based methods (Korving et al., 2019).

Another chemical method is the electrocoagulation (EC) process which is based on the principles of oxidation/reduction, flotation, and coagulation. In the EC method, charged ions can move between the submerged anode and cathode in an electrolyte solution when the current is applied. Through oxidation, the anode undergoes electrochemical dissolution generating metal ions *in situ*. The rate of cation production can be regulated by changing the current density. Various mono- and polymeric species are produced due to the hydrolysis of these cations, ultimately forming amorphous metal (hydroxides). The coagulants generated are highly compatible with counter ions and dispersed particles, facilitating the removal of contaminants such as P. Furthermore, the bubbles of $H_2(g)$ generated by the cathode through the process of water reduction can aid in capturing suspended materials and eliminating macro flocs that resist gravity (Hu et al., 2023).

Chemical methods effectively remove P, with most approaches involving precipitation, sorption and ion exchange mechanisms (Cornel & Schaum, 2009; Yuan et al., 2015). Solids separation procedures such as filtration and clarification remove these solid precipitates. For instance, iron removes P as iron ions are first oxidized into form of iron, which rely on factors

such as catalytic activity, oxygen level, pH, and the presence of inhibitory elements like sulphur. The reaction requires high concentrations of oxygen (Fe_2 +/0-15 g O2) (Scott, 1978). Upon contact with pyrophosphate and tripolyphosphate, iron ions form strong complexes that are subsequently removed by adsorption in form of iron hydroxo-phosphate (Jenkins et al., 1971). There are factors such as pH of iron, temperature, and phosphate concentration all influence the precipitation of oxidative products, thus affecting the efficient elimination of phosphorus (Nir et al., 2009). Another coagulant is aluminum that removes P in high amount. Arvin & Petersen (1980) established that aluminum precipitation relies on factors such as the activity of calcium, water's bicarbonate activity, pH, and the metal to phosphate concentration. Phosphate precipitation takes place when calcium is combined in the form of lime, as $Ca(OH)_2$. Calcium carbonate is formed when additional calcium ions react with natural alkalinity in wastewater. As the pH rises above 10, the excess calcium reacts with P, forming hydroxylapatite (Arvin & Jenkins, 1985).

Liquid ion exchange methodology is another effective method for removing P, in which impure acidified coagulants are mixed with organic soluble extract. The metal ions then adsorb to this liquid extract, forming organic-metal complexes. A hydrophobic organic solvent such as kerosene, is used to dissolve the extract when saturated with metal and remove it from the acidified sludge. Following the mixing process of the organic solution with an acid solution, the extract and coagulant are recovered from the solution (Al-Enezi et al., 2004).

Another useful chemical method involves the utilization of hydrogels, which offer a various technique for removing P from wastewater treatment. The principle of hydrogels in P removal depends on various factors such as surface charges, network structure, and hydrophilic nature. The inherent characteristics and gelatin chemistry of hydrogels enable selective absorption of substances, facilitating both P removal and recovery while enabling wastewater purification (Hu et al., 2023). Over time various types of hydrogels have emerged as evident from the literature. For instance, a self-cross-linked hydrogel exhibits exceptional swelling capacity, enabling the recovery of 10.72 mg/g of P from sewage (Dai et al., 2020). A poly (allylamine)-based hydrogel developed by Kofinas and Kioussis (2003) removes 98% of P from aquaculture wastewater in less than 2 hours. This hydrogel demonstrates a versatile nature across diverse wastewater compositions (Kofinas & Kioussis, 2003). Mahmoud et al. (2019) developed a gelatin and chitosan-based binary polymer network hydrogel, which works best in environments with high P concentrations, recovering 30 mg/g more of P compared to self-cross-linked hydrogels (Mahmoud et al., 2019).

1.4. Side Effects of Using Chemical Methods

Although chemical methods of phosphate removal from wastewater are efficient, these procedures still have certain limitations. During wastewater treatment, P is precipitated by the addition of salt, resulting in the removal of solids either by filtration or settling under gravity. However, elimination of chemically bonded phosphorus can be problematic, making efficient recovery of P impossible for further use (Yuan et al., 2015). Presence of high metal content can lead to environmental consequences, thus limiting the reuse of chemical sludge (Driver et al., 1999).

Liu & Zhou (2022) utilized chemical coagulants, resulting in 68 % elimination of P from glycine wastewater. However, the presence of residual metal ions in excess in generated sludge by these coagulants have been associated with adverse health effects and increased operating costs (Liu & Zhou, 2022). Researchers have revealed that aluminum ions play a central role in the brain's cholinergic signaling system, alter the signaling pathways of phosphoinositide, and bind to beta-amyloid, possibly leading to Alzheimer's syndrome in humans (Gauthier et al., 2000).

1.5. Biological Methods

Bacteria needed in biological methods for P removal, they possess the ability to absorb and assimilate P from the environment more than their growth requirements. Polyphosphate-accumulating organisms (PAOs) are the most suitable bacteria for this role as they can store P in polyphosphate (PolyP) form. Their P accumulation and conversion to PolyP involve two phases, aerobic and anaerobic. During the aerobic phase, PolyP is synthesized, while in the anaerobic phase, phosphorus is released from PolyP hydrolysis to generate energy (Fernando et al. 2019).

1.5.1. Enhanced Biological Phosphorus Removal (EPBR)

Biological P elimination, particularly Enhanced Biological Phosphorus Removal (EPBR), is a cost-efficient and economically viable alternative to chemical management. Compared to chemical precipitation, EBPR offers advantages such as the elimination of chemicals and chemical sludge during the precipitation process. In EBPR, after separation from the liquid phase, excess bio-P sludge with high P content (bio-P sludge) is produced. This process is

optimal for P recovery due to the absence of environmental toxic chemicals and the higher bioavailability of sludge compared to chemical sludge treatment methodologies (Zaletova et al., 2018).

1.5.2. Activated Sludge Process (ASP)

In Biological Nutrient Removal (BNR), activated sludge (AS), also known as suspended solids culture, has been shown to be effective for removing N and P. Sludge is recirculated between reactors in various settings. AS-plants have the advantage of having each process within its own reactor, making it relatively easy to manage (Zhang et al., 2017).

Due to the recirculation of sludge, BNR operates as a continuous system. In this process, settled sludge is pumped into the anaerobic region, where phosphate accumulating organisms (PAOs) can accumulate organic substrate and release P under advantageous conditions. After transporting the liquid bulk into the aerobic zone, PAOs utilize the stored poly-β-hydroxy-alkanoates (PHA) to absorb phosphorous and replenish their glycogen reserves. Compared to the moving bed biofilm reactor (MBBR), the load of suspended solids is ten times higher (Ødegaard, 2006), which requires massive space and extra cost for managing secondary removal of P. In the AS procedure, N and EBPR removal can simply be implemented by adding an anoxic chamber. Numerous full-scale procedures have been developed which helps in collective elimination of N and P with the help of AS methods with different configurations (Mannina et al., 2019).

1.5.3. Membrane Bio Reactor (MBR)

The MBR method is an improved version of conventional AS (CAS). This system enhances treatment efficiency and process control by integrating membranes, biological processes, and either ultra- or microfiltration. However, MBR processes face similar constraints to AS plants, in addition to the high cost of membrane fouling (Leyva-Díaz et al., 2017). The wash-out effects of PAOs may have an unfavorable outcome on the system's efficiency. As maintaining full recirculation of PAOs in typical ASP systems without any loss in the sludge is challenging (Sriwiriyarat & Randall, 2005).

1.5.4. Sequencing Batch Reactor (SBR)

The sequence batch reactor (SBR) is another wastewater treatment system that contains sequence of tanks. The fundamental principle of SBR is that every tank functions cyclically.

Each tank functions as a batch reactor and undergoes a process of filling, treatment, settling, and removal process. The treatment cycle begins when the tank is filled to a predetermined level at a specified time. The wastewater undergoes chemical and biological processes to remove impurities and purify the water during the treatment phase. After the treatment phase, pollutants are allowed to settle down at the bottom. The cleared supernatant sits on top of the settled solids after the settling process and removed from the tank. The tank turns on to start a new cycle once the supernatant has been removed. One advantage of the SBR system is the sequential nature of all treatment processes within a single tank, with contrasts with the traditional continuous-flow system. This system utilizes several tanks for different phases of treatment, and the wastewater continuously passes through these tanks necessarily. The SBR system operates on a time-based methodology that provides operational flexibility. The size and capacity of a system depend on the total time required for a cycle to complete. SBS systems are designed to effectively manage varying loads and flow rates by adjusting tank volume and cycle time (Irvine et al., 1989).

1.5.5. Moving Bed Biofilm Reactor (MBBR)

Biofilm reactors aid in removing nutrients such as P and N, along with organic matter. Before the advent of moving bed biofilm reactors (MBBRs), old technologies including fluidized-bed reactors, biological trickling filters, rotating biological contactors, and aerated submerged fixed-film biofilm reactors were used, each with their own pros and cons (Rittmann et al., 1982). Updated EBPR technology is a valuable upgrade of conventional activated sludge (CAS) plants developed with modification of existing technologies (Salvetti et al., 2006). As a hybrid between activated sludge and biofiltration, the MBBR incorporates some beneficial features of both (Aguado et al., 2006).

The biofilm method function someway similarly to a sequencing batch reactor (SBR) (Helness & Ødegaard, 2001). In EBPR, PAO consumes degradable organic matter, known as volatile fatty acids (VFAs), and under anaerobic circumstances release PO_4. Microbes accumulate and store PO_4 in aerobic zones and utilize these internal reserves of energy for cell growth and expansion. This ability of PAOs can be exploited by providing alternating anaerobic and aerobic environments, which allows them to produce and store more PO_4 compared to other microbes. Under anoxic circumstances, some PAO bacteria receive electrons from nitrate rather than oxygen (Zhou et al., 2021).

1.5.6. Hias Method for Sludge Treatment

The Hias process is an innovative wastewater treatment technology that combines EBPR (Enhanced Biological Phosphorus Removal) and moving-bed biofilm-reactor (MBBR) developed at Hias IKS in Ottestad, Norway, using Kaldnes K3 biofilm carriers as bio-medium (Saltnes *et al.*, 2017; Villard et al. 2023). The technique can recover 50% of phosphorus in a reusable form. Biofilm is a complex heterogeneous micro ecosystem of microbial communities (Flemming et al., 2021). There are several stages through which biofilm formation works, starting with adsorption of nutrients and macromolecules, cell transportation, adhesion, and irreversible attachment. The ability of microorganisms to produce energy in reactor setting is dependent on variations in the availability of electron donors and/or acceptors (Falkowski et al., 2008). Biofilms are communities of microorganisms that contain over a thousand different bacterial species (Angell et al., 2016). Several factors contribute to shifts in gradients within biofilms, affecting the distribution of nutrients and other substances. These shifts depend on variables such as rates of solute diffusion, microbial activity, and the physical structure of the biofilms (Flemming et al., 2016).

The bioreactor is divided into 10 zones. The first three reactor zones are anaerobic, and the remaining seven zones are aerobic. This allows for the alternating conditions required by the PAOs, utilizing a belt conveyor. Similar to active sludge EBPR techniques, the biofilm-carriers in the tenth zone are carried back to the first reactor zone, removing the need to recycle sludge for inoculation. The tenth zone effluent is returned to a sedimentation tank to separate and collect the P. The Hias IKS in Ottestad facility has the capacity to handle wastewater from 200,000 person equivalents (PE). The Wastewater Treatment Plant (WWTP) is responsible for receiving municipal wastewater from four Norwegian municipalities, namely Hamar,stange Løten and Ringaker. The treated wastewater is eventually released into Norway's largest lake, Mjøsa. Over an extended period, the Hias process has demonstrated remarkable stability and consistently achieves a P removal efficiency exceeding 95% (Saltnes *et al.*, 2017).

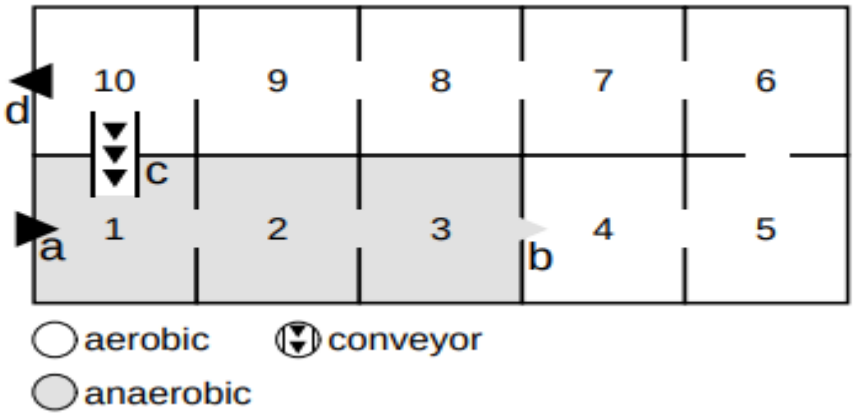


Figure 1.1. Graphic summary of Hias procedure. An inlet system installed for pre-treated water. (a), modified connected system in anaerobic zones for carrying biofilm carriers and water (grey) to an aerobic area (white) (b), a system for taking back biofilm carriers to the anaerobic area devoid of water (c) and for the treated water an outlet stream installed (d), (Revised from Rudi et al. 2019; HIAS IKS. 2015)

1.6. PAOs in Activated Sludge System

The process of biologically removing P from wastewater relies on the enrichment of PAOs within activated sludge. Within the activated sludge system, the dominant bacteria are aerobic heterotrophs responsible for breaking down and ultimately converting organic compounds found in wastewater into carbon dioxide. The overall heterotrophic population of bacteria remains stable, however, the dominant bacteria in terms of metabolic activity still vary depending on different environments or habitats within the three zones of plants (Seviour et al., 2003). EBPR systems operate at full-scale and in lab conditions, using *Ca. Accumulibacter* as a major genus of PAO (Samarasiri et al., 2019).

The most studied PAO full-scale activated sludge and in laboratory scale is "Ca. *Accumulibacter".* Based on *Accumulibacter ppk1* gene, it has been phylogenetically separated into two type I and II sub lineages, individually comprising several clades. Up to now, many clades have been proposed, for example clades IIA to II-I and clades IA to IF (Wang et al., 2021). Another bacterium, *Microlunatus phosphovorus*, was initially isolated in 1995, and its whole genome sequence analyzed in 2010. This species is also capable of accumulating polyP. In EBPR systems, it accounts for 9.0% of PAOs and 2.7% of all microorganisms (Kawakoshi et al., 2012).

1.6.1. Bacterial Enzymes responsible in PolyP Metabolism

Polyphosphate kinase (PPK) enzyme is responsible for polyP synthesis in most organisms. ATP's terminal phosphate is transferred to the growing chain of polyP—a reaction catalyzed by the PPK enzyme. A few pathogens, as well as most microbial genomes analyzed to date, appear to have apparent polyphosphate kinase (PPK) homologs. Using molecular techniques, it is possible to retrieve PPK genes from uncultivated organisms because PPK sequences are highly conserved among these bacteria (Mitchell et al., 2002).

Under unfavorable conditions many bacteria accumulate polyP for growth. Several genes identified through genome analysis are involved in the polyP metabolism. These include *ppk1* and *ppk2*, encoding polyphosphate kinase (PPK). Another gene, *ppx2* hydrolyzes the polyP terminal residues to produce inorganic phosphate (Pi) and is encoded by the exopolyphosphatase (PPX). Additionally, there is the *ppgk* gene, which encodes the polyphosphate glucokinase (PPGK). This gene is responsible for phosphorylating glucose using polyP. The *pap* gene encodes for a polyphosphate: AMP phosphotransferase (PAP), an enzyme involved in synthesizing ADP using AMP and polyP. The polyP/ATP-dependent NAD kinase (PPNK) is encoded by the *ppnk* gene, which synthesizes $NADP^+$ using polyP and ATP (Zhong et al., 2018).

Another extracellular hydrolytic enzyme called phosphatase (PO4ase) catalyzes reactions that release Pi from organic-bound P. Phosphatase is the only enzyme that cycles phosphorus (P) in the environment. It exists in nature in the forms of alkaline phosphatase, acid phosphomonoesterase and phosphodiesterases. Phosphatases exhibit different reaction capabilities depending on their substrates. Phosphatases can be utilized as biochemical tests under anaerobic conditions to assess the stability of anaerobic digesters. These enzymes are involved in the removal of heavy metal contaminants (Anupama et al., 2008).

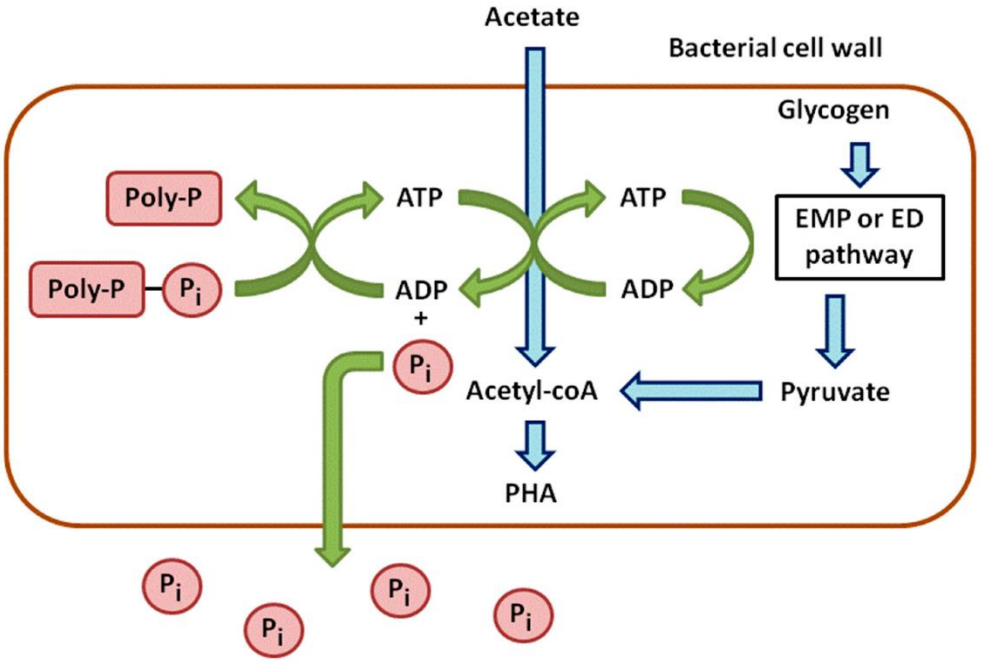


Figure 1.2. Aerobic metabolism of PAOs (adapted from & Dudgeon et a., 2010).

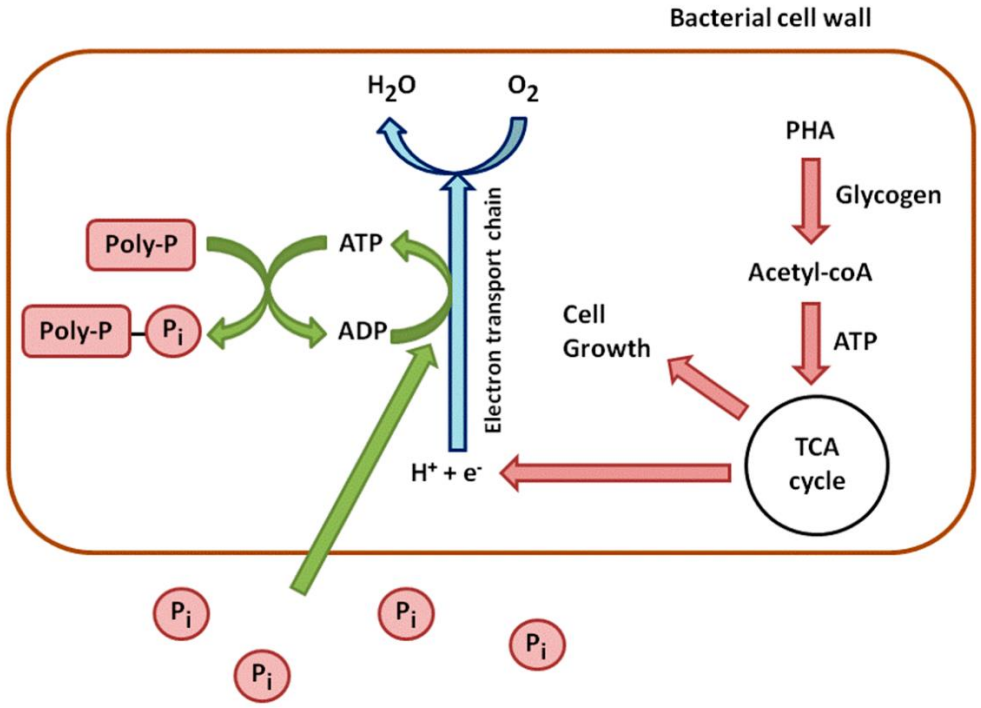


Figure 1.3. Anaerobic metabolism of PAOs (adapted from & Dudgeon et a., 2010).

1.7. Aim of the study

The overall objective of this study was to determine the Accumulibacter species involved in EBPR in the Hias process using the expression of the polyphosphate kinase (*ppk1*) gene as a genetic marker.

To achieve this, the study focuses on the following specific objectives:

- To analyse the *ppk1* sequence variants to determine conserved unique sequences for phylogenetic analysis and for primer design.
- To design primers for the specific amplification of each of the *ppk1* sequences identified in the metagenome study.
- To establish and optimize RT-qPCR protocols for expression of each of these ppk1 sequences, validated by cloning and sequencing the amplicons.
- To use the optimized protocol to analyse *ppk1* expression throughout one EBPR cycle using RT-qPCR.

2. Material and Methods

2.1. Study design and area

A laboratory facility at the Hamar, in Inland Norway University of Applied Sciences, Norway, was used for this research study and the project was experimental. The aims and objectives of this study was accomplish as demonstrated in experimental workflow in Figure 3.1.

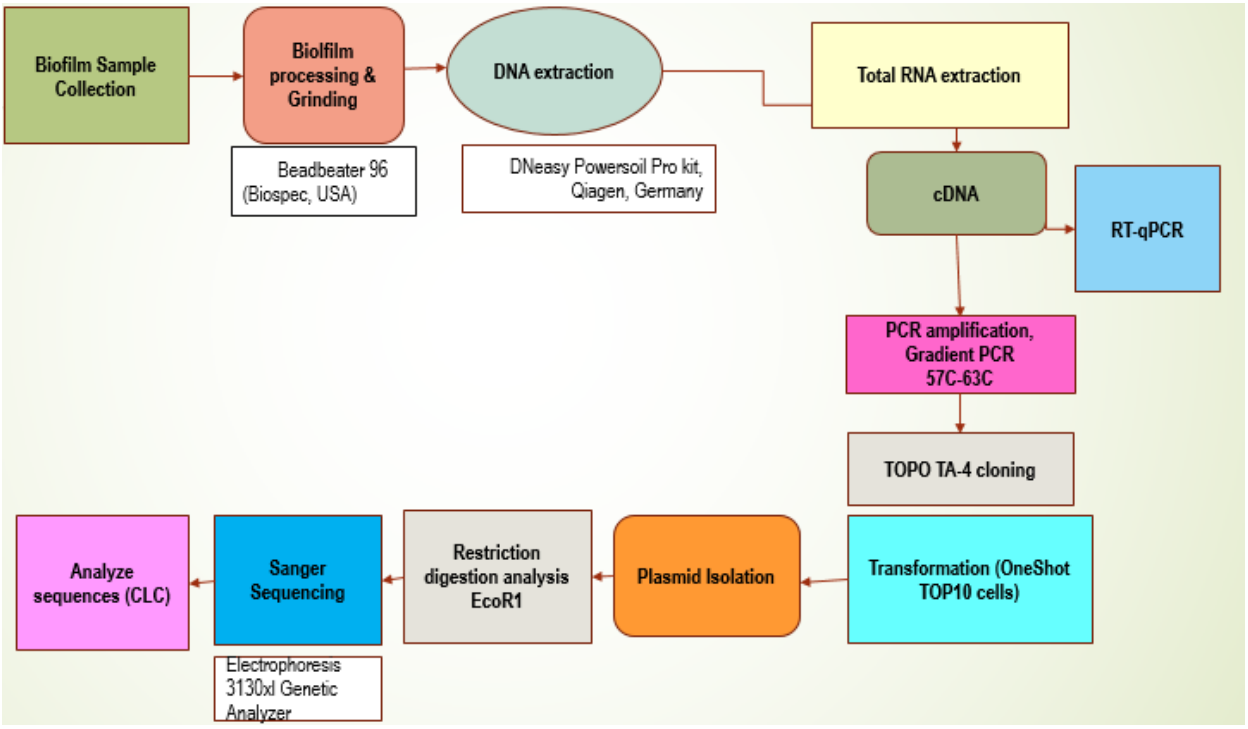


Figure 2.1. Experimental workflow of this study to determine the *Accumulibacter* species involved in EBPR in the Hias process utilizing the polyphosphate kinase gene expression as a genetic marker."

2.2. Primer design to target Accumulibacter *ppk1* sequences

To determine the *Accumulibacter* population in HIAS reactor zone, the *ppk1* gene, a phylogenetic marker, was amplified samples using PCR. Gene-specific primers were designed by Wenche johansen based on the sequences of the eight *Ca*. Accumulibacter *ppk1* sequences obtained from previously metagenomics study described in (Appendix-B). Additionally, one degenerate primer set were also designed to target and amplify all eight sequences. The primer

length was set to be 20 and 21 nucleotides long, Tm (DNA melting temperature) was set to be between 60℃ - 65.7℃, and a GC content ranging from 57-65%. The expected amplicon size range was 234bp. (Appendix-A, Table A).

2.3. Bioinformatics analysis

The phylogenetic analysis of 8 *Ca. Acumulibacter* PPK1 sequences obtained from the previous metagenomics study (Appendix-B) was performed to infer their relationship and their divergence from a common ancestor. MEGA 11 software was used for constructing the phylogenetic tree. A neighbour joining method with a bootstrap value of 1000 was applied, and *E.Coli* designated as an outer group aiming to infer evolutionary relationship within the *Ca. Acumulibacter* species.

The Pfam domain composition of the 8 sequences within the *Ca. Acumulibacter* species was analysed to identify domains composition and to identify conserved regions. The *ppk1* gene nucleotide sequence translated into protein sequence and Pfam domains compositions was analysed using CLC Workbench (Pfam A v35). Full lengths Pfam domain analysis was performed to understand the functional aspects of the PPK1 across 8 *Ca*. Accumulibacter strains (Appendix C).

2.4. Sample collection and processing

Three biological replicates of biofilm carriers were taken randomly from each distinct reactor zone of the Hias Process, e.g. three anaerobic (1, 2, 3) and aerobic (4, 6, 8, and 10) zones. Samples were transferred to a double Ziplock bag. The bags were sealed before placing the samples into liquid nitrogen (at -196°C). To prepare the mortar, 100% ethanol was burned off, and then it was stored at -80°C in liquid nitrogen. One biofilm carrier was directly transferred from liquid nitrogen to the pre-cooled mortar. Following this, the biofilm carriers containing the biofilm were crushed in liquid nitrogen. Any large, non-grindable plastic pieces were removed. The resulting powdered material was divided into three equal portions, which were then transferred into a 1.5 mL micro centrifuge tube and promptly stored at -80°C for subsequent nucleic acid isolations.

2.5. DNA Isolation

DNA was isolated from 21 biofilm samples for the specific amplification of each of ppk1 primer pairs identified in the metagenome study of 8 *Ca. Accumulibacter* sequences and to optimize PCR conditions. DNeasy PowerSoil Pro kit was used for DNA extraction (DNeasy Powersoil Pro kit, Qiagen, Germany). Around 500 mg of crushed biofilm powder from 7 different zones (zones 1, 2, 3, 4, 6, 8 & 10) was added to each of 7 micro centrifuge tubes. 800 µL of CD1 solution was added to each tube and the samples were vortexed thoroughly to ensure proper mixing. The samples were centrifuged at 15,000 x g for 1 minute. The supernatant was transferred to a sterile micro centrifuge tube (2 ml tube provided with the kit). 200 µL of CD2 solution was added in each tube, followed by centrifugation at 15,000 x g for 1 minute. Subsequently, 600 µL of CD3 solution was added to the sample, and 650 µL of the resulting lysate was loaded onto an MB Spin Column and centrifuged at 15,000 x g to bind the DNA to the column. 500 µL of C5 solution was added to the MB Spin column, followed by centrifugation at 15,000 x g for 1 minute. A subsequent centrifugation step was done at 16,000 x g for 2 minutes. In the final step, DNA was eluted from the MB Spin column by adding 100 µl of C6 solution and centrifuged at 15,000 x g for 1 minute. The extracted DNA was analyzed by 1% agarose gel electrophoresis. The concentration and purity of DNA were evaluated spectrophotometrically using NanoDrop instrumentation (Thermo Scientific, USA).

2.6. RNA Isolation

RNA was isolated from 21 biofilm samples taken from various zones (as mentioned in section 3.4). The lysing reaction proceeded upon addition of the samples to tubes containing 1 mL of TRIzol™ reagent (4 °C) and PowerBead Pro tubes (containing silica beads) (Invitrogen, USA). The Mini Beadbeater96 (Biospec, USA) was used to crush the samples. Beating was repeated three times (per run: 40 seconds) with one-minute rest on ice. Following the bead-beating process, the suspension was transferred to a new tube after a 5-minute ice incubation.

In the phase separation step, 1 mL of TRIzol™ reagent was mixed with 300 µL of chloroform and 30 µL of water. The mixture was then centrifuged at 12,000 × g for 15 minutes at 4°C. Three phases resulted from the mixture: a lower red phenol-chloroform phase, an interphase, and a colorless upper aqueous phase. The RNA-containing colorless upper aqueous phase (about 600 µL) was transferred to a fresh tube. After mixing, an equivalent volume of 70%

ethanol was added. Following this, 700 µL of wash buffer flow-through was discarded, and the procedure was repeated.

The volume of 700 µL wash buffer1 was added and centrifuged for 15 seconds at 12,000 × g. Then, 500 µL of wash buffer 2 was added and the centrifugation step was repeated to further purify the sample. After the last washing step, the sample was centrifuged for 1 minute at 12,000 × g for membrane drying. RNA was eluted by adding 50 µL of RNase-free water followed by centrifugation at 12,000 × g for 2 minutes, the isolated RNA was analyzed through 1% agarose gel electrophoresis. Concentration and purity were evaluated spectrophotometrically using NanoDrop instrumentation (Thermo Scientific, USA). The samples were stored at -80 °C.

2.7. Optimization of PCR conditions for RT-qPCR

PCR Gradient Polymerase Chain Reaction (PCR) was used to determine the optimal annealing temperature range from 56°C to 64°C in PCR to amplify ppk1 sequences using primer pairs designed in section 3.2. The gradient PCR procedure allowed the identification of each primer pair's annealing temperature, yielding the most efficient amplification for the identification of *ppk1* for further optimization of conditions for the RT-qPCR reaction. The cDNA was amplified using all primer pairs targeting *ppk1* following the thermal profile protocol for PCR. The cDNA template used for PCR amplification was diluted 1:10.

The following table presents the PCR setup reactions; 8 reactions for PCR amplification using seven different primer sets and with an annealing temperature range 57°C to 63°C; and one reaction for negative control reaction (NTC).

Table 2.1. Experimental setups of PCR: Master Mix (MM) preparation for PCR amplification of the reverse transcribed cDNA.

Initial conc.	Component/ Tubes	Final conc.	Volm. (µl) (1rxn)	Volm. (µl) (8 rxn)
25mM	MgCl2	2.5mM	2	16
10×	Buffer B1	1×	2	16
10 mM	dNTPs	0.2 mM	0.4	3.2
10 µM	Forward Primer	150nM	0.3	2.4
10 µM	Reverse Primer	150nM	0.3	2.4

5 U/mL	HOT FIREpol DNA pol.	0.04U/µl	0.15	1.2
H20			13.85	110.8
	Total volume		**19 µl**	**152 µl**

Table 2.2. Gradient PCR setup amplifying cDNA fragments using 8 primer pairs targeting ppk1.

Initial Denaturation	Step 1: 95°C for 15 minutes
Amplification (35 cycles)	Step 01: 95°C for 30 seconds (Denaturation) Step 02: 57°C to 63°C for 30 seconds (Annealing)
Extension	Step 01: 72°C for 20 seconds

2.8. TOPO TA Cloning

The PCR-amplified cDNA containing free deoxyadenosine (A) at the 3' ends was cloned into a TOPO vector. The vector used in this study was pCR™4-TOPO® vector significantly simplifies the process of TA cloning by providing linearized single 3´ thymidine (T) overhangs. It comes with covalently bounded Topoisomerase (Solis BioDyne, 2020). Therefore, this process eradicates the need for ligase or restriction enzyme as Topoisomerase executes both the cutting and pasting purposes. It also aids in the addition of single A to the 3´ ends of amplified products. The linearized vector, with its overhanging 3´ T residues allows the insert with 3` A overhangs to seamlessly ligate and circularize the construct (Bernard et al., 1993, 1994). TOPO-TA cloning and sequencing were performed to verify the specific PCR amplification of *ppk1* cDNA. After gradient PCR amplification, only C&A primer pairs targeting *ppk1* were selected for cloning into the TOPO® vector. To prepare the cloning reaction, the amplified PCR product of 1 µl was mixed with water, salt solution (1 µL), and TOPO® vector 10 ng/µL.

Table 2.3. Preparation of ligation reaction of PCR amplified fragments into the TOPO® vector based on the protocol of TOPO® TA Cloning® Kit for Sequencing.

TOPO Cloning Reaction	
Components	Volume
PCR product	0.5-1 µl
Salt Solution	1 µl
TOPO vector	1 µl
PCR grade water	Upto 3 µl
Total	6 µl

2.8.1. Transformation by OneShot Top 10 competent cells

The One Shot® TOP10 and DH5α™-T1R competent cells comprising *E. coli* vials were thawed on ice and separated into two tubes, with 25 µL in each. The process is called transformation, in which the competent cells use their capacity for taking up foreign DNA (*ppk1* insert) competently. Then, 25 µL of OneShot Top 10 chemically competent cells (Invitrogen, USA) were transformed by adding 2 µL of the ligation reaction. The cells were initially incubated on ice for 5–30 minutes on ice. Afterward, they are heat-shocked at 42°C for 30 seconds without shaking. After heat-shocking, the tubes were instantly transferred back to ice, and 250 µL of S.O.C. medium was added to each tube. The tubes are then incubated at 37°C with horizontal shaking (200 rpm) for 1 hour. The transformed cells were cultured overnight on LB plates supplemented with 50 µg/mL Kanamycin and were subjected to inoculation with volume of 10 µl, 50 µl, and 100 µl of the transformation mixture. Colonies were initially cultured overnight at 37°C, but due to the small size of colonies, an additional 24-hour incubation was implemented. After incubation, six well-isolated colonies from each of the two transformed reactions were picked and identified.

2.8.2. Purification of Plasmid DNA

Three positive colonies were introduced into plastic sterile culture tubes containing 3 mL of LB broth with 50 µg/mL of kanamycin for each transformation response. The samples were incubated for 24 hours at 37°C with agitation at 225 rpm. Subsequently, following the manufacturer's instructions, the Plasmid DNA (30 µL eluted out using PCR water) was isolated

using the PureYield Plasmid Miniprep System (Promega, USA). The concentration of plasmid DNA was determined using a NanoDrop Spectrophotometer (Thermo Scientific, USA).

Three Positive clones per transformation reaction were identified through restriction digestion analysis to further confirm the presence of the *ppk1* gene inserts. The restriction digestion analysis reaction mix consisted of ~250 ng of isolated plasmid DNA mixed with 1X EcoRI Buffer and EcoRI enzyme (10 U) (USA, NEB). The reaction mix was incubated at 37°C for 1 hour (see reaction details on Appendix-E, Table B). 1% TAE agarose was used to analyze the restriction digestion reaction.

2.8.3. Cycle sequencing & capillary electrophoresis

The BigDye Terminator v3.1 Cycle Sequencing Kit (USA, Applied Biosystems) was used for sequencing. The sequencing reaction mixture consisted of 200 ng of plasmid DNA from 6 plasmid samples amplified using A&C primer pairs, 5.0 µL of H_2O, 2.0 µL of sequencing buffer (5X), and T7 forward primer and T3 sequencing primer at a concentration of 0.1µg/µL each. The following thermocycler protocol was applied:

Table 2.4. PCR program used for performing the Sanger sequencing using T7 forward primer and T3 sequencing primers for PCR amplification.

Big Dye Cycle Sequencing Steps:

Initial Denaturation	Step 01: 96°C for 60 seconds
Amplification (15 cycles)	Step 01: 96°C for 10 seconds Step 02: 50°C for 5 seconds Step 03: 60°C for 1 minute and 15 seconds
Extension (5 cycles)	Step 01: 96°C for 10 seconds Step 02: 50°C for 5 seconds Step 03: 60°C for 1 minute and 30 seconds
Post-Dissociation Steps(∞)	Step 01: 96°C for 10 seconds Step 02: 50°C for 5 seconds Step 03: 60°C for 2 minutes Step 04: 4°C hold indefinitely (Cooling)

2.8.4. Purification of the extension products

After performing the ethanol precipitation technique, eight extension sequencing products were purified. This involved adding 2 µL of 3M NaOAc (pH 5.2), 96% EtOH, 2 µL of 125 mM EDTA, and 52 µL of H_2O to the reaction mixture, followed by incubation at room temperature for 15-minutes. The samples were subsequently centrifuged at maximum speed for 30 minutes at 4°C. After washing the resultant DNA pellet with 70% ethanol, it was centrifuged again for ten minutes at maximum speed. Following the aspiration of the supernatant, the pellet was allowed to dry for half an hour at room temperature. In the final step, the pellet was re-suspended in 10 µL of deionized formamide. Cycle sequencing and capillary electrophoresis was carried out using Applied Biosystems' 3130xl Genetic Analyzer (USA). Upon receiving the sequencing data, CLC Workbench was used for aligning and removing conflicts, differences, or erros in the sequence process that do not align with each other due to error in sanger sequencing. Initially, individual sequencing reads were aligned with the consensus sequence, which represents the most common nucleotide at each position derived from multiple alignments. This alignment facilitated the identification of any mismatches between individual reads and the consensus sequences. The data then underwent alignment to the reference sequence to identify conflicts. Conflicts were then resolved by selecting the nucleotide that matched the consensus sequence at each potion. Subsequently, the sequences were Blasted against the NCBI nucleotide databases to identify the target bacterial species in Hias bioreactor.

2.9. RT-qPCR reaction

A quantitative polymerase chain reaction analysis using reverse transcriptase (RT-qPCR) was performed to assess the *ppk1* gene expression level in each reactor zone. RNase-free water was added to the 21 RNA samples to dilute them to 100 ng/µL, and NanoDrop (Thermo Scientific, USA) was used to confirm the concentration. For RT-qPCR analysis, a total of 1 µg of RNA was extracted. The RNA was primed with random hexamers and then reverse-transcribed to cDNA using the First-Strand Synthesis System SuperScript IV (Invitrogen, USA).

2.10. Reverse Transcribed RNA to cDNA

Extracted RNA samples were reverse transcribed to cDNA, utilizing the First-Strand Synthesis System SuperScript IV kit (USA, Invitrogen). Genomic DNA was initially digested by

incubating 1 µg of total RNA with DNaseI Reaction Buffer (10X) and DNaseI (Amp Grade, 1U/µL) for 30 minutes at 37 °C. After adding 25 mM EDTA to the reaction mixture to activate DNaseI, the mixture was incubated at 65 °C for 10 minutes.

For the reverse transcription of RNA samples, ~ 450 ng of total RNA was mixed with 50 µM random hexamer primers and 10 mM dNTP. The reaction mixture was heated at 65 °C for 5 minutes and then incubated for 1 minute on ice. The reverse transcription reaction was set up by adding 40 units/µL RNaseOUT™ Recombinant RNase Inhibitor, First-Strand Buffer (5X), and 200 units/µL SuperScript™ III RT enzyme. 0.1M of DTT (Dithiothreitol) was mixed in the tube and heated for 15 minutes at 70 °C. Furthermore, a control reaction without reverse transcription was set up by adding First-Strand Buffer (5X) and 0.1M of DTT.

In the RT-qPCR method, overall, two control reaction was set up during cDNA synthesis, one control with no reverse transcriptase (without SuperScript™ III RT enzyme), and one control reaction with no template (NTC). Using the RNAse-free water the cDNA samples were diluted to 1:10 ratio. The cDNA was amplified by using A and C primer pairs at a concentration of 10µM along with HOT FIREPol, EvaGreen and qPCR Supermix (Estonia, Solis Biodyne) (See Master Mix in Appendix F, Table B).

The Applied Biosystems 7500 Fast Real-Time PCR System (USA) was used to amplify the samples using Microamp Fast Optical 96 Well Reaction Plates with a volume of 0.1 mL. The amplification process began with an initial denaturation at 95°C for 15 minutes in the RT-qPC. This was followed by 40 amplification cycles involving denaturation at 95°C for 30 seconds, annealing for 20 seconds at 58°C (optimized by prior gradient PCR analysis), and extension for 32 seconds at 72°C. After the cycling phase, a post dissociation analysis was conducted consisting of the following steps: denaturation for 15 seconds at 95°C and annealing for 1 minute at 58°C.

3. Results

3.1. Bioinformatics analysis of *Ca. Accumulibacter* PPK1 protein sequences

In a previous metagenomics study of the Hias biofilm, 8 different sequences encoding the *ppk1* gene annotated to various species within the genus *Ca. Accumulibacter* were identified and retrieved (See in Appendix-B). The *ppk1* gene nucleotide sequence translated into protein sequence and Pfam domains compositions was analyzed using CLC Workbench (Pfam A v35).

3.1.1. *Ca.* Accumulibacter PPK1 domain structure and composition

The Pfam domain composition of the 8 sequences was analysed to identify domain composition and conserved regions (Fig. 3.1). All sequences showed similar domain architecture, with the exception of sequences 1, 2, 3, and 4, the domain analysis results display that the PP kinase C domains in sequences 1-4 are shorter than the PP_kinas_C domain of sequences 5-8. All sequences contained the PP-kinase_N_terminal domain composed of 107 aa, and PP_Kinase domain composed of 138 aa length and the PP_kinase_domain 353 aa length in only 5, 6, 7 and 8 number of *Ca. Accumulibacter* Pfam domain sequences.

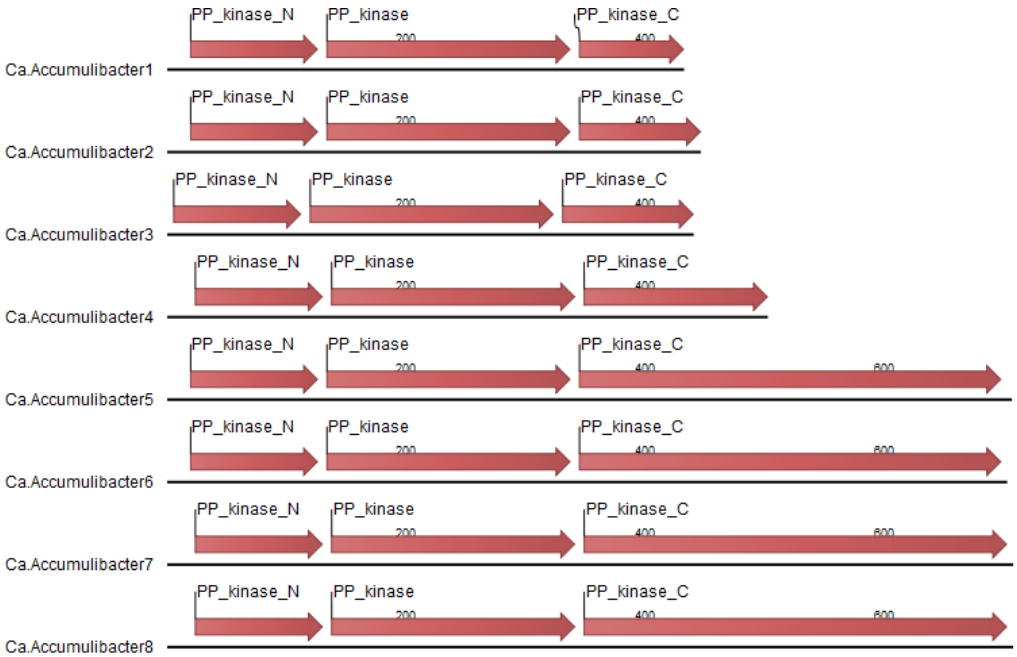


Figure 3.1. Pfam Domain Composition of eight Ca. Accumulibacter PPK1 Sequences: The presented PPK1 sequences were analysed using previous metagenomics study data. The Pfam domain composition is composed of three subdomains (PP_kinase_N, PP_kinase,

PP_kinase_C). The N-terminal PP_kinase_N and PP_kinase domains are identical in all the sequences, while the C-terminal PP_kinase_C domain varies among these sequences, with sequences 1 to 4 have comparatively small sequence length of the PP_kinase_C domain. (PP_kinase_N_ N-terminal polyphosphate kinase domain, PP_kinase_ Polyphosphate kinase domain, PP_kinase_C_ C-terminal polyphosphate kinase domain).

3.1.2. Phylogenetic analysis of *Ca*. Accumulibacter PPK1 sequences

All domains were included for construction of phylogenetic tree based on similar Pfam domain compositions among the 8 different *Ca*. Acumuliabcter sequences. A phylogenetic analysis was performed to evaluate the evolutionary relationship between these sequences. A phylogenetic analysis of the full-length *Ca.* Accumulibacter PPK1 sequences 1-8 was performed to investigate their evolutionary relationship using MEGA 11 soft wear (Fig. 3.2). The *ppk1* gene from *E. coli* was used as outgroup sequence.

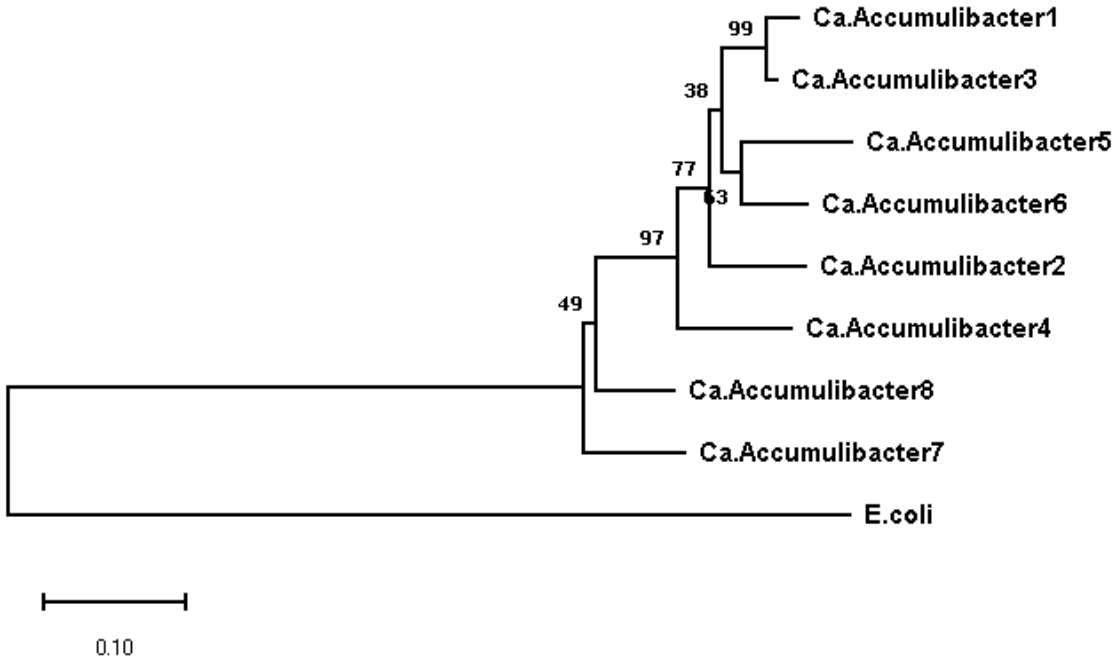


Figure 3.2. Phylogenetic analysis depicting the evolutionary relationship of PPK1 sequences among 8 *Ca. Accumulibacter* strains: The phylogenetic tree was constructed by MEGA 11 using

a neighbour-joining method with bootstrap support value based on 1000 replicates, utilizing PPK1 protein sequences taken from previous metagenomics study (See in Appendix B). E. coli was used as an out-group to root the tree. The strains are grouped into distinct clades: clades 1 and 3, clades 5 and 6, clades 2 and 4, and clades 8 and 7 having varying degrees of relatedness.

The bootstrap is a simulation technique to estimate sample size and analyse data (Walters, S. J., 2004). It involves repeatedly selecting random samples from the original data, with replacement, and is useful for hypothesis testing (p-values), standard error (SE), and confidence interval (CI) estimation (Davison & Hinkley (1997). If we get this observation 100 times out of 100, then this supports our result.

The phylogenetic analysis of full-length domains of PPK1 sequences among eight *Ca. Accumulibacter* strains, reveals distinct evolutionary relationships with varying bootstrap support. *E. coli* serves as a distant reference point, emphasizing the divergence of *Ca.* Accumulibacter strains. The clade comprising of *Ca*. Accumulibacter1 and *Ca*. Accumulibacter3 shows high degree of confidence, with a bootstrap value of 99, indicating high confidence value of evolutionary relationship. The clade including *Ca*. Accumulibacter5 and *Ca*. Accumulibacter6 with a bootstrap value of 63 suggests some confidence in the existence of this branch. The broader clade branch including *Ca.* Accumulibacter2 and *Ca.* Accumulibacter4 is strongly supported with a bootstrap value of 97. Contrary to this, the clade encompassing *Ca*. Accumulibacter8 and *Ca*. Accumulibacter7 exhibits a low confidence and support, with a bootstrap value of 49. These values indicate robust evolutionary relationships among *Ca*. Accumulibacter1 and *Ca*. Accumulibacter3 strains and potential variability in others, providing insights into the evolutionary diversification within the *Ca.* Accumulibacter genus.

3.2. PCR amplification of cDNA fragments (*ppk1)*

Eight primer sets were designed (Appendix A). Each primer set was designed to specifically amplify different regions of the *ppk1* sequences from various *Accumulibacter* species. Agarose gel electrophoresis was performed to analyse all the PCR reactions. The reverse transcribed cDNA, derived from zones 4 and 6, was used as the template for PCR amplification. These primer sets were tested for their specific amplification using cDNA generated form total RNA isolated from a specific biofilm samples.

Figure 3.3 display the results of PCR amplification using primer pairs A, C, E, F, and G at two different annealing temperatures (57℃ & 62℃). The expected 234 bp fragment amplified using A&G primer sets, was observed at both temperatures. However, very low intensity of PCR amplified products was observed using the G primer pair at both temperatures of 234bp. PCR amplification using primer sets C, E, F, G did not result in any amplification of DNA, therefore, gradient PCR was performed to determine suitable annealing temperature.

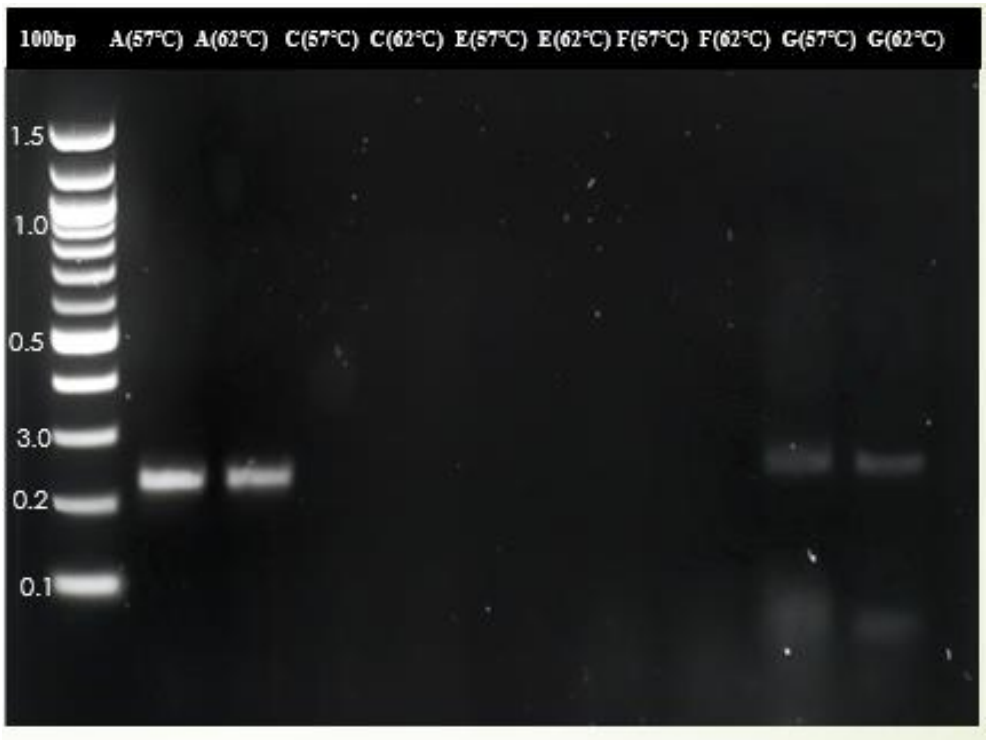


Figure 3.3. Agarose gel electrophoresis of PCR amplified fragments: Five primer pairs (A, C, E, F and G) were amplified in PCR at 57℃ and 62℃ annealing temperatures. The amplicons were used in replicates and marked as follows: A(57℃), A(62℃), C(57℃), C(62℃), E(57℃), E(62℃), F(57℃), F(62℃), G(57℃), G(62℃).The molecular weight marker used was a 100bp DNA ladder.

To optimize the PCR conditions further, gradient PCR was performed to determine the suitable annealing temperature for each of 8 primer sets used for *ppk1* amplification in this research. For cDNA amplification, 7 primer pairs (A, B C, E, F, G, H & M) were used in a gradient PCR with an annealing temperature range of 57℃ to 63℃ (Figure 3.4). A band of the expected size of 234bp was amplified using primer pair A at all annealing temperatures. The cDNA fragments amplified using primer pair H exhibited annealing within temperature range of 57℃ to 63℃,

however, another band of approximately 500bp was detected, indicative of non-specific amplification. Hence, the H primer pair was excluded from further analysis. The remaining primer sets either exhibited non-specific amplification or failed to amplify, as a result, primer pairs E, F, G, H, and M primer pairs were excluded from further analysis. All no template control reactions exhibited an absence of bands, confirming that no template cDNA was present in the reactions. However negative control of G primer pairs exhibited contamination. In the previous experiment (at 57℃ and 62℃ using standard PCR), amplification of a fragment using primer sets C were observed. Although there were no amplification observed while using primer sets C in gradient PCR at 57℃ to 63℃ annealing temperature likely due to use of new cDNA templates, variations in the primer binding efficiency occur.

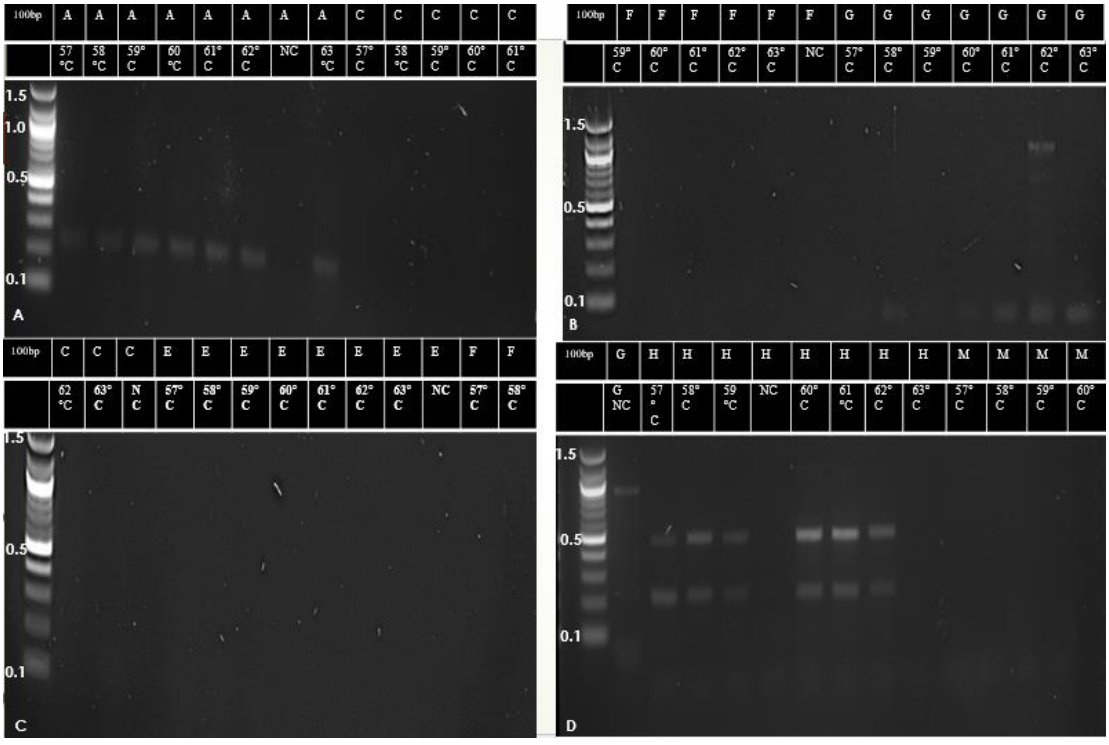


Figure 3.4. Agarose gel electrophoresis of Gradient PCR amplified fragments: Seven primer pairs (A, B C, E, F, G, H and M) were amplified in PCR at annealing temperatures of 57℃ to 63℃. Amplification of each primer pairs with annealing temperature have been presented above in all panels A, B, C and D. The molecular weight marker was a 100bp DNA ladder.

The results of successful PCR amplification of two tested primer pairs (A and C), yielding the expected 234bp size, out of the six primer pairs examined at 58 °C presented in Figure 3.5. Other all PCR reactions, appear to contain nonspecific amplified DNA fragments of various sizes in the agarose gel, indicating potential off-target amplification were excluded from this study. Hence, only two (A and C) out of eight tested primer pairs leads to the amplification of PCR products. Consequently, these two primers were selected for further analysis in this study,

Overall, among the eight tested primer pairs, only primer sets A and C resulted in the specific amplification of a band of the expected size of 234bp (Fig 3.5).

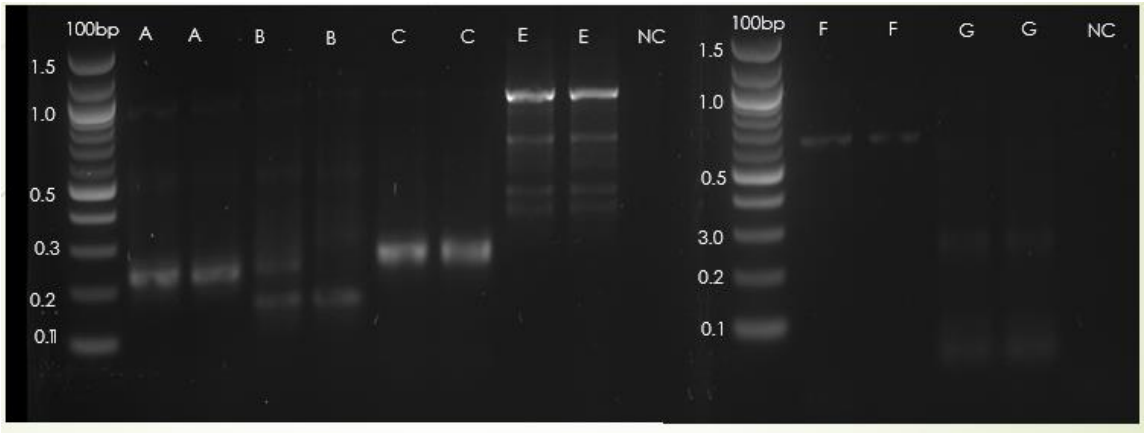


Figure 3.5. Agarose gel electrophoresis of PCR amplified ppk1 cDNA fragments: Six primer pairs A, B C, E, F, G (All samples in replicates) were amplified at 58℃ annealing temperatures. The molecular weight marker used was a 100bp DNA ladder. NC = Negative control (no template control). Replicates of Primer pairs: A - lane 2 and 3; B - lane 4 and 5; C - lane 6 and 7; E - lane 8 and 9; F - lane 12 and 13; G - lane 14 and 15.

3.3. Cloning of PCR amplicons for Sequencing

Two products amplified via PCR using primer pairs A & C were cloned individually into the ToPo-TA cloning vector.

The three random colonies were picked up from each of two transformation reactions (in a total of 6 E. coli clones) and grown overnight on agar plate containing kanamycin and plasmid DNA was isolated from each colony (Figure 3.6). The isolated plasmids were then subjected to restriction digest analysis and sequencing to verify the nature of the cloned insert. Gel

electrophoresis results showed that out of 6 putative positive colonies, 5 were confirmed as positive by restriction digestion analysis. The Gel electrophoresis results indicated two DNA fragments within the desire band size approximately about 234 bp, in addition to the 3956 bp size of the plasmid DNA. Three putative positive clones using primer pair A were selected for restriction digestion analysis to validate the presence of *ppk1* amplicon in panel A figure 3.6. Among the three putative positive clones, clone 1 and 3 resulted in the expected size of 234 bp and 3956 bp, indicating cleavage of the circular plasmid DNA at two different positions. Due to presence of multiple non-specific bands in clone no in 2 Panel A, it leads to the exclusion of that sample in this study

However, all cloning reactions of PCR amplified products using primer pair C resulted in positive clones. Additionally, multiple nonspecific bands were detected in all clones (Refer to figure 3.6, panel B). Cloning reactions showed a comparatively very low quantity of DNA fragments, with all clones showing a band size of 3956 bp of plasmid and 234bp of *ppk1* insert. Generally, except for clone No.2 from primer pairs A (Panel A,2nd clone), all other five clones were considered positive. Overall, from Panel A clone 1 and 3, panel B clone 1, and 3 plasmids were selected for Sanger sequencing to determine the nucleotide sequence and to identify the nature of *Ca. Accumulibacter* strains.

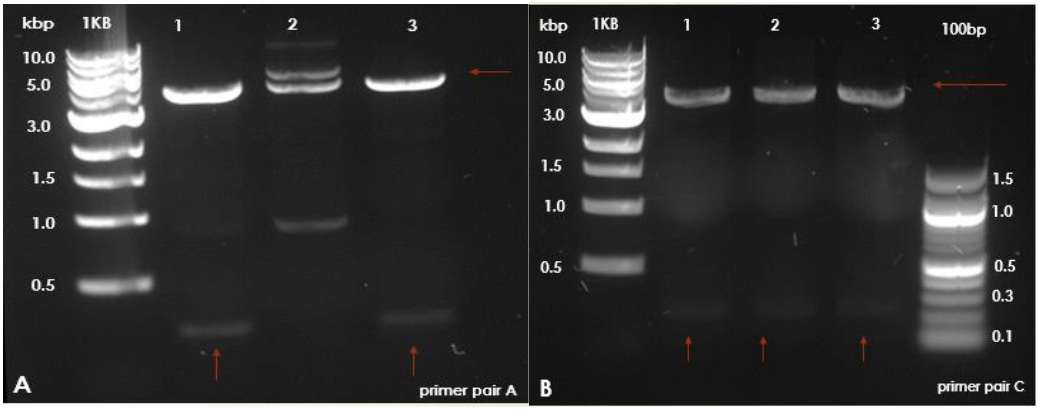


Figure 3.6. Agarose gel electrophoresis of restriction digestion analysis (EcoRI) isolated plasmids: Cleavage reactions for restriction digestion analyses were performed to confirm the presence of positive clones using primer pairs A (Panel A) and primer pairs C (Panel B). Each number 1, 2 and 3 in panel A and B represent three individual clones. A 100bp ladder is included for size reference on right side (Lane 1) in Panel A represents primer pairs A: The left side displays a 1KB ladder (Lane 1) panel B. The right side display a 100bp ladder (Lane 5). Number 1, 2, and 3 present fragments from three restriction digestion reactions in both panels.

3.4. Determining DNA Sequence via Sanger sequencing

To determine the nucleotide sequence of the cloned *ppk1* gene fragments, the inserts from each two cloning reactions of restriction-digested positive plasmids from section 3.2 were sequenced in both forward and reverse direction using Sanger sequencing, resulting in a total of eight reactions. High quality sequence data from Sanger sequencing were generated (up to 930 bp) containing 234 bp size of *ppk1* gene (Appendix F; Fasta sequences).

The assembled sequences of two plasmids were primarily evaluated through BLAST analysis to identify the nature of the sequence. Results showed that all sequences plasmid created from amplicons of primer pair A corresponded to the *ppk1* gene exhibited a high degree of homology to the *Ca*. Accumulibacter Phosphatis sequence, with 85.96% identity (see Appendix D, Figure B and C). Similarly, the plasmid generated form amplicons of primer pair C also showed 86.32% identity with *Ca. Accumulibacter Phosphatis.* The obtained sequencing chromatograms were analysed to determine the nucleotide sequence of the *ppk1* gene inserts.

Each sequence was individually assembled to generate consensus sequences to confirm the presence of insert. Conflicts were resolved using CLC Workbench software. The aligned consensus sequences were inspected manually to identify any differences among them. In addition, each consensus sequence was analysed for actual gene length by aligning it with the corresponding reference sequences. Using the CLC workbench tool, these conflicting residues were edited and removed from the aligned sequences to ensure the accuracy of the consensus sequence (Sequence information on Appendix F).

It is concluded that despite using two different primer sets and generating of multiple plasmid clones, only one *ppk1* gene sequence was obtained of *Ca. Accumulibacter Phosphatis* following cloning and sequencing. This shows that *Ca. Accumulibacter Phosphatis* present in high amount in Hias bioreactor and accumulate P by highest rate.

3.5. RT qPCR analysis of *ppk1* gene expression in reactor cycle

The purpose of RT-qPCR analysis was to investigate if expression of the *ppk1* gene varies throughout the different reactor zones.

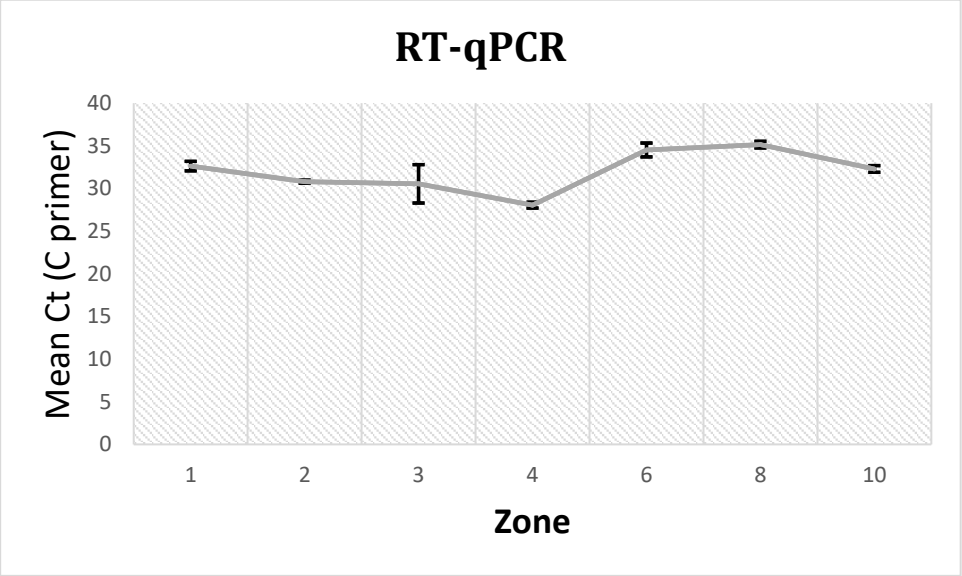


Figure 3.7. Graphical representation of ppk1 gene expression level: Graph displaying error bars representing the mean cycle threshold (Ct) values of amplicons containing primer sets C., indicating the ppk1 expression level throughout the Hias reactor zones.

RT-qPCR was employed to evaluate gene expression profiles of *ppk1* across various zones throughout one reactor cycle. Samples were collected from seven different zones to analyse the expression levels of *ppk1*-specific transcripts amplified using primer sets A and C. Three technical biological replicates of cDNA samples of each 21 biofilm samples were taken. So overall, three cDNA synthesized using one total RNA as template (3 ×21=63). Moreover, cycle threshold (Ct) values form aerobic and anaerobic zones were analysed. The standard deviation of Ct values was calculated to assess the variability in gene expression levels across samples. Furthermore, to visualize the distribution of Ct values and the degree of deviation within the dataset, a graph was plotted (Figure 3.7). Post-dissociation plots depicted single peaks representing single product amplification.

The fluctuation of Ct values across different zones indicates variations in the abundance of the *ppk1* transcripts among samples from various reactor zones. These Ct value patterns signify that in anaerobic zone 1 (Ct > 32.6), Ct values above 32 suggest lower concentration of *ppk1* transcripts reveals downregulation. In zone 2 and 3 (Ct > 30), a slight decrease in Ct value to 33 suggests a slightly higher abundance of the *ppk1* transcripts compared to zone 1, signifying that there is an up regulation of *ppk1* gene expression in zone 2-3.In aerobic zone 4 (Ct = 28), the decrease in Ct value to 28 suggests a substantial increase in the abundance of the *ppk1* transcripts equated to zone 3 proposing an up regulation of *ppk1* gene expression as compared to all zones. Conversely, in aerobic zones 6 and 8 (Ct > 34), Ct values increasing to above 34 suggest a decrease in the abundance of the *ppk1* transcripts compared to Zone 4, so *ppk1* expression were downregulated. In zone 10 (Ct = 32), the Ct value returning to 32 suggests a slight increase in the abundance of the *ppk1* compared to zones 6 and 8. Overall, the fluctuations in Ct values across different zones indicate changes in the expression or abundance of ppk1 in samples. The absence of amplification in the no-RT control reaction suggests that there is no contaminating gDNA in the reaction. Therefore, it can be inferred that the expression of the *ppk1* gene varies throughout the Hias biofilm reactor zones and highly regulated in zone 4 but downregulated in all other aerobic zones, which could be influenced by factors such as environmental conditions, nutrient availability, microbial composition, or metabolic activity across the reactor zones.

A post-dissociation analysis of the RT-qPCR results revealed dual amplification peaks using primer pair A (see Appendix G). Adjustments to the annealing temperature were made, increasing it to 61 ^{o}C instead of 58 ^{o}C in the RT-qPCR assay. Nonetheless results persist. Subsequently, to investigate the nature of the amplification curves of dual peaks specifying presence of 2 products in samples, RT-qPCR amplified 4 transcripts at 61 ^{o}C from zone 10 were taken. After restriction digestion analysis, two positive clones were selected for sequencing. The BLAST analysis confirmed that the one positive clone sequence exhibited homology with a *Sulfuritalea hydrogenivorans* strain sk43H, while the second positive clone sequence showed homology with of *ppk1* gene of *Ca. Accumulibacter phosphatis.*

The presence of dual peaks in the RT-qPCR results could be due the reason that the gene-specific primers were not used for cDNA synthesis, so all RNA present in the sample were reverse transcribed to cDNA. Moreover, another cDNA from another bacterium has been amplified of same size in that reaction. Cloning and sequencing of the resultant amplified products of RT-qPCR confirmed the presence of these two distinct sequences, providing

assurance that the dual peaks observed in the RT-qPCR results were not artifacts but signified genuine amplification of two distinct DNA fragments present in amplicon A, underlining the Hias samples complexity. Hence, *ppk1* transcript dynamics using primer set A were not investigated because of non-specific amplification.

4. Discussion

The nutrients in the wastewater discharge not only contaminate the waterbodies but also cause serious global problems, including medical issues related to their toxicity (Seviour et al., 2003). Therefore, it is important to treat effluent discharges from industries and urbanization process to improve water quality. One of the main components of the discharge water is phosphorus, due to its extensive use in agriculture and fertilizer production (Witek-Krowiak et al., 2022). This necessitates the removal of phosphorus using an effective method. The use of bioreactors is considered an eco-friendly and effective method, wherein phosphorus-accumulating organisms (PAOs) remove phosphorus through their inherent biochemical process (Ødegaard, 2016). PAOs require alternating aerobic and anaerobic conditions to appropriately absorb and store phosphorus. To provide PAOs with optimum growth conditions, the bioreactor is equipped with alternating aerobic/anaerobic zones (Seviour et al., 2003). The aim of this study is to determine the PAO *Accumulibacter* species involved in the Hias process of different EBPR zones using the expression of phosphate kinase (*ppk1*) gene as a genetic biomarker.

4.1. Sequence variant and phylogenetic analysis of *ppk1* gene

The *ppk1* gene was selected as a genetic marker to study the *Ca*. Accumulibacter species and the P accumulation activity in the Hias biofilm. To achieve this aim, the first objective of the study was to analyse the ppk1 sequence variants for the presence of conserved domains. Eight different sequences of *Ca*. Accumulibacter were obtained from metagenomics study, which revealed identical PP_kinase_N (107-138 aa), PP_kinase (204 aa), and PP_kinase_C (400-600 aa) domains among all sequences but comparatively short sequence length of C-terminal PP_kinase_C domains in 1, 2, 3 and 4 sequences (Figure 3.1). Consistent with the previous study, the N domain consists of 2-106 aa residues forming three α-helices. The middle domain (PP_kinase), also known as Head domain consists of 107-321 aa residues and forms the outward head. The C domain, made up of mix of β-sheet and α-helices, consists of 322-627 aa residues and forms the surface on which the N domain lies. These catalytic domains are found to have similarities with the structure of lipid phosphatase and phospholipase D (Achbergerová & Nahálka, 2011).

A phylogenetic analysis of eight different *PPK1* sequences from individual *Ca. Accumulibacter* species exhibited distinct clades: *Ca. Accumulibacter* 1 and 3, 5 and 6, 2 and 4, and 8 and 7. Most branches had bootstrap values above 60, indicating good confidence, except for the *Ca*. Accumulibacter 8 and 7 clades, which had a value below 50, suggesting uncertainty in their relationship.

4.2. Primer Designing and Optimization of RT-qPCR Protocol

After target sequence selection, careful primer designing, and validation are required for specific and efficient amplification in RT-qPCR (Basu, 2015). If the targeted gene is conserved across multiple species, as in our study where the *ppk1* gene is conserved across multiple Ca. Accumulibacter strains observed from sequence and phylogenetic analysis, primers targeting this gene can be used to detect a wide range of microbes in biofilms (Notredame et al., 2000). For this purpose, eight primer sets were designed to target and amplify various Ca. Accumulibacter *ppk1* sequence retrieved from the metagenome study.

To optimize the RT-qPCR protocol, it is necessary to determine the optimum annealing temperature, which can deviate greatly during the experiment. Therefore, it is required to test a range of annealing temperatures to find the optimal conditions (Promega Corporation, 2009). The optimal annealing temperature of eight specific *ppk1* primer sets to assess the gene expression profiles of ppk1 gene was determined using gradient PCR in this experiment. Gel electrophoresis results of PCR amplified fragments using eight primer pairs (A, B, C, E, F, G and M) at different annealing temperature range from 57℃ to 62℃ showed a single 234 bp fragment by using amplicons comprising only primer pair A and C. No successful amplification was observed with primer sets B, E, F, G, and M (Figure 3.3) only A and C primer pairs were studied due to their ability to amplify 234bp fragments in PCR. One possible explanation oof no amplification in (B, E, F, G, H and M primer pairs) is that the *ppk1* transcripts were very low in abundance, resulting in only two specific *ppk1* transcripts being amplified.

4.3. Sequence Validation by Cloning and Sequencing

To measure *Ca. Accumulibacter* strains *ppk1* gene expression in the Hias biofilm reactor applying RT-qPCR methodology, two primer pairs (A and C) among the initial eight primer

sets showed the expected 234 bp fragments and were selected for RT-qPCR, with an annealing temperature of 58°C. The resultant PCR amplicons using (A and C primer sets) were cloned using the TOPO TA cloning kit. The TOPO vector contains an antibiotic selection marker named kanamycin, which aids in the selective bacterial growth harbouring the ligated plasmids (transformed bacteria) (Bernard et al., 1993, 1994). Since the objective was to validate the sequence insert containing the *ppk1* gene, it was necessary to select the bacteria containing the recombinant plasmid. After restriction digestion analysis the gel electrophoresis results showed 234 bp plasmid DNA fragment corresponding to the *ppk1* gene insert and a 3956 bp fragment corresponding to the PCR4-TOPO vector. Overall, six putative positive clones were assessed by digestion analysis, of which five positive clones displayed target specific cleavage of the DNA insert of 234 bp size. The subsequent Sanger sequencing and BLAST analysis confirmed that the verified *ppk1* gene is specific to *Ca. Accumulibacter phosphatis* strain.

Sanger sequencing was used in this study because of its reliability and cost-effectiveness. Overall, four positive plasmids DNA were sequenced. The assembled sequences were BLASTed, and the results showed that all the sequences generated from amplicon using primer pairs A belonged to the gene *ppk1* and exhibited 85.96% homology to the *Ca. Accumulibacter Phosphatis* sequence. An identical homology of 86.32% was observed in plasmids generated from amplicons using the C primer pairs (Appendix D). This homology percentage indicates a significant match to the target gene *ppk1* of *Ca. Accumulibacter phosphatis* sequence. This sequence validation using cloning, Sanger sequencing, and BLAST analysis confirmed that the tested A and C primer sets are specific to *ppk1* gene. The specificity of the primers ensures that the RT-qPCR data will accurately reflects the gene expression levels (Rodríguez et al., 2015). This justifies That *Ca.Accumulinacter phosphatis* species present abundantly as compared to other bacterial species and important for P accumulation in the Hias proses.

4.4. Population dynamics and Expression Analysis in EBPR Cycle

All sequences containing amplicons of the *ppk1* gene from *Ca. Accumulibacter Phosphatis* strains showed a high accumulation of the *ppk1* gene in the Hias reactor. The abundance of *Ca. Accumulibacter phosphatis* in the Hias bioreactor is supported by another study, stated the occurrence of *Ca.Accumulibacter phosphatis* abundantly and helps in accumulation of

polyphosphate (polyP) in five EBPR activated sludge reactors operated with different carbon sources (Fukushima et al., 2007).

RT-PCR combined with qPCR is a method that measures RNA levels by using the cDNA, allowing for rapid quantification and detection of gene expression fluctuations (Yilmaz et al., 2012). Sequencing results confirmed the existence of Ca. *Accumulibacter Phosphatis*. Variation in the abundance of *Ca. Accumulibacter* might be overlooked due to the absence of normalization in the RT-qPCR data. To ensure accurate and dependable qPCR results, it is recommended to use stably expressed reference genes for normalization. This ensures that the observed changes are caused by biological differences rather than technical variability, as per the guidelines set forth by MIQE (Bustin et al., 2009). However, biofilms present in the Hias reactor, are intricate microbial communities made up of a variety of species and have different gene expression patterns. It is challenging to identify reference genes with stable expression in all samples and situations due to their complexity. Additionally, it is difficult to achieve similar amplification efficiency for a reference gene in a complex microbial community. Variations in amplification efficiency have the potential to introduce bias, making it difficult to correctly normalize the data and investigate the expression of microbial genes. As a result of these challenges, no normalization was performed to ensure the reliability of the RT-qPCR method.

Without normalization, the different quantities of RNA can affect the RT-qPCR results (Nolan et al., 2006). To compensate for the lack of normalization in the RT-qPCR, the same amount of total RNA was used from each of the 21 samples for the cDNA synthesis reaction, as measured using a NanoDrop spectrophotometer. However, there is a risk of human error when pipetting the exact amount of RNA, and the NanoDrop spectrophotometer due to its simplicity and speed is not the most accurate method for RNA quantification, particularly measuring the low concentrations (Imbeaud et al., 2005). More precise methods, such as Qubit fluorometry or an Agilent Bioanalyzer could have been used to address this problem (Dang et al., 2016), but unfortunately, they were not performed in this study. To ensure the reliability of the results, the technical replicates were consistent, with the cycle threshold (Ct) values varying by only 0.5 cycles or less. The significance of consistency of technical replicates in reliability and reproducibility of RT-qPCR data is emphasized by Taylor et al. (2010). This suggests that the RT-qPCR results are reliable and reproducible.

The mean Ct values obtained from RT-qPCR indicate the expression levels of the *ppk1* gene across different reactor zones of Hias process. These Ct values demonstrate varying gene expression levels among the 21 cDNA samples, ranging from about 28 to 35 (Figure 3.1).

Lower Ct values correspond to higher gene expression. The Ct values between zones 1, 2, and 3 (anaerobic) fluctuate between 30 and 32.6, revealing upregulation of *ppk1* transcripts. This implies that the level of *ppk1* increased in anaerobic environments, suggesting that *Ca. Accumulibacter phosphatis* is actively involved in processes leading up to phosphate uptake. The *Ca. Accumulibacter phosphatis* breaks down polyP in the anaerobic phase to generate energy, releasing phosphate into the surrounding environment and absorbing volatile fatty acids stored as intracellular polyhydroxyalkanoates (PHAs) (Burow et al., 2007). The bacteria appear to be preparing for the next aerobic phase, in which they will take up and store phosphate as polyP, based on the high expression of *ppk1* in anaerobic conditions. A slightly lower Ct value of 28 observed in zone 4 suggests comparatively higher *ppk1* expression compared to other zones. In zone 4 an upregulation of *ppk1* during the early aerobic phase. Ct values above 32-34 in zones 6, 7, and 8 indicate down regulation of *ppk1* transcripts expression. The increased Ct values represent the downregulation of *ppk1* gene expression in aerobic zones 6 to 10 in aerobic environment. During the aerobic stage, *Ca. Accumulibacter phosphatis* utilizes the stored PHAs to produce the energy needed for phosphate uptake from the environment and storage of polyP (Oehmen et al., 2007). *Ppk1* plays a role in polyP synthesis, so the expected outcomes were upregulation of the *ppk1* gene during the aerobic phase. However, there is also a *ppk2* gene involved in the production of PPK2 enzymes, which are also responsible for hydrolysis of polyP (Rao et al., 2009). When polyP levels are high, PPK2 activity increases to break down polyP, thereby reducing the cellular requirement for further polyP synthesis. This decrease in the need for polyP synthesis causes ppk1 expression to be downregulated through a negative feedback mechanism (Chuang et al., 2013). The cyclical upregulation of the *ppk1* gene during the anaerobic phase and its downregulation during the aerobic phase are crucial for the EBPR process, which regulates phosphate levels in wastewater treatment plants. Despite variation in conditions across different zones of the IKS HIAS reactor, significant fluctuations in the *ppk1* gene accumulation pattern were observed throughout the reactor zones. This finding confirms that microbial population show variability in expression of gene related to P accumulation, regardless of the aerobic and anaerobic conditions in the reactor zones.

5. Conclusion

In conclusion, the study successfully analysed the *ppk1* sequence variants to design primers and optimize the RT-qPCR protocol. The amplified primer products were validated by cloning and sequencing, confirming their specificity to the *ppk1* gene of Ca.*Accumulibacter phosphatis*. Expression analysis using RT-qPCR identified distinct expression patterns of the *ppk1* gene across different reactor zones of the EBPR cycle. The cyclic upregulation and downregulation of ppk1 across anaerobic and aerobic phases demonstrated the metabolic activities of PAOs in P uptake and storage in the Hias reactor zones. One drawback of this study was the normalization in RT-qPCR data, which can result in variability due to different RNA quantities. This issue was addressed by using consistent RNA quantities; however, more advanced RNA quantification methods could have enhanced accuracy. Additionally, the observed cyclic upregulation of *ppk1* in anaerobic phases and downregulation in aerobic phases in *Accumulibacter* needs further investigation in other PAOs. This would help determine whether the observed negative feedback on *ppk1* expression is truly caused by *ppk2* or other factors are at play. Overall, this study contributes to the broader understanding of microbial communities in wastewater treatment and their effective utilization for P removal and storage.

6. References

Aboulhassan, M. A., Souabi, S., Yaacoubi, A., & Baudu, M. (2006). Removal of surfactant from industrial wastewaters by coagulation flocculation process. *International Journal of Environmental Science & Technology*, *3*(4), 327–332. https://doi.org/10.1007/BF03325941

Achbergerová, L., & Nahálka, J. (2011). Polyphosphate-an ancient energy source and active metabolic regulator. *Microbial cell factories*, 10, 1-14.

Aguado, D., Ferrer, A., Seco, A., & Ferrer, J. (2006). Comparison of different predictive models for nutrient estimation in a sequencing batch reactor for wastewater treatment. *Chemometrics and Intelligent Laboratory Systems*, *84*(1–2), 75–81. https://doi.org/10.1016/j.chemolab.2006.03.009

Al-Enezi, G., Hamoda, M. F., & Fawzi, N. (2004). Ion Exchange Extraction of Heavy Metals from Wastewater Sludges. *Journal of Environmental Science and Health, Part A*, *39*(2), 455–464. https://doi.org/10.1081/ESE-120027536

Alexander, G., & Stevens, R. (1976). Per capita phosphorus loading from domestic sewage. *Water Research*, *10*(9), 757–764. https://doi.org/10.1016/0043-1354(76)90093-2

Angell, I. L., Hanssen, J. F., & Rudi, K. (2017). Prokaryote species richness is positively correlated with eukaryote abundance in wastewater treatment biofilms. *Letters in applied microbiology*, 65(1), 66-72.

Anupama, V. N., Amrutha, P. N., Chitra, G. S., & Krishnakumar, B. (2008). Phosphatase activity in anaerobic bioreactors for wastewater treatment. *Water Research*, *42*(10–11), 2796–2802. https://doi.org/10.1016/j.watres.2008.02.014

Arvin, E., & Jenkins, D. (1985). Biological removal of phosphorus from wastewater. *Critical Reviews in Environmental Control*, *15*(1), 25–64. https://doi.org/10.1080/10643388509381726

Badger, D. D., & Thomason, D. E. (2020). Economic and Environmental Impacts of Using Municipal Sewage Effluent for Agricultural Production. In *Ground Water Quality and Agricultural Practices* (pp. 111–126). CRC Press. https://doi.org/10.1201/9781003069782-9

Bernard, Philippe, Philippe Gabarit, El Mustapha Bahassi, and Martine Couturier. 1994. "Positive-Selection Vectors Using the F Plasmid ccdB Killer Gene." *Gene* 148(1):71–74. doi: 10.1016/0378-1119(94)90235-6.

Bernard, Philippe, Katherine E. Kézdy, Laurence Van Melderen, Jan Steyaert, Lode Wyns, Martin L. Pato, Patrick N. Higgins, and Martine Couturier. 1993. "The F Plasmid CcdB Protein Induces Efficient ATP-Dependent DNA Cleavage by Gyrase." *Journal of Molecular Biology* 234(3):534–41. doi: 10.1006/jmbi.1993.1609.

Barnes, D., Bliss, P. J., Grauer, B., & Robins, K. (1985). Pretreatment of high strength wastewater by an anaerobic fluidised bed process. Part II response to organic load transients. *Environmental Technology Letters*, *6*(1–11), 73–78. https://doi.org/10.1080/09593338509384320

Bhagowati, B., & Ahamad, K. U. (2019). A review on Lake Eutrophication dynamics and recent developments in lake modeling. *Ecohydrology & Hydrobiology*, *19*(1), 155–166. https://doi.org/10.1016/j.ecohyd.2018.03.002

Burow, L. C., Kong, Y., Nielsen, J. L., Blackall, L. L., & Nielsen, P. H. (2007). Abundance and ecophysiology of Defluviicoccus spp., glycogen-accumulating organisms in full-scale wastewater treatment processes. *Microbiology*, 153(1), 178-185.

Chuang, Y. M., Belchis, D. A., & Karakousis, P. C. (2013). The polyphosphate kinase gene ppk2 is required for Mycobacterium tuberculosis inorganic polyphosphate regulation and virulence. *MBio*, 4(3), 10-1128.

Corbridge, D. E. C. (2016). *Phosphorus*. CRC Press. https://doi.org/10.1201/b12961

Cordell, D., Drangert, J.-O., & White, S. (2009). The story of phosphorus: Global food security and food for thought. *Global Environmental Change*, *19*(2), 292–305. https://doi.org/10.1016/j.gloenvcha.2008.10.009

Cordell, D., & White, S. (2013). Sustainable Phosphorus Measures: Strategies and Technologies for Achieving Phosphorus Security. *Agronomy*, *3*(1), 86–116. https://doi.org/10.3390/agronomy3010086

Cornel, P., & Schaum, C. (2009). Phosphorus recovery from wastewater: needs, technologies and costs. *Water Science and Technology*, *59*(6), 1069–1076. https://doi.org/10.2166/wst.2009.045

Dai, L., Wang, Y., Li, Z., Wang, X., Duan, C., Zhao, W., Xiong, C., Nie, S., Xu, Y., & Ni, Y. (2020). A multifunctional self-crosslinked chitosan/cationic guar gum composite hydrogel and

its versatile uses in phosphate-containing water treatment and energy storage. *Carbohydrate Polymers*, *244*, 116472. https://doi.org/10.1016/j.carbpol.2020.116472

Dang, J., Mendez, P., Lee, S., Kim, J. W., Yoon, J. H., Kim, T. W., ... & Kim, I. J. (2016). Development of a robust DNA quality and quantity assessment qPCR assay for targeted next-generation sequencing library preparation. *International journal of oncology*, 49(4), 1755-1765.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application* (No. 1). Cambridge university press.

Desmidt, E., Ghyselbrecht, K., Zhang, Y., Pinoy, L., Van der Bruggen, B., Verstraete, W., Rabaey, K., & Meesschaert, B. (2015). Global Phosphorus Scarcity and Full-Scale P-Recovery Techniques: A Review. *Critical Reviews in Environmental Science and Technology*, *45*(4), 336–384. https://doi.org/10.1080/10643389.2013.866531

Driver, J., Lijmbach, D., & Steen, I. (1999). Why Recover Phosphorus for Recycling, and How? *Environmental Technology*, *20*(7), 651–662. https://doi.org/10.1080/09593332008616861

Elser, J. J. (2012). Phosphorus: a limiting nutrient for humanity? *Current Opinion in Biotechnology*, *23*(6), 833–838. https://doi.org/10.1016/j.copbio.2012.03.001

Fernando, Eustace Y., Simon Jon McIlroy, Marta Nierychlo, Florian-Alexander Herbst, Francesca Petriglieri, Markus C. Schmid, Michael Wagner, Jeppe Lund Nielsen, and Per Halkjær Nielsen. 2019. “Resolving the Individual Contribution of Key Microbial Populations to Enhanced Biological Phosphorus Removal with Raman–FISH.” *The ISME Journal* 13(8):1933–46. doi: 10.1038/s41396-019-0399-7

Flemming, H.-C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S. A. & Kjelleberg, S. 2016 Biofilms: an emergent form of bacterial life. Nat. Rev. Microbiol. 14, 563.

Flemming, H.-C., Baveye, P., Neu, T. R., Stoodley, P., Szewzyk, U., Wingender, J., & Wuertz, S. (2021). Who put the film in biofilm? The migration of a term from wastewater engineering to medicine and beyond. *Npj Biofilms and Microbiomes*, *7*(1), 10. https://doi.org/10.1038/s41522-020-00183-3

Fukushima, T., Uda, N., Okamoto, M., Onuki, M., Satoh, H., & Mino, T. (2007). Abundance of Candidatus ‘Accumulibacter phosphatis’ in enhanced biological phosphorus removal activated sludge acclimatized with different carbon sources. M*icrobes and environments*, 22(4), 346-354.

Gauthier, E., Fortier, I., Courchesne, F., Pepin, P., Mortimer, J., & Gauvreau, D. (2000). Aluminum Forms in Drinking Water and Risk of Alzheimer's Disease. *Environmental Research*, *84*(3), 234–246. https://doi.org/10.1006/enrs.2000.4101

Ge, J., Meng, X., Song, Y., & Terracciano, A. (2018). Effect of phosphate releasing in activated sludge on phosphorus removal from municipal wastewater. *Journal of Environmental Sciences*, *67*, 216–223. https://doi.org/10.1016/j.jes.2017.09.004

Granéli, E., Weberg, M., & Salomon, P. S. (2008). Harmful algal blooms of allelopathic microalgal species: The role of eutrophication. *Harmful Algae*, *8*(1), 94–102. https://doi.org/10.1016/j.hal.2008.08.011

HIAS IKS. "Method for biological purification of waste water". NO20140660. 2015.

Heistad, A., Paruch, A. M., Vråle, L., Ádám, K., & Jenssen, P. D. (2006). A high–performance compact filter system treating domestic wastewater. *Ecological Engineering*, *28*(4), 374–379. https://doi.org/10.1016/j.ecoleng.2006.06.011

Helness, H., & Ødegaard, H. (2001). Biological phosphorus and nitrogen removal in a sequencing batch moving bed biofilm reactor. *Water Science and Technology : A Journal of the International Association on Water Pollution Research*, *43*(1), 233–240.

Hu, H., Tong, Y. W., & He, Y. (2023). Current insight into enhanced strategies and interaction mechanisms of hydrogel materials for phosphate removal and recovery from wastewater. *Science of The Total Environment*, *892*, 164514. https://doi.org/10.1016/j.scitotenv.2023.164514

Imbeaud, S., Graudens, E., Boulanger, V., Barlet, X., Zaborski, P., Eveno, E., ... & Auffray, C. (2005). Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic acids research*, 33(6), e56-e56.

Irvine, R. L., Ketchum, L. H., & Asano, T. (1989). Sequencing batch reactors for biological wastewater treatment. *Critical Reviews in Environmental Control*, *18*(4), 255–294. https://doi.org/10.1080/10643388909388350

Jenkins, D., Ferguson, J. F., & Menar, A. B. (1971). Chemical processes for phosphate removal. *Water Research*, *5*(7), 369–389. https://doi.org/10.1016/0043-1354(71)90001-7

Kauppi, L. (1985). The Contribution of Agricultural Loading to Eutrophication in Finnish Lakes. *Water Science and Technology*, *17*(6–7), 1133–1140. https://doi.org/10.2166/wst.1985.0208

Walters, S. J., & Campbell, M. J. (2005). The use of bootstrap methods for estimating sample size and analysing health-related quality of life outcomes. *Statistics in medicine*, *24*(7), 1075–1102. https://doi.org/10.1002/sim.1984

Kawakoshi, A., Nakazawa, H., Fukada, J., Sasagawa, M., Katano, Y., Nakamura, S., Hosoyama, A., Sasaki, H., Ichikawa, N., Hanada, S., Kamagata, Y., Nakamura, K., Yamazaki, S., & Fujita, N. (2012). Deciphering the Genome of Polyphosphate Accumulating Actinobacterium Microlunatus phosphovorus. *DNA Research*, *19*(5), 383–394. https://doi.org/10.1093/dnares/dss020

Kofinas, P., & Kioussis, D. R. (2003). Reactive Phosphorus Removal from Aquaculture and Poultry Productions Systems Using Polymeric Hydrogels. *Environmental Science & Technology*, *37*(2), 423–427. https://doi.org/10.1021/es025950u

Korving, L., Van Loosdrecht, M., & Wilfert, P. (2019). Effect of Iron on Phosphate Recovery from Sewage Sludge. In *Phosphorus Recovery and Recycling* (pp. 303–326). Springer Singapore. https://doi.org/10.1007/978-981-10-8031-9_21

Leyva-Díaz, J. C., Martín-Pascual, J., & Poyatos, J. M. (2017). Moving bed biofilm reactor to treat wastewater. *International Journal of Environmental Science and Technology*, *14*(4), 881–910. https://doi.org/10.1007/s13762-016-1169-y

Li, B., & Brett, M. T. (2015). The relationship between operational and bioavailable phosphorus fractions in effluents from advanced nutrient removal systems. *International Journal of Environmental Science and Technology*, *12*(10), 3317–3328. https://doi.org/10.1007/s13762-015-0760-y

Liu, D., & Zhou, S. (2022). Application of chemical coagulation to phosphorus removal from glyphosate wastewater. *International Journal of Environmental Science and Technology*, *19*(4), 2345–2352. https://doi.org/10.1007/s13762-021-03164-x

Mahmoud, G., Abdel Khalek, M., Shoukry, E., Amin, M., & Abdulghany, A. (2019). Removal of phosphate ions from wastewater by treated hydrogel based on chitosan. *Egyptian Journal of Chemistry*, *0*(0), 0–0. https://doi.org/10.21608/ejchem.2019.9627.1646

Malmqvist, B., & Rundle, S. (2002). Threats to the running water ecosystems of the world. *Environmental Conservation*, *29*(2), 134–153. https://doi.org/10.1017/S0376892902000097

Mannina, G., Rebouças, T. F., Cosenza, A., & Chandran, K. (2019). A plant-wide wastewater treatment plant model for carbon and energy footprint: Model application and scenario analysis. *Journal of Cleaner Production*, *217*, 244–256. https://doi.org/10.1016/j.jclepro.2019.01.255

Mitchell, V. G., Mein, R. G., & McMahon, T. A. (2002). Utilising Stormwater and Wastewater Resources in Urban Areas. *Australasian Journal of Water Resources*, *6*(1), 31–43. https://doi.org/10.1080/13241583.2002.11465208

Nir, T., Arkhangelsky, E., Levitsky, I., & Gitis, V. (2009). Removal of phosphorus from secondary effluents by coagulation and ultrafiltration. *Desalination and Water Treatment*, *8*(1–3), 24–30. https://doi.org/10.5004/dwt.2009.686

Nolan, T., Hands, R. E., & Bustin, S. A. (2006). Quantification of mRNA using real-time RT-PCR. *Nature protocols*, 1(3), 1559-1582.

Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology*, 302(1), 205-217.

Ødegaard, H. (2006). Innovations in wastewater treatment: –the moving bed biofilm process. *Water Science and Technology*, *53*(9), 17–33. https://doi.org/10.2166/wst.2006.284

Ødegaard, H. (2016). A road-map for energy-neutral wastewater treatment plants of the future based on compact technologies (including MBBR). *Frontiers of Environmental Science & Engineering*, *10*(4), 2. https://doi.org/10.1007/s11783-016-0835-0

Oehmen, A., Lemos, P. C., Carvalho, G., Yuan, Z., Keller, J., Blackall, L. L., & Reis, M. A. (2007). Advances in enhanced biological phosphorus removal: from micro to macro scale. *Water research*, 41(11), 2271-2300.

Paruch, A. M., Mæhlum, T., Haarstad, K., Blankenberg, A.-G. B., & Hensel, G. (2016). Performance of Constructed Wetlands Treating Domestic Wastewater in Norway Over a Quarter of a Century – Options for Nutrient Removal and Recycling. In *Natural and Constructed Wetlands* (pp. 41–55). Springer International Publishing. https://doi.org/10.1007/978-3-319-38927-1_3

Promega Corporation. (2009). Protocols & applications guide. Retrieved October 21, 2013, from http://www.promega.com/~/media/files/resources/paguide/letter/paguide_us.pdf?la=en

Rao, N. N., Gómez-García, M. R., & Kornberg, A. (2009). Inorganic polyphosphate: essential for growth and survival. *Annual review of biochemistry*, 78, 605-647.

Rittmann, B. E., Strubler, C. E., & Ruzicka, T. (1982). Anaerobic-Filter Pretreatment Kinetics. *Journal of the Environmental Engineering Division*, *108*(5), 900–912. https://doi.org/10.1061/JEEGAV.0001350

Rodríguez, A., Rodríguez, M., Córdoba, J. J., & Andrade, M. J. (2015). Design of primers and probes for quantitative real-time PCR methods. *PCR primer design*, 31-56.

Rudi, Knut, Inger Andrea Goa, Torgeir Saltnes, Gjermund Sørensen, Inga Leena Angell, and Sondre Eikås. 2019. "Microbial Ecological Processes in MBBR Biofilms for Biological Phosphorus Removal from Wastewater." *Water Science and Technology* 79(8):1467–73. doi: 10.2166/wst.2019.149.

Saltnes, T., Sorensen, G. & Eikas, S. 2017 Biological nutrient removal in a continuous biofilm process. Water Pract. Technol. 12, 797–805.

Salvetti, R., Azzellino, A., Canziani, R., & Bonomo, L. (2006). Effects of temperature on tertiary nitrification in moving-bed biofilm reactors. *Water Research*, *40*(15), 2981–2993. https://doi.org/10.1016/j.watres.2006.05.013

Samarasiri, B. K. T., Rathnasiri, P. G., & Fernando, D. (2019). Development of an Enzymatic Hydrolysis Pretreatment Strategy to Improve Batch Anaerobic Digestion of Wastewater Generated in Desiccated Coconut Processing Plants. *2019 Moratuwa Engineering Research Conference (MERCon)*, 314–319. https://doi.org/10.1109/MERCon.2019.8818847

Schindler, B. K., Weiss, T., Schütze, A., Koslitz, S., Broding, H. C., Bünger, J., & Brüning, T. (2013). Occupational exposure of air crews to tricresyl phosphate isomers and organophosphate flame retardants after fume events. *Archives of Toxicology*, *87*(4), 645–648. https://doi.org/10.1007/s00204-012-0978-0

Schoumans, O. F., Bouraoui, F., Kabbe, C., Oenema, O., & van Dijk, K. C. (2015). Phosphorus management in Europe in a changing world. *AMBIO*, *44*(S2), 180–192. https://doi.org/10.1007/s13280-014-0613-9

Scott, B. C. (1978). Parameterization of Sulfate Removal by Precipitation. *Journal of Applied Meteorology*, *17*(9), 1375–1389. https://doi.org/10.1175/1520-0450(1978)017<1375:POSRBP 2.0.CO;2

Seviour, R. J., Mino, T., & Onuki, M. (2003). The microbiology of biological phosphorus removal in activated sludge systems. *FEMS Microbiology Reviews*, *27*(1), 99–127. https://doi.org/10.1016/S0168-6445(03)00021-4

Seviour, Robert J., Takashi Mino, and Motoharu Onuki. 2003. "The Microbiology of Biological Phosphorus Removal in Activated Sludge Systems." *FEMS Microbiology Reviews* 27(1):99–127. doi: 10.1016/S0168-6445(03)00021-4.

Smith, V. H., & Schindler, D. W. (2009). Eutrophication science: where do we go from here? *Trends in Ecology & Evolution*, *24*(4), 201–207. https://doi.org/10.1016/j.tree.2008.11.009

Sriwiriyarat, T., & Randall, C. W. (2005). Performance of IFAS wastewater treatment processes for biological phosphorus removal. *Water Research*, *39*(16), 3873–3884. https://doi.org/10.1016/j.watres.2005.07.025

Strayer, D. L., & Dudgeon, D. (2010). Freshwater biodiversity conservation: recent progress and future challenges. *Journal of the North American Benthological Society*, *29*(1), 344–358. https://doi.org/10.1899/08-171.1

Tarayre, C., Nguyen, H. T., Brognaux, A., Delepierre, A., De Clercq, L., Charlier, R., ... & Delvigne, F. (2016). Characterisation of phosphate accumulating organisms and techniques for polyphosphate detection: a review. *Sensors*, 16(6), 797.

Taylor, S., Wakem, M., Dijkman, G., Alsarraj, M., & Nguyen, M. (2010). A practical approach to RT-qPCR—publishing data that conform to the MIQE guidelines. *Methods*, 50(4), S1-S5.

Tiwari, P. K., A. K. Misra, and Ezio Venturino. 2017. "The Role of Algae in Agriculture: A Mathematical Study." *Journal of Biological Physics* 43(2):297–314. doi: 10.1007/s10867-017-9453-8.

Villard, D., Nesbø Goa, I. A., Leena Angell, I., Eikaas, S., Saltnes, T., Johansen, W., & Rudi, K. (2023). Spatiotemporal succession of phosphorous accumulating biofilms during the first year of establishment. *Water Science & Technology*, *88*(2), 381–391. https://doi.org/10.2166/wst.2023.214

Vollenweider, R. A. (1975). Input-output models. *Schweizerische Zeitschrift Für Hydrologie*, *37*(1), 53–84. https://doi.org/10.1007/BF02505178

Wang, Y., Gao, H., & F Wells, G. (2021). Integrated omics analyses reveal differential gene expression and potential for cooperation between denitrifying polyphosphate and glycogen

accumulating organisms. *Environmental microbiology*, 23(6), 3274–3293. https://doi.org/10.1111/1462-2920.15486

Wei, Y., Cui, H., Hu, Q., Bai, Y., Qu, K., Sun, J., & Cui, Z. (2022). Eutrophication status assessment in the Laizhou Bay, Bohai Sea: Further evidence for the ecosystem degradation. *Marine Pollution Bulletin*, *181*, 113867. https://doi.org/10.1016/j.marpolbul.2022.113867

Witek-Krowiak, Anna, Katarzyna Gorazda, Daniel Szopa, Krzysztof Trzaska, Konstantinos Moustakas, and Katarzyna Chojnacka. 2022. "Phosphorus Recovery from Wastewater and Bio-Based Waste: An Overview." *Bioengineered* 13(5):13474–506. doi: 10.1080/21655979.2022.2077894.

Yuan, Q., Sparling, R., & Oleszkiewicz, J. (2015). Polyhydroxybutyrate Production from Municipal Wastewater Activated Sludge with Different Carbon Sources. *Air, Soil and Water Research*, *8*, ASWR.S27218. https://doi.org/10.4137/ASWR.S27218

Zaletova, N., Zaletov, S., & Bulychev, I. (2018). Potential of biological phosphorus removal. *MATEC Web of Conferences*, *178*, 09021. https://doi.org/10.1051/matecconf/201817809021

Zhang, Z., Pan, S., Huang, F., Li, X., Shang, J., Lai, J., & Liao, Y. (2017). Nitrogen and Phosphorus Removal by Activated Sludge Process: A Review. *Mini-Reviews in Organic Chemistry*, *14*(2), 99–106. https://doi.org/10.2174/1570193X14666161130151411

Zhong, C., Fu, J., Jiang, T., Zhang, C., & Cao, G. (2018). Polyphosphate metabolic gene expression analyses reveal mechanisms of phosphorus accumulation and release in Microlunatus phosphovorus strain JN459. *FEMS Microbiology Letters*, *365*(6). https://doi.org/10.1093/femsle/fny034

Zhou, Y., Kiely, P. D., Kibbee, R., & Ormeci, B. (2021). Effect of polymeric support material on biofilm development, bacterial population, and wastewater treatment performance in anaerobic fixed-film systems. *Chemosphere*, *264*, 128477. https://doi.org/10.1016/j.chemosphere.2020.128477

Zhu, W., Yang, Z., & Wang, L. (1996). Application of ferrous-hydrogen peroxide for the treatment of H-acid manufacturing process wastewater. *Water Research*, *30*(12), 2949–2954. https://doi.org/10.1016/S0043-1354(96)00197-2.

Basu, C. (Ed.). (2015). PCR primer design. New york: Humana Press.

Appendix

Appendix-A

Table A. *List of 8 Primer pairs for ppk1 gene amplification targeting specific individual sequences of Ca. Accumulibacter. Expected amplicon size is 234bp.*

Anneal to ppk1 sequences	Primer sets Name	Oligo sequence (5' to 3')	Tm	GC%	Oligo names	Expected size	Start to end position
1 and 6	A	CTACCGTCTGGTCAGCGACGA	65.7	61.9%	A-F	234bp	273-506
1	A	CTCTTGTTGAGTACGCGCGG	60.3	60.0%	A-R	234bp	273-506
3	B	CTACCGGCTGGTCAGCGACGA	68.2	66.5%	B-F	234bp	231-464
3, 4, 6, 7	B	CTCTTGTTGAGCACGCGCGG	63.2	65.0%	B-R	234bp	231-464
5	C	CTACCGTCTGGTCAGTGACGA	63.3	57.1%	C-F	234bp	273-506
5	C	CTCTTGTTAAGTACGCGCGG	57.8	55.%	C-R	234bp	273-506
7	E	CTACTCTCTGGTCAGCGACGA	62.7	57.1%	E-F	234bp	284-518
3, 4, 6, 7	E	CTCTTGTTGAGCACGCGCGG	62.7	65.0%	E-R	234bp	284-518
8	F	CTATCGCCTGGTCAGCGACGA	65.7	61.9%	F-F	234bp	284-518
8	F	CTCTTGTTGAGCACCCGCGG	63.4	65.0%	F-R	234bp	284-518
4	G	CTACCGCCTGATCAGCACCGA	65.8	61.9%	G-F	234bp	284-518
3, 4, 6, 7	G	CTCTTGTTGAGCACGCGCGG	63.2	65.0%	G-R	234bp	284-518
2	H	CTACCGACTGGTCAGCGACGA	65.7	61.9%	H-F	234bp	506-273
2	H	CTCTTGTTGAGCACCCGCGG	63.4	65.0%	H-R	234bp	273-506
1,2,3,4,5,6,7,8	All	CTAYYSNCTGRTCAGYRMCGA		53.8%	M-F	234bp	
1,2,3,4,5,6,7,8	All	CTCTTGTTRAGYACBCGCGG		58.8%	M-R	234bp	

Appendix-B

8 Ca. Accumulibacter. Sequences targeting *ppk1 gene* attain from previous Metagenomics study

1. Sequence1.
   ATGACTGCTCAAACCGATATCGCGCAGGCCAGCGCGGCACACTTCCCGACCCAGA
   ATTTTCTGAACCGCGAACTAGGCATCCTGGCCTTCAACCGCCGGGTACTGGCGCA
   GGCTAAAAATCCTCGCACACCGCTGCTCGAACGACTGCGTTTTCAATGCATCGTC
   AGCAGCAATCTCGACGAATTTTTCGAAATCCGGGTGGCCGGCCTCAAGGAACAGC

TCAAGCTGAACTCGGTTGCCGTCACCCCGGACGGTCTGACGGCGCGCGCCGCCTA
CCGTCTGGTCAGCGACGAGGCGCACTCGCTGGTTGAAGAGCAATACGCCCTGCTC
AACGACGAGATCCTGCCCAAGCTGGCTGCCGAGGGCATTCATTTTCTGCGCCGTT
CGGCATGGACCGAAGCGCAACGCGAGTGGATCAGGAACTATTTTTTCCGCGAAGT
GATGCCGGTACTGACGCCGATCGGTCTTGATCCGTCGCACCCCTTTCCGCGCGTA
CTCAACAAGAGCCTGAATTTCGCCGTTGAACTCGATGGCAAGGACGCCTTCGGAC
GCAGTTCCGGCGCCGCCATCGTGCAAGCGCCCCGTGTCCTGCCGCGCGTCATTCG
CCTGCCGCAGGAACTCTGCGACTGCGAGTACGGTTTCGTGTTTTTCTCGTCCGTGC
TGCACGCTTTTGTCGGTGAACTGTTCGCCGGCATGAACGTCCTCGGTTGTTACCAG
TTCCGCGTCACACGCAACAGCGACCTGTTCGTCGATGAAGAAGAGGTAAAGAAC
CTGCGCGCCAAGATTCAGGGCGAACTGCCGCAGCGGCACTTTGGCGATGGAGTG
CGCCTGGAAGTGGCGGACAACTGTTCAGAGGCCATGGCCGGCTTCCTGCTCGCCC
AGTTCAACCTGAGCGAGACCGATCTCTACCGCGTTCCCGGCCCGGTCAATCTGGT
GCGCCTGATGCAGGTTCCGGACTGGGTCGCCCGCGACGACCTAAAATTCAATCCA
TTCACCCCCGGCGTTCCCAAGGCGCTGCAGAAATGCCACTCGATTTTTGACGCCA
TTCGTGCTGGCGACATCCTGCTGCACCACCCTTACCAGAGCTTCACTCCGGTGATC
GAACTGCTCGACCAGGCGGCGACCGACACGCGGGTGGTGGCGATCAAGATGACG
GTTTACCGCACCGGAACCGATTCCGTGCTGATGCAGTCGCTGCTGCGCGCGGCGC
AGAACGGCAAGGAAGTCACCGTCGTCGTCGAACTGATGGCGCGCTTTGACGAAG
AGGCCAACATCGGCTGGGCAACCAAGCTGGAA

2. Sequence 2

ATGAATATTCGTAATGATATCGCTGCCGGCAGCGCCGCCGTTTTCCCTGCCGAAA
ACTTTCTTAACCGCGAACTGGGCATCCTGGCGTTCAACCGCCGCGTGCTTTCGCA
ATGCAGGAACCCGCGCACGCCGCTGCTCGAGCGCCTGCGTTTCATTTGCATCGTC
AGCAGCAATCTCGATGAATTCTTCGAGATCCGCGTCGCCGGCCTCAAGGAACAGG
TCAAGCTGAACTCGGTCGGGGTTACGGCCGACGGCCTGACCGCGCGCGCCGCCTA
CCGACTGGTCAGCGACGAAGCCCACGCGCTGGTCGAGGAGCAGTATTCGCTGCTC
AACAACGAGATCCTGCCGAATCTCGCCGCCGAAGGCATCCGGTTCATCCGCCGCA
GCGAATGGAGCGAAGCGCAGCTCGAATGGATCAAGGGCTACTTTTTGCGCGAGG
TCATGCCGGTTCTGACGCCGATTGGTCTCGATCCGTCGCACCCCTTCCCGCGAGTG
CTCAACAAGAGTCTGAACTTCGCCGTCGAACTCGAAGGCAAGGACGCTTTCGGGC
GCAGTTCCGGCGCCGCCATCGTGCAGGCGCCGCGCGTCCTGCCGCGCGTCATTCG
CCTGCCCGAGGAACTTTGCGATTGCGAATACGGCTTTGTGTTTTTCTCATCGGTAT

TGCACGCCTTCGTCGGCGAGCTGTTCGCCGGAATGAACGTCCTCGGCTGTTATCA
GTTCCGCGTTACGCGTAACAGCGATCTCTTCGTCGACGAAGAAGAAGTCAAGAAC
CTGCGCACCAGCATCCAGGGCGAGTTGCCGCAGCGGCACTTCGGCGACGGTGTCC
GGCTGGAAGTGGCGGACAACTGCTCGGAAGTCATGGCCGAGTTTCTTCTCGCGCA
ATTCAACCTCAGCGCGACCGATCTCTATCGTGTCGCCGGGCCGGTGAACCTGGTG
CGCCTGATGCAGGTACCCGACTGGGTGCTGCGCAACGATCTGAAGTTCCAGCCTT
TCCATCCGGGAACGCCCAAGGCGCTGCAAAAATGCCACAACATCTTTGATGCCAT
ACGCGGTAACGACATTCTCCTTCACCACCCCTATCAGAGCTTCACTCCGGTCATCG
AACTGATCGACCAGGCGGCTACCGATGCTCAGGTCGTGGCGATCAAGATGACCGT
TTATCGTACCGGGACCGACTCGGTGCTGATGCAGTCGCTGCTGCGCGCCGCGCAG
AACGGCAAGGAGGTCACCGTCGTCGTCGAACTGATGGCGCGCTTCGACGAAGAA
GCGAACATCGGCTGGGCAACCAAGCTGGAGGAAGTCGGCGCCCATGTCGTCTAT
GGCGTCGTCGGCTACAAG

3. Sequence3.

TTCCCGACGCAGAATTTTCTTAACCGCGAACTGGGCATCCTGGCCTTCAACCGCC
GGGTGCTGGCGCAGGCAAAAAACCCGCGCACACCGCTGCTTGAACGGCTGCGTTT
TCAATGCATCGTCAGCAGCAATCTCGACGAATTTTTCGAAATCCGGGTGGCCGGC
CTCAAGGAACAGCTCAAGCTGAACTCGGTGGCCGTGACCCCGGATGGATTGACG
GCGCGTGCCGCCTACCGGCTGGTCAGCGACGAGGCGCACTCGCTGGTAGAAGAG
CAATACGCCCTGTTCAACGACGAAATCCTGCCTAAACTGGCCGACGAGGGCATTC
ACTTTCTGCGTCGCTCGGCATGGACCGACGCACAGCGCGAGTGGATCAGCGACTA
TTTTTTCCGCGAAGTGATGCCGGTGCTGACCCCGATCGGTCTTGATCCGTCGCACC
CCTTTCCGCGCGTGCTCAACAAGAGCCTGAATTTCGCCGTTGAACTCGATGGCAA
GGACGCCTTCGGACGCAGTTCCGGCGCGGCCATCGTGCAGGCGCCCCGCGTCTTG
CCGCGCGTCATTCGCCTGCCGAAGGAACTCTGCGACTGCGAATACGGATTCGTGT
TTTTCTCGTCGGTGCTGCACGCTTTTGTCGGTGAACTTTTTGCCGGCATGAACGTC
CTCGGCTGTTACCAGTTCCGCGTCACGCGCAACAGCGACCTCTTCGTCGATGAGG
AGGAGGTCAAGAACCTGCGCGCCAAGATTCAGGGCGAACTGCCGCAACGGCACT
TCGGCGACGGCGTGCGCCTGGAAGTGGCGGACAACTGTTCGGAGGCAATGGCCG
GTTTCCTGCTCGCCCAGTTCAACCTGAGCGAAACCGATCTCTACCGCGTTCCCGGC
CCGGTCAACCTGGTTCGTCTGATGCAGGTTCCGGACTGGGTTGCCCGCGACGACC
TCAAGTTCATTCCATTCACCCCCGGCGTCCCCAAGGCGCTGCATAAATGCAACAC
GATTTTTGACGCCATCCGCGCCGGGGACATCCTGTTGCACCACCCCTACCAGAGC

TTCACACCGGTAATCGAAGTGCTTGATCAGGCGGCAATCGACACGCGGGTGGTGG
CCATCAAAATGACGGTTTACCGCACCGGGACCGATTCCGTGCTGATGCAGTCGCT
GCTGCGCGCGGCACAGAACGGCAAGGAAGTCACCGTCGTCGTCGAACTGATGGC
ACGCTTTGACGAAGAGGCCAACATCGGCTGGGCAACCAAGCTGGAAGAGGTCGG
CGCGCACGTCGTCTATGGCGTCGTCGGCTACAAGACGCACGCCAAGATGCTGATG
ATC

4. Sequence 4.

ATGACGACCGTAACACTGACCACCGAAACCCTTCCCCACGAGACGCCCACCGCTT
TTCCGCCGGACTACTTTCTGAATCGCGAAATCGGCCTGCTGGCCTTCAACCGGCG
CGTCCTCTGGCAGGCGAAGAACCCGCGCACGCCGCTGCTCGAACGTCTGCGCTTT
CTGTGCATCGTCAGCAGCAACCTCGACGAGTTTTTTGAGATCCGCGTCGCCGGCC
TCAAGGAGCAGGTCAAGCTCAAGTCGGTTGTCGCCACCACCGACGGGCTGACGG
CGCGCGCCGCCTACCGCCTGATCAGCACCGAGGCGCATGCCATCGTCGAGGAGC
AGTACGCGCTGTTCAATAACGAGCTGCTGCCCAAGCTTGCCGCCGAAGGCATCAA
GTTCATCCGCCGCAGCGAGTGGACCGAGGCCCAGCTCGAATGGATACGCGGCTAT
TTTTTCCGCGAAGTGATGCCGGTACTGACGCCCATCGGTCTGGATCCCTCGCACCC
TTTCCCGCGCGTGCTCAACAAGAGCCTCAATTTCGCCGTCGAACTTGAGGGTAAG
GACGCTTTCGGCCGCAGTTCGGGGGCCGCGATCGTCCAGGCGCCGCGCGTCCTGC
CGCGCATCATCCGCCTGCCGCGCGAGCTGTGCGCCTGCGAATACGGCTTCGTCTT
CTTCTCGTCGGTCCTCCATGCCTTTGTCAGCGAGCTCTTCGCCGGGATGAATGTGC
TCGGCTGTTACCAGTTCCGCGTCACGCGCAACAGCGACCTCTTCGTCGACGAAGA
AGAAGTCAAGAACCTGCGCACCAAAATTCAGGGCGAACTGCCGCAGCGGCACTT
TGGCGACGGCGTTCGCCTCGAAGTTGCCGATAATTGTTCGCAGACCATGGCCGAT
TTTCTGCTGGCGCAATTCGCCTTACGCGACACCGACCTCTATCGCGTCGCCGGGCC
GGTCAACCTCGTTCGCCTGATGCAGGTGCCCGATTGGGTGCTGCGCAACGACCTC
AAATTTCCGTCCTTCCAGCCGGGAATACCGAAAGCCTTGCAGAAAGGACACGCG
ATTTTCGACGCCGTCGCCAAGGGCGACATCCTGCTCCATCACCCCTACCAGAGCT
TCGTTCCGATCATCGATCTGCTCGACCAGGCGGCAAGCGACGCGCAGGTGGTAGC
GATAAAAATGACCGTTTACCGCACGGGAACCGACTCGGTACTGATGCAATCGCTG
CTGCGCGCCGCGCAGAACGGCAAGGAAGTGACGGTCGTCGTCGAACTGATGGCG
CGCTTTGACGAAGAGGCCAATATCGGCTGGGCGACCAAGCTCGAGGAAGTCGGA
GCGCACGTCGTTTATGGCGTCGTCGGCTATAAAACGCACGCCAAGATGCTGATGA
TCGTTCGCCGCGAGCAGACACAACTGCGTCGCTACGTCCACCTGGGAACCGGCAA

CTACCATCCCCGGACGGCGCGCCTGTACTCCGATTTCGGGCTGCTCACCTGCAAC
GAAGAAATCGGCGCCGACGTCAAC

5. Sequence 5

ATGACTTATCCTCCTGATATTGCGCACGCCAGCGGGACCCGCTTCCCTTCTGAAA
ACTTCCTGAACCGCGAACTGGGCATCCTGGCATTCAATCGGCGGGTGCTTTCACA
GGCCAGAAATCCGCTCGTCCCATTGCTTGAACGACTGCGATTTCTGTGCATCGTC
AGCAGCAATCTTGACGAATTCTTCGAAATCCGTGTGGCTGGCCTCAAGGAGCAGG
TCAAGCTGAATTCCGTAGCCGTTACCACGGACGGATTGACGGCGCGTGCGGCCTA
CCGTCTGGTCAGTGACGAAACACATTCGCTGGTTGAGGAAAAGTACGCCCTGCTC
AACGACGAGATCCTTCCCAAACTGGCCGAGGAAGGCATTGTCTTCCTGCGGCGCT
CGGCATGGACCGTCGAACAACGCGAGTGGATCAAGGACTATTTTTTCCGCGAAGT
CATGCCGGTCCTGACCCCGATCGGTCTTGATCCATCGCATCCCTTCCCGCGCGTAC
TTAACAAGAGTCTCAATTTTGCCGTCGAACTCAAGGGCAAGGATGCCTTCGGCCG
CAGCTCCGGAGCAGCAATCGTCCAGGCCCCCCGAGTCCTGCCGCGCGTCATCCGG
CTACCACGTGAACTGTGCGATTGCGCGTACGGCTTTGTGTTCTTCTCGTCGGTCCT
GCATGCCTTCGTCGGTGAGCTGTTTGCCGGCATGAACGTCCTGGGCTGTTACCAG
TTTCGTGTCACGCGCAACAGCAATCTCTTCGTCGATGAAGAAGAAGTGAAAAACC
TGCGCGCAAAGATTCAGGGCGAACTGCCGCAGCGGCACTTCGGCGATGGCGTGC
GGCTGGAGGTTGCCGACATCTGCTCGGATGCCATGGCCGAGTTTCTCCTCTCCCA
GTTCAATCTGGGAGCTACCGATCTTTACCGTGTTCCGGGACCGGTCAATCTGGTG
CGTCTGATGCAGGTGCCGGACTGGGTGGAACGAACCGACCTCAAGTTCACGCCGT
TCAGGCCGGGTTTCCCGAAATCGTTGCACAAAGGGAATTCCATCTTTGACGATAT
CCGCGCCGGCGATACCCTGTTGCACCATCCCTACCAGAGTTTCTCACCGGTAATTT
CGCTGCTCGAACTGGCCGCAAGCGACCCGCAGGTTGCAGCCATCAAGATGACGG
TGTACCGCACCGGGACCGATTCCGTATTGATGAAGTCATTGATTCTTGCTGCGCA
AAACGGCAAGGAAGTCACGGTGATCCTCGAACTGATGGCGCGTTTTGATGAGGA
AGCGAATATCGGTTGGGCCACCAAGCTGGAACAGGTCGGCGCCCATGTCATTTAT
GGCGTCGTCGGCTATAAAACCCACGCCAAGATGCTGTTGATTGTTCGCCGGGAAG
AAGGGAAGACCGGCAGCATGTTGCGTCGCTACGTGCATCTGGGAACCGGAAATT
ACCACCCGAAGACCGCCCGTCTCTATTCAGACTTCGGACTGCTTACCTGCAATGA
AGAAATCGGTGACGATGTCAATGAAGTATTCAAGCAGCTCACCGGACTCGGGCT
GGCCCAATCTCTGAAGCATCTCTGGCAGGCTCCTTTTTCCTTGCAGCAGAACCTGC
TGTCCGCAATTCAGCAAGAGGCCGAGGCCGCGCGTTCCGGAGGGCGTTCGCGCAT

CATCGCCAAGATGAATTCGCTGCTCGAACCAGGGATCATCAGCGCGCTTTATGAG
GCATCGCAGGCCGGAGTGAAAATCGACCTGGTCGTTCGTGGCGTTTGCGCTTTGC
GGCCGGGAATCAAGGGACTGTCGGAGAATATTCGGGTTCGTTCAATCATCGGCCG
GCTGCTCGAACACCATCGGGTATTCTATTTCTACGCCAGCGGCGCGGAGAAGGTT
TACCTGTCAAGCGCCGACTGGATGGAGCGAAACTTCTTTCGCCGCATCGAACTCG
CCTTCCCGATACTTGACAAGAAGCTCAAGCGTCGCGTGATTACCGAAGGATTGCA
GATCTATTTCGCCGACAATCAGCAGGCCTGGGAAATGGACGCCAACGGCAGCTAT
CGACTGCGCCGCTGCGCGCGCGGCAAACCGCGCTGCGCGCAAAGCGAGTTAATC
GAACTGCTCAAACCCAGAGAACCTTCGTGA

6. Sequence 6

ATGACTACTCGCACCGATATCCCGCCGGCCAGTGCCGCGCATTTCCCCTCTGAAA
ACTTCCTGAACCGGGAGCTGGGCATCCTGGCGTTCAACCGCCGCGTCCTGTCGCA
GGCCAAAAACCCGCTCACACCGCTGCTCGAACGTCTACGTTTTCTCTGCATTGTCA
GCAGCAATCTCGACGAATTTTTCGAAATCCGCGTCGCCGGCCTCAAGGAACAGCT
CAAGCTCAACTCGGTAGCGGTCACCACCGACGGCCTGACCGCACGTGCCGCCTAC
CGTCTGGTCAGCGACGAAGCTCACTCGCTGGTTGAAGAGCAATACGCCTTGCTCA
ACGATGAAATCCTGCCCAGGTTTGCCGAGGAGGGCATTCACTTTCTGCGGCGCTC
GGCATGGACTACCGGGCAACGCGAATGGATCAAGGACTATTTTTTCCGCGAGGTC
ATGCCGGTGCTGACCCCGATCGGTCTTGATCCCTCGCACCCCTTCCCGCGCGTGCT
CAACAAGAGCCTTAATTTCGCTGTTGAACTCGATGGCAAGGACGCCTTCGGACGC
AGCTCAGGCGCCGCCATCGTGCAAGCCCCGCGGGTTCTGCCGCGCGTCATTCGCC
TGCCGCAGGAACTGTGCGATTGCGAGTACGGTTTCGTATTTTTCTCCTCGGTACTG
CACGCGTTTGTCGGCGAACTGTTCTCGGGCATGAACGTCCTCGGTTGTTACCAGTT
CCGCGTCACACGCAATAGCGACCTGTTCGTCGATGAAGAAGAGGTGAAGAACCT
GCGCACCAAAATTCAGGGCGAACTGCCGCAACGGCATTTTGGCGACGGCGTGCG
GCTGGAGGTTGCGGACAATTGTTCGGAGGTCATGGCCGATTTCCTGCTCTCCCAG
TTCAACCTGGGGGAGACCGATCTCTATCGGGTTGCCGGCCCGGTTAACCTTGTTC
GCCTGATGCAAGTACCCGATTGGGTCGCCCGCGACGATCTCAAGTTCGCTCCTTT
CCGGCCCGGATTCCCCAAGTCGCTGCACAAGGGGCAGATCATTTTTGCCGATATT
CGCACCAACGATGTCCTGCTGCATCATCCCTACCAGAGTTTTGCGCCGGTCATCTC
GCTGATCGACCAGGCAACGACAGACCCGCAAGTCGTGGCGATCAAGATGACGGT
TTACCGTACCGGAACCGACTCCGTGCTGATGCAATCGCTGCTGCACGCTGCGCAG

AACGGTAAAGAGGTCACCGTGGTCGTCGAGTTGATGGCGCGCTTTGACGAAGAG
GCCAACATCGGCTGGGCAACCAAGCTGGAAGAAGTAGGAGCCCACGTCGTCTAT
GGCGTCGTCGGCTACAAGACGCACGCCAAGATGCTGATGATCGTTCGTCGTGAAG
AAGGAAAAACGGGCAGCCTGCTGCGGCGCTATGTGCATCTGGGAACCGGTAACT
ACCATCCGAAAACGGCACGTCTGTACTCCGATTTCGGGCTGCTCACCAGCAATGA
AGAGATCTGTTCCGATGTCAATGAGGTGTTCAAGCAGCTTACTGGGCTTGGTCGC
GCCCAGACCCTGAAGCATCTCTGGCAGGCACCCTTCTCCTTGCAGCAAAACATCC
TTGCCGCCATCAAGGTAGAGGCCGCTGCCGCAGTTACCGGCCAGCGCTCGCGCAT
CATTGCGAAATTGAATTCACTGCTTGAGCCCGAAACCATCAGCGCGCTGTATGAG
GCTTCGCAAGCCGGCGTAAAGATTGACCTGATCGTGCGCGGCGTTTGCGCGTTAC
GACCGGGAATCAAGGGCCTGTCGGAGAACATTCGCGTTCGCTCGATTATCGGCCG
CCTGCTTGAACATCATCGCGTGTTCTATTTTTACGCGGGCGGCGAGGAGAAGGTT
TATCTCTCCAGTGCCGACTGGATGGAGCGCAACTTCTTCCGCCGCATCGAGCTCG
CCTTCCCGATACTCGACAAGAAGCTTAAACGCCGGGTCATCGCCGAAGGCCTGCA
AATCTATCTGACCGACAACCAGCAGGCCTGGGAGATGGACGCCAACGGCAGTTA
CCATTTGCGCCGTGGTTCGCCCGACAAACCACGCTGCGCGCAGAGCGAACTGATT
GAACTGCTCAAACCCTGA

7. Sequence 7

ATGAACACCGCCGCCACACCCACCGAAACCCCTCCCTCCAAGGCCACTGACGGCT
TCCCGTCGGACTACTTTCAGAATCGCGAGATGGGGCTGCTTGCTTTCAATCGGCG
CGTGTTGTGGCAGGCCAAGAACCACAAAACCCCACTGCTGGAACGTTTGCGCTTC
TTGTGCATCGTCAGCAGCAACCTCGACGAGTTCTTCGAAATCCGGGTTGCCGGCG
TCAAGGAGCAACTCAAACTGGGATCGGTAGCGCTCACCACCGACGGCAAGACGG
CACGCGAAGTCTACTCTCTGGTCAGCGACGAAACGCACGCCCTGGTCAAGGAAC
AGTACGCACTGTTTAATGAAGAGATCCTGCCGCAGCTGGCGAGTGAAGGCATTCG
CTTCCTCAAGCGCGGCGACTGGAGCCCCGCCCAGCGCGAATGGATCAGCGCCTTC
TTCTTCCGCGAAGTGATGCCGGTGATCACCCCGATCGGCCTCGACCCCTCGCACC
CCTTCCCGCGCGTGCTCAACAAGAGCCTCAACTTTGCCGTCGAACTCGAAGGTCG
TGATGCTTTCGGGCGCAGCTCGAACGCGGCCATCGTGCAGGCGCCACGTGTTCTG
CCACGCGTCATCCGGCTGCCACGCGAACTTGGTGACAGCGAATACTGCTTCATCT
TCCTGTCGTCGATCCTGCATGAGTTCGTCCACGAGCTGTTCGCCGGCATGAAAGT
CCTTGGCTGCTACCAGTTCCGCGTCACGCGCAACAGCAACCTGTTCGTCGACGAA
GAAGCGGTCAAGAATCTGCGCGCCAAGATCCAGGGCGAATTGCCGCAGCGACAC

TTTGGTAATGCGGTTCGCCTCGAGGTCGCCAACAACTGCTCGGAAGCAATGGCCG
AGTTCCTGCTCGGCCACTTCGACCTCACCGAGCGCGACCTCTTCCGCGTCGCCGG
GCCGGTCAATCTGGTCCGCCTGATGCAGGTTCCCGACTGGGTACTGCGCGACAAC
CTCAAGTTCCAACCCTTCAAACCCGGCACACCGAAAGTTCTGCAGAAGAATGCCA
ACGTGTTCGACGCCATCCGCGGCGGCGACATCCTGCTGCACCATCCCTACCAGAG
CTTCAATCCGGTCATCGAATTACTCGATCAGTCAGCCGCCGACCCGCAGGTGGTG
GCCATCAAGATGACCGTTTACCGTACCGGCACTGACTCGGTGCTGATGCAGTCGC
TGTTGCGCGCGGCACAGAACGGCAAGGAAGTCACGGTGGTCGTCGAGCTGATGG
CGCGCTTTGACGAGGAAGCGAACATCGGCTGGGCGACCAAGCTTGAAGAAGTCG
GCGCGCACGTCATTTACGGCGTTGTTGGCTACAAGACGCACGCCAAGATGCTGAT
GATCGTCCGCCGCGAGGAAAGCAAGGCCGGCGGCAGCCTCCTGCGCCGCTACGT
GCACCTCGGTACCGGCAACTACCACCCCAAGACGGCCCGCCTGTACTCCGATTTT
GGTCTGCTCACCTGCAACGAAGATATCGGTGCCGATGCCAACGAGGTGTTCAAGC
AGCTTACCGGCCTCGGCCGCGCGCAGACCCTGAAGCACCTCTGGCAAGCGCCTTT
CTCGCTGCAACCGAATGTCGTCGCCGGGATCCGTGCCGAAGCCGAGGCAGCACG
CACCGGCAAGCGTTCACGAATCATCGCCAAGATGAATTCGCTGCTCGAACCGGAG
ACCATCGCTGCGCTCTACGAAGCCTCGCAGGCCGGCGTCAAGATCGACCTCATCG
TCCGCGGCGTCTGCGGCTTGCGTGCCGGCGTCAAGGACCTGTCCGAGAACATTAA
CGTCTGGTCGATCATCGGTCGGCAACTCGAACATCATCGGGTGTTCTATTTCTTCG
CCGACGGCGAAGAAAAAGTCTATCTATCGAGCGCCGACTGGATGGACCGCAATT
TCTTCCGCCGCATCGAACTCGCCTTTCCGATCCTCGACCGCAAGCTGAAGCGCCG
GGTAATCACCGAGGGGCTGCAGATCTACCTCACCGACAATACCCTGGCCTGGGAA
CTCGGAGCGGATGGCAGCTATCATCAGCGGCGCGGCTCACGGACCAGCCCGCATT
CGTCGCAAGCCGAACTGATCGAACTGCTGAAGGACTGA

8. Sequence 8

ATGACTACAGAAGCCCCGACCGCCGATACCCCTGCCCAGCAGTCCCCGATTACCT
TCCCGCCGGAATACTTCCAGAATCGTGAGATGAGCCTGCTCGCGTTCAATCGGCG
TGTTCTCTGGCAGGCCAAGAACCCGCGGACTCCGCTGCTCGAACGTCTGCGCTTC
CTGTGCATCGTCAGCAGCAATCTCGACGAGTTCTTCGAGATTCGGGAGGCGGGGA
TCAAGGAACAGCTGAAGCTCGGATCGGTGGCGATCACCACCGACGGCAAGACGG
CGCGCGAGGTCTATCGCCTGGTCAGCGACGAAACGCATGCGATCGTCAAGGAGC
AGTACGCGCTCCTCAACGACGAAATCCTGCCCCTGCTCGCGGCCGAAGGCATCCG
TTTCGTCAAGCGCGCAGAATGGAACGTCGCGCAGCGTGAATGGATCAGCGACTTC

TTCTTCCGCGAAGTGATGCCGGTAATCACGCCGATCGGTCTCGACCCCTCGCATC
CCTTCCCGCGGGTGCTCAACAAGAGCCTCAACTTCGCCGTCGAACTCGAAGGTCG
CGATGCCTTTGGCCGCAGTTCCGGGGCAGCCATCGTCCAGGCACCGCGCGTCCTG
CCGCGGGTCATCCGGCTGCCGCGTGAACTCGGGGACGCCGAGTACACCTTCGTCT
TCCTGTCGTCGATCCTGCACGAGTTCGTCCACGAGCTCTTCGCGGGCATGAAGGT
CCTCGGCTGCTACCAGTTCCGGGTCACGCGCAACAGCGACCTGTTCGTCGACGAG
GAGGAAGTCAAGAACCTGCGCGCCAAGATCCAGGGCGAATTGCCACAACGTCAC
TTTGGCGATGCCGTCCGGCTCGAAGTCGCCAACAGCTGCTCTGAGGCGATGACGC
AGTTCCTCCTCGGACAGTTCAATCTTACCGAAAGCGACCTGTTTCGAGTCGCCGG
TCCGGTCAATCTGGTGCGCCTGATGCAGGTGCCCGACTGGGTGGTGCGCAACGAC
CTGAAGTTCCCCCCCTTCACGCCGGGCATACCGAAGGCCTTGCAGAAATGCCACA
GCGTTTTCGACAGCATCCGCGGCGGCGACATCCTGCTGCACCACCCCTACCAGAG
CTTCAATCCGGTAATCGAGTTGCTCGAACAGGCGGCAACTGACCCGCAGGTGGTG
GCGATCAAGATGACCGTCTACCGCACCGGCACCGACTCCGTCCTGATGCAATCGC
TGCTGCGCGCGGCGCAGAACGGCAAGGAAGTGACGGTGGTCGTCGAACTGATGG
CCCGCTTCGACGAGGAAGCGAACATCGGCTGGGCGACCAAACTCGAGGAAGTCG
GCGCGCACGTCGTCTATGGGGTCGTCGGTTACAAGACGCACGCCAAGATGCTGAT
GATCGTGCGGCGCGAAGAAGGCAAGGCAGGCAGCAGCCTCCTGCGCCGCTACGT
GCATCTGGGCACCGGGAACTACCATCCGAAAACGGCCCGTCTGTACTCGGACTTC
GGTCTGCTGACCTGCAACGAAGAGATCGGCGCCGATACCAACGAGGTCTTCAAG
CAGCTCACCGGTCTGGGTCGTGCGCAGACGCTGACGCACTTGTGGCAGGCACCGT
TCACACTGCAGCCGAACGTCGTCGCCGGCATCCGCGCGGAAGCCGATGCGGCGC
GCTCAGGCAAGCGTGCGCGCATCATCGCCAAGATGAATTCGCTACTCGAACCGGA
GACGATCGCGGCGCTTTATGAAGCGTCGCAGGCCGGCGTGAAGATCGACCTGATC
GTGCGCGGCGTCTGTGGCCTGCGTGCCGGTGTCAAGGACCTGTCCGACAACATCA
GTGTCCGCTCGATCATCGGCCGGCAACTCGAGCATCACCGGGTCTTCTATTTCTAC
GCCGGTGGCGAGGAGAAGGTCTATCTGTCGAGCGCTGACTGGATGGAGCGTAAC
TTCTTCCGGCGCATCGAACTGGCTTTTCCGGTGCTCGACCGCAAACTGAAGCGTC
GGGTGATCGCCGAGGGGCTGCAGATCTATCTCGCCGACAACACCCTGGCCTGGGA
ACTCGGACCGGATGGGACCTACCACCAGCGGCGCGGCTCACGCACCAGCCGGCA
CGCGTCGCAGAGCGAGTTGATCGAACTGCTCCAGTCCTGA

Appendix-C

The 8 *ppk1* DNA sequences translated to protein sequences of *Ca.Accumulibacter* by CLC. Workbench.

>Ca.Accumulibacter1

MTAQTDIAQASAAHFPTQNFLNRELGILAFNRRVLAQAKNPRTPLLERLRFQCIVSSN
LEFFEIRVAGLKEQLKLNSVAVTPDGLTARAAYRLVSDEAHSLVEEQYALLNDEILPK
LAAEGIHFLRRSAWTEAQREWIRNYFFREVMPVLTPIGLDPSHPFPRVLNKSLNFAVE
LDGKDAFGRSSGAAIVQAPRVLPRVIRLPQELCDCEYGFVFFSSVLHAFVGELFAGMN
VLGCYQFRVTRNSDLFVDEEEVKNLRAKIQGELPQRHFGDGVRLEVADNCSEAMAG
FLLAQFNLSETDLYRVPGPVNLVRLMQVPDWVARDDLKFNPFTPGVPKALQKCHSIF
DAIRAGDILLHHPYQSFTPVIELLDQAATDTRVVAIKMTVYRTGTDSVLMQSLLRAA
QNGKEVTVVVELMARFDEEANIGWATKLE

>Ca.Accumulibacter2

MNIRNDIAAGSAAVFPAENFLNRELGILAFNRRVLSQCRNPRTPLLERLRFICIVSSNL
DEFFEIRVAGLKEQVKLNSVGVTADGLTARAAYRLVSDEAHALVEEQYSLLNNEILP
NLAAEGIRFIRRSEWSEAQLEWIKGYFLREVMPVLTPIGLDPSHPFPRVLNKSLNFAVE
LEGKDAFGRSSGAAIVQAPRVLPRVIRLPEELCDCEYGFVFFSSVLHAFVGELFAGMN
VLGCYQFRVTRNSDLFVDEEEVKNLRTSIQGELPQRHFGDGVRLEVADNCSEVMAEF
LLAQFNLSATDLYRVAGPVNLVRLMQVPDWVLRNDLKFQPFHPGTPKALQKCHNIF
DAIRGNDILLHHPYQSFTPVIELIDQAATDAQVVAIKMTVYRTGTDSVLMQSLLRAAQ
NGKEVTVVVELMARFDEEANIGWATKLEEVGAHVVYGVVGYK

>Ca.Accumulibacter3

FPTQNFLNRELGILAFNRRVLAQAKNPRTPLLERLRFQCIVSSNLDEFFEIRVAGLKEQ
LKLNSVAVTPDGLTARAAYRLVSDEAHSLVEEQYALFNDEILPKLADEGIHFLRRSA
WTDAQREWISDYFFREVMPVLTPIGLDPSHPFPRVLNKSLNFAVELDGKDAFGRSSG
AAIVQAPRVLPRVIRLPKELCDCEYGFVFFSSVLHAFVGELFAGMNVLGCYQFRVTR
NSDLFVDEEEVKNLRAKIQGELPQRHFGDGVRLEVADNCSEAMAGFLLAQFNLSETD
LYRVPGPVNLVRLMQVPDWVARDDLKFIPFTPGVPKALHKCNTIFDAIRAGDILLHHP
YQSFTPVIEVLQAAIDTRVVAIKMTVYRTGTDSVLMQSLLRAAQNGKEVTVVVELM
ARFDEEANIGWATKLEEVGAHVVYGVVGYKTHAKMLMI

>Ca.Accumulibacter4

MTTVTLTTETLPHETPTAFPPDYFLNREIGLLAFNRRVLWQAKNPRTPLLERLRFLCIV
SSNLDEFFEIRVAGLKEQVKLKSVVATTDGLTARAAYRLISTEAHAIVEEQYALFNNE
LLPKLAAEGIKFIRRSEWTEAQLEWIRGYFFREVMPVLTPIGLDPSHPFPRVLNKSLNF
AVELEGKDAFGRSSGAAIVQAPRVLPRIIRLPRELCACEYGFVFFSSVLHAFVSELFAG
MNVLGCYQFRVTRNSDLFVDEEEVKNLRTKIQGELPQRHFGDGVRLEVADNCSQTM
ADFLLAQFALRDTDLYRVAGPVNLVRLMQVPDWVLRNDLKFPSFQPGIPKALQKGH
AIFDAVAKGDILLHHPYQSFVPIIDLLDQAASDAQVVAIKMTVYRTGTDSVLMQSLLR
AAQNGKEVTVVVELMARFDEEANIGWATKLEEVGAHVVYGVVGYKTHAKMLMIV
RREQTQLRRYVHLGTGNYHPRTARLYSDFGLLTCNEEIGADVN

>Ca.Accumulibacter5

MTYPPDIAHASGTRFPSENFLNRELGILAFNRRVLSQARNPLVPLLERLRFLCIVSSNL
DEFFEIRVAGLKEQVKLNSVAVTTDGLTARAAYRLVSDETHSLVEEKYALLNDEILP
KLAEEGIVFLRRSAWTVEQREWIKDYFFREVMPVLTPIGLDPSHPFPRVLNKSLNFAV
ELKGKDAFGRSSGAAIVQAPRVLPRVIRLPRELCDCAYGFVFFSSVLHAFVGELFAGM
NVLGCYQFRVTRNSNLFVDEEEVKNLRAKIQGELPQRHFGDGVRLEVADICSDAMA
EFLLSQFNLGATDLYRVPGPVNLVRLMQVPDWVERTDLKFTPFRPGFPKSLHKGNSIF
DDIRAGDTLLHHPYQSFSPVISLLELAASDPQVAAIKMTVYRTGTDSVLMKSLILAAQ
NGKEVTVILELMARFDEEANIGWATKLEQVGAHVIYGVVGYKTHAKMLLIVRREEG
KTGSMLRRYVHGTGNYHPKTARLYSDFGLLTCNEEIGDDVNEVFKQLTGLGLAQSL
KHLWQAPFSLQQNLLSAIQQEAEAARSGGRSRIIAKMNSLLEPGIISALYEASQAGVKI
DLVVRGVCALRPGIKGLSENIRVRSIIGRLLEHHRVFYFYASGAEKVYLSSADWMER
NFFRRIELAFPILDKKLKRRVITEGLQIYFADNQQAWEMDANGSYRLRRCARGKPRC
AQSELIELLKPREPS

>Ca.Accumulibacter6

MTTRTDIPPASAAHFPSENFLNRELGILAFNRRVLSQAKNPLTPLLERLRFLCIVSSNLD
EFFEIRVAGLKEQLKLNSVAVTTDGLTARAAYRLVSDEAHSLVEEQYALLNDEILPRF
AEEGIHFLRRSAWTTGQREWIKDYFFREVMPVLTPIGLDPSHPFPRVLNKSLNFAVEL
DGKDAFGRSSGAAIVQAPRVLPRVIRLPQELCDCEYGFVFFSSVLHAFVGELFSGMNV
LGCYQFRVTRNSDLFVDEEEVKNLRTKIQGELPQRHFGDGVRLEVADNCSEVMADF
LLSQFNLGETDLYRVAGPVNLVRLMQVPDWVARDDLKFAPFRPGFPKSLHKGQIIFA
DIRTNDVLLHHPYQSFAPVISLIDQATTDPQVVAIKMTVYRTGTDSVLMQSLLHAAQ

NGKEVTVVVELMARFDEEANIGWATKLEEVGAHVVYGVVGYKTHAKMLMIVRREE
GKTGSLLRRYVHLGTGNYHPKTARLYSDFGLLTSNEEICSDVNEVFKQLTGLGRAQT
LKHLWQAPFSLQQNILAAIKVEAAAAAVTGQRSRIIAKLNSLLEPETISALYEASQAGV
KIDLIVRGVCALRPGIKGLSENIRVRSIIGRLLEHHRVFYFYAGGEEKVYLSSADWME
RNFFRRIELAFPILDKKLKRRVIAEGLQIYLTDNQQAWEMDANGSYHLRRGSPDKPRC
AQSELIELLKP

>Ca.Accumulibacter7

MNTAATPTETPPSKATDGFPSDYFQNREMGLLAFNRRVLWQAKNHKTPLLERLRFLC
IVSSNLDEFFEIRVAGVKEQLKLGSVALTTDGKTAREVYSLVSDETHALVKEQYALFN
EEILPQLASEGIRFLKRGDWSPAQREWISAFFFREVMPVITPIGLDPSHPFPRVLNKSLN
FAVELEGRDAFGRSSNAAIVQAPRVLPRVIRLPRELGDSEYCFIFLSSILHEFVHELFAG
MKVLGCYQFRVTRNSNLFVDEEAVKNLRAKIQGELPQRHFGNAVRLEVANNCSEAM
AEFLLGHFDLTERDLFRVAGPVNLVRLMQVPDWVLRDNLKFQPFKPGTPKVLQKNA
NVFDAIRGGDILLHHPYQSFNPVIELLDQSAADPQVVAIKMTVYRTGTDSVLMQSLLR
AAQNGKEVTVVVELMARFDEEANIGWATKLEEVGAHVIYGVVGYKTHAKMLMIVR
REESKAGGLLRRYVHLGTGNYHPKTARLYSDFGLLTCNEDIGADANEVFKQLTGLG
RAQTLKHLWQAPFSLQPNVVAGIRAEAEAARTGKRSRIIAKMNSLLEPETIAALYEAS
QAGVKIDLIVRGVCGLRAGVKDLSENINVWSIIGRQLEHHRVFYFFADGEEKVYLSSA
DWMDRNFFRRIELAFPILDRKLKRRVITEGLQIYLTDNTLAWELGADGSYHQRRGSR
TSPHSSQAELIELLKD

>Ca.Accumulibacter8

MTTEAPTADTPAQQSPITFPPEYFQNREMSLLAFNRRVLWQAKNPRTPLLERLRFLCI
VSSNLDEFFEIREAGIKEQLKLGSVAITTDGKTAREVYRLVSDETHAIVKEQYALLND
EILPLLAAEGIRFVKRAEWNVAQREWISDFFFREVMPVITPIGLDPSHPFPRVLNKSLN
FAVELEGRDAFGRSSGAAIVQAPRVLPRVIRLPRELGDAEYTFVFLSSILHEFVHELFA
GMKVLGCYQFRVTRNSDLFVDEEEVKNLRAKIQGELPQRHFGDAVRLEVANSCSEA
MTQFLLGQFNLTESDLFRVAGPVNLVRLMQVPDWVVRNDLKFPPFTPGIPKALQKCH
SVFDSIRGGDILLHHPYQSFNPVIELLEQAATDPQVVAIKMTVYRTGTDSVLMQSLLR
AAQNGKEVTVVVELMARFDEEANIGWATKLEEVGAHVVYGVVGYKTHAKMLMIV
RREEGKAGSSLLRRYVHLGTGNYHPKTARLYSDFGLLTCNEEIGADTNEVFKQLTGL
GRAQTLTHLWQAPFTLQPNVVAGIRAEADAARSGKRARIIAKMNSLLEPETIAALYE
ASQAGVKIDLIVRGVCGLRAGVKDLSDNISVRSIIGRQLEHHRVFYFYAGGEEKVYLS

SADWMERNFFRRIELAFPVLDRKLKRRVIAEGLQIYLADNTLAWELGPDGTYHQRRG
SRTSRHASQSELIELLQS

>Escherichia_coli;

MKKKSPHINREISWLSFNERVLQEAEDRSTPLIERLKFLGIFSNNRDEFYRVRVATVRR
LAKLGKKAMDIYGEAPQDLLTRLQRKVIEQQNKFEEIYQELLKELAINDVYLINEKHL
SIHQQQYVHDYFIKEVMPTLFPIMIDDNKPFPYMKDKASYLYLKMEWINNPKKNKYA
LIEIPSRVISRFIILPDQGAKKYVIMLDDVIRYNTDLIFEVFGYKTVEAYNIKLTRDAEL
DIDNDVSKSMIEKVSKSLKARKQGQPVRFVYDAAMPNDMLRYIMRKLGMAKKDNA
IPGGRYHNFKDFISFPNLGVKTLVYNHPPAIEHKYFIDHHQTILKVIKNRDILLHYPYHS
FNHIVNLLREASIDPVVESIKITLYRVADSSRIANALINAVKNGKKVTVLVELQARFDE
ENNIYWANKLSEEGARVIYGVPGLKVHSKLFLITTKESGKEVKYAHIGTGNFNEKTA
RIYTDFTLLTCNKLITAEVSKVFDFYENNFVVEQFKHLAVAPFYMRKLFVQLIDHEITL
AKNKKDGSIILKMNSLVDKEMIDKLYEASRAGVKITLIVRGACSLVTEMEGWSDNIIA
YSIVDKYLEHARVFIFGNGGNEKIYISSADWMSRNLDSRSEVAVPIYDKDIRKQLKDII
NIQLSGNTKVRIIDRKQENHYKKPQKGEKKLRVQDEVYSYLQKDTERYLTVPKNDKL
IMN

Appendix-D

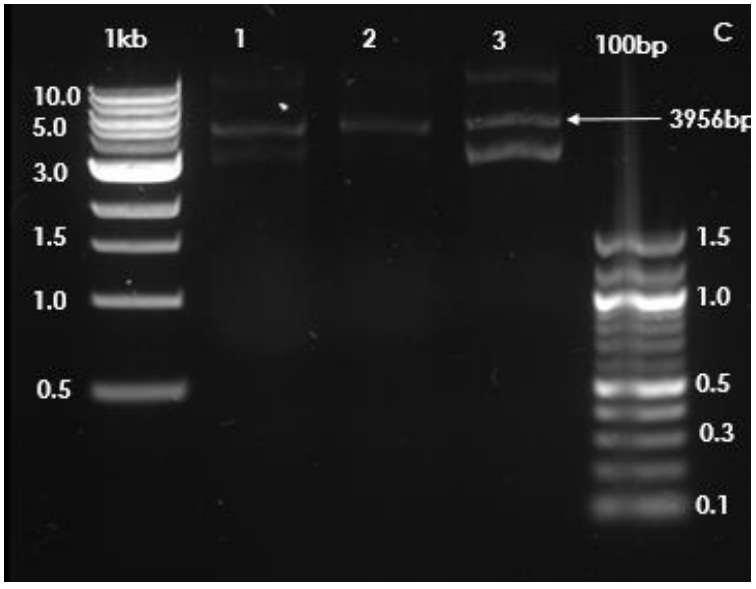


Figure A. Figure presenting restriction digestion of plasmid DNA (Primer pair C):

1 KB ladder on the left side and 100 bp ladder on the right side for reference.

select all 21 sequences selected GenBank Graphics Distance tree of results MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Candidatus Accumulibacter phosphatis clone LB_132 polyphosphate kinase gene, partial cds	Candidatus Accumulibacter phosphatis	252	502	100%	7e-62	85.96%	1114	KP738034.1
Candidatus Accumulibacter phosphatis clone JA_17 polyphosphate kinase gene, partial cds	Candidatus Accumulibacter phosphatis	161	321	98%	1e-34	79.22%	1114	KP738104.1

Figure B. *BLAST analysis of nucleotides sequences obtained from Sanger sequencing of cloned inserts (primer pair A).*

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Candidatus Accumulibacter phosphatis clone LB_132 polyphosphate kinase gene, partial cds	Candidatus Accumulibacter phosphatis	255	255	99%	3e-63	86.32%	1114	KP738034.1
Candidatus Accumulibacter phosphatis clone JA_17 polyphosphate kinase gene, partial cds	Candidatus Accumulibacter phosphatis	159	159	98%	2e-34	78.97%	1114	KP738104.1

Figure C. BLAST analysis of nucleotides sequences obtained from Sanger sequencing of cloned inserts (primer pair C).

Appendix-E

Table B. *Preparation of qPCR Reaction Master Mix (MM) to study the gene expression profiles of ppk1 in Hias reactor.*

Components	Initial conc	Final conc	Volume Per run (µl) (1rxn)
EvaGreen qPCR mix	5X	1X	3µl
Forward Primer	10µM	0.2µM	0.3 µl
Reverse Primer	10µM	0.2µM	0.3 µl
H20	-	-	10.4µl
Total Volume			14 µl
Template			1µl (1:10diluted)

Appendix-F

Sequences retrieved after Sanger Sequencing

Assembling of *ppk1* amplicons sequences to generate consensus sequence in CLC Main Workbench 7.9.3. (Nucleotides in green colour is the insert sequence of *ppk1* gene) PRIMER (A&C) both Forward and reversed sequences

GCATTGGCGGTTTCCTGGTCAGATAGCCAGTAGCCTGACATCATCGGGTTCAG
GCACGGTTCTGCGGAACTGACCTTCTAACGTTGAAAGAATCTAGGTGAAGATC
TTTTGATTATCTCATGCCTGACATTTATATTCCCCAGACATCAGGTTAAATGGC
GTTTTTGATGTCATTTTCGCGGTGGCTGAGATCAGCCACTTCTTCCCCGATAAC
GGAGACCGGGCACACTGGCCCATATCGGTGGTCATCATGCGCCAGCTTTCATC
CCCGATATGCACCACCGGGTAAAGTTCACGGGAGACTTTATCTGACAGCAGAC
GTGCACTGGCCAGGGGGATCACCATCCGTCGCCCCGGCGTGTCAATAATATCA
CTCTGTACATCCACAAACAGACGATAACGGCTCTCTCTTTTATAGGTGTAAACC
TTAAACTGCCGTACGTATAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCG
GGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGAT
TAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCC
AGTGAATTGTAATACGACTCACTATAGGGCGAATTGAATTTAGCGGCCGCGAA
TTCGCCCTTCTCTTGTTGAGTACGCGCGGGAAGGGATGCGATGGATCAAGACC
GATCGGGGTCAGGACCGGCATGACTTCGCGGAAAAAATAGTCCTTGATCCACT
CGCGTTGTTCGACGGTCCATGCCGAGCGCCGCAGGAAGACAATGCCTTCCTCG
GCCAGTTTGGGAAGGATCTCGTCGTTGAGCAGGGCGTACTTTTCCTCAACCAGC
GAATGTGTTTCGTCGCTGACCAGACGGGAGAAGGCGATTCGTAACTGCGCCCC
CCCCTCCCCAACGTTC

TATCACGGGGCCGGGGAATCCGGCCCTTCTCTTGTTGAGTACGCGCGGGAAGGGA
TGCGATGGATCAAGACCGATCGGGGTCAGGACCGGCATGACTTCGCGGAAAAAA
TAGTCCTTGATCCACTCGCGTTGTTCGACGGTCCATGCCGAGCGCCGCAGGAAGA
CAATGCCTTCCTCGGCCAGTTTGGGAAGGATCTCGTCGTTGAGCAGGGCGTACTT
TTCCTCAACCAGCGAATGTGTTTCGTCGCTGACCAGACGGGAGAAGGGCGAATTC

GTTTAAACCTGCAGGACTAGTCCCTTTAGTGAGGGTTAATTCTGAGCTTGGCGTA
ATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACA
ACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT
AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCG
TGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
GGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGG
CGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCACGG
GATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCC
GTAAAAGGCCGCGTTGCTGGCGTTTTCCATAGCTCGCCCCCTGACGAGCATCACA
AAAAATCGACGCTCAGTCAGAGGTGGCGAAACCCGACAGACTATAAGATACAAG
CGTTCCCCTGGAGCTTCCTTCGTGCGCCTCTCCTGTTCGACCTGCCGCTACGGGAT
ACTTGTTCCGCCCTATC CTACCGTCTGGTCAGCGACGA

FASTA sequence of *ppk1* gene, using PRIMER (A)

>seq1 *ppk1* gene (Primer pairs A)
CTCTTGTTGAGTACGCGCGGGAAGGGATGCGATGGATCAAGACCGATCGGGGT
CAGGACCGGCATGACTTCGCGGAAAAAATAGTCCTTGATCCACTCGCGTTGTT
CGACGGTCCATGCCGAGCGCCGCAGGAAGACAATGCCTTCCTCGGCCAGTTTG
GGAAGGATCTCGTCGTTGAGCAGGGCGTACTTTTCCTCAACCAGCGAATGTGTT
TCGTCGCTGACCAGACGGGAC

(*ppk1* gene) Using A primer pairs

AATTCTTTCCCCCGAATAACGGGAGACCGGCACACTGGCCATATCGGTGGTCA
TCATGCGCCCAGCTTTCATCCCCGATATGCACCACCGGGTAAAGTTCACGGGA
GACTTTATCTGACAGCAGACGTGCACTGGCCAGGGGGATCACCATCCGTCGCC
CCGGCGTGTCAATAATATCACTCTGTACATCCACAAACAGACGATAACGGCTC
TCTCTTTTATAGGTGTAAACCTTAAACTGCCGTACGTATAGGCTGCGCAACTGT
TGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGG
GGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCAC
GACGTTGTAAAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAA
TTGAATTTAGCGGCCGCGAATTCGCCCTTCTACCGTCTGGTCAGCGACGAAACA
CATTCGCTGGTTGAAGGAAAAGTACGCCCTGCTCAACGAACGAGAATCCTTCC
CAAACTGGCCGAGGAAGGCATTGTCTTCCTGCGGCGCTCGGCATGGACCGTCG

AACAACGCGAGTGGATCAAGGACTATTTTTTCCGCGAGGTCATGCCGGTCCTG
ACCCCGATCGGTCTTGATCCCTCGCATCCCTTCCCGCGCGTACTCAACAAGGAG
GAAAGGGCGGAATTCGTTTTAAACTTCCG

GGTGCCTCCTTTCTTCTACCGTCTGGTCAGCGACGAAACACATTCGCTGGTTGA
AGGAAAAGTACGCCCTGCTCAACGAACGAGAATCCTTCCCAAACTGGCCGAGG
AAGGCATTGTCTTCCTGCGGCGCTCGGCATGGACCGTCGAACAACGCGAGTGG
ATCAAGGACTATTTTTTCCGCGAGGTCATGCCGGTCCTGACCCCGATCGGTCTT
GATCCCTCGCATCCCTTCCCGCGCGTACTCAACAAGAGAAGGGCGAATTCGTTT
AAACCTGCAGGACTAGTCCCTTTAGTGAGGGTTAATTCTGAGCTTGGCGTAATC
ATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAA
CATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT
AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGT
CGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGT
ATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGG
CTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGA
ATCAGGGGATAACGCAGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCA
GGAACCGTTAAAAGGGCCGCGTTGCTGGCGTTTTTCCATAGGCTTCGCCCCCCC
TGACGAGCATCACAAAAATCGACGCTCAAGTTCAGAGGTGGGCTGAAACCGG
AACAGGGACTAAAGAATCCAGGCGTTTTCCCCCTGGAAAGCTCCCCTCGTGGC
GCTCTCCTGTTCGACCCCTGGCCGCTTAACCGGAATATTGTGT

>seq2 *ppk1* gene (Primer pairs A)

CTACCGTCTGGTCAGCGACGAAACACATTCGCTGGTTGAAGGAAAAGTACGCC
CTGCTCAACGAACGAGAATCCTTCCCAAACTGGCCGAGGAAGGCATTGTCTTC
CTGCGGCGCTCGGCATGGACCGTCGAACAACGCGAGTGGATCAAGGACTATTT
TTTCCGCGAGGTCATGCCGGTCCTGACCCCGATCGGTCTTGATCCCTCGCATCC
CTTCCCGCGCGTACTCAACAA

Ppk1 gene. C PRIMER.

ATCGACTAATTACCTCGTGGCGGCTTAGCGTTCCCATGTCAGGATAGCCCAGAT
AAGCTGACAATCAATTCCGGAGGGTCAAGCAACCGTTCATGCGGACTGCCTTT

CTACCGTGAAAAAGGATCTAAGGTGAAGATCCCTTTTTTGATAAATCTCATGCC
TGACATTATTATTCCCCAGAAACATCAGGTAAATGGCGTTTTTTGATGTCATTT
TCGCCGGTGGCTGAGATCAGCCACTTCTTCCCCGATAACGGAGACCGGCACAC
TGGCCATATCGGTGGTCATCATGCGCCAGCTTTCATCCCCGATATGCACCACCG
GGTAAAGTTCACGGGAGACTTTATCTGACAGCAGACGTGCACTGGCCAGGGGG
ATCACCATCCGTCGCCCCGGCGTGTCAATAATATCACTCTGTACATCCACAAAC
AGACGATAACGGCTCTCTCTTTTATAGGTGTAAACCTTAAACTGCCGTACGTAT
AGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACG
CCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCA
GGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGA
CTCACTATAGGGCGAATTGAATTTAGCGGCCGCGAATTCGCCCTTCTCTTGTTA
AGTCGCGCGGGAAGGGGATGCGATGGATCAAGACCGATCGGGGTCAGGACCG
GCATGACTTCGAGGAAAAAATAGTCCTTGATCCACTCGCGTTGTTCGACGGTCC
ATGCCGAGCGCCGCAGGAAGACAATGCCTTCCTCGGCCAGTTTGGGAAGGATC
TCGTCGTTGAGCAGGGCGTACTTTTCCTCAACCAGCGAATGTGTTTCGTCACTG
ACCAGACGGTAGAGGCGGAATTTCCTTAACCG

GGCCGGGCGGCCGAAAACCGCCCCTTTCTCTTGTTAAGTCGCGCGGGAAGGGG
ATGCGATGGATCAAGACCGATCGGGGTCAGGACCGGCATGACTTCGAGGAAA
AAATAGTCCTTGATCCACTCGCGTTGTTCGACGGTCCATGCCGAGCGCCGCAG
GAAGACAATGCCTTCCTCGGCCAGTTTGGGAAGGATCTCGTCGTTGAGCAGGG
CGTACTTTTCCTCAACCAGCGAATGTGTTTCGTCACTGACCAGACGGTAGAAGG
GCGAATTCGTTTAAACCTGCAGGACTAGTCCCTTTAGTGAGGGTTAATTCTGAG
CTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCAC
AATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCT
AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGT
CGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGA
GGCGGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCG
CTCGGGTCGTTCGGCTGCGGGCGAGCCGGTATCAGCTCACTCCAAAAGGGCGG
GTAATACGGGTTATCCACCAGAAATCAGGGGTATAACCGCAACGAAAAGAACT
TGTGAGCCCAAAGGGCCACGCAAAAGTGCCCTGGAACCGTTAAAAAAGGCGC
GTTGGCCTGGCGTTTTTTCCAATAGGCTCCGCCCCCCCTTGACAGAAGAT

>seq3 ppk1 gene (Primer pairs C)

CTCTTGTTAAGTCGCGCGGGAAGGGGATGCGATGGATCAAGACCGATCGGGGTC
AGGACCGGCATGACTTCGAGGAAAAAATAGTCCTTGATCCACTCGCGTTGTTCGA
CGGTCCATGCCGAGCGCCGCAGGAAGACAATGCCTTCCTCGGCCAGTTTGGGAAG
GATCTCGTCGTTGAGCAGGGCGTACTTTTCCTCAACCAGCGAATGTGTTTCGTCAC
TGACCAGACGGTAG

Ppk1 gene using C PIMER

TGCCGTTTTTTTTCCCTGTTCCAAGAATAGCCCCAGTAAGCTGGACAATTCATC
CGGGGGTCCAGCACCGTTTCTGCGACTGGCTTTCTACCGTGAAAAGGGATCTA
GGGTGAAAGATCCCTTTTTTGATAATCTCATGCCTGACATTTATATTCCCCAGA
ACATCAGGGTTAATGGGCGTTTTTTGATGTCATTTTCGCGGTGGCTGAGATCAG
CCACTTCTTCCCCCGATAACGGAGACCGGCACACTGGCCATATCGGTGGTCATC
ATGCGCCAGCTTTCATCCCCGATATGCACCACCGGGTAAAGTTCACGGGAGAC
TTTATCTGACAGCAGACGTGCACTGGCCAGGGGGATCACCATCCGTCGCCCCG
GCGTGTCAATAATATCACTCTGTACATCCACAAACAGACGATAACGGCTCTCTC
TTTTATAGGTGTAAACCTTAAACTGCCGTACGTATAGGCTGCGCAACTGTTGGG
AAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGA
TGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACG
TTGTAAAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAATTGA
ATTTAGCGGCCGCGAATTCGCCCTTCTCTTGTTAAGTCGCGCGGGAAGGGATGC
GATGGATCAAGACCGATCGGGGTCAGGACCGGCATGACTTCGAGGAAAAAAT
AGTCCTTGATCCACTCGCGTTGTTCGACGGTCCATGCCGAGCGCCGCAGGAAG
ACAATGCCTTCCTCGGCCATGTTTGGGAAGGATCTCGTCGTTGAGCAGGGCGT
ACTTTTCCTCAACCAGCGAATGTGTTTCGTCACTGACCAGACGGTAGGAAGGC
CGAATTTTGGAAAAT

GGCAGGGGCCGCGGAATTCCGCCCTTCTCTTGTTAAGTCGCGCGGGAAGGGAT
GCGATGGATCAAGACCGATCGGGGTCAGGACCGGCATGACTTCGAGGAAAAA
ATAGTCCTTGATCCACTCGCGTTGTTCGACGGTCCATGCCGAGCGCCGCAGGA
AGACAATGCCTTCCTCGGCCATGTTTGGGAAGGATCTCGTCGTTGAGCAGGGC
GTACTTTTCCTCAACCAGCGAATGTGTTTCGTCACTGACCAGACGGTAGAAGG
GCGAATTCGTTTAAACCTGCAGGACTAGTCCCTTTAGTGAGGGTTAATTCTGAG

CTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCAC
AATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCT
AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGT
CGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGA
GGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGC
TCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACG
GTTATCCACAGAATCAGGGGATAACGCAGAAAGAACATGTGAGCAAAAGGCC
AGCAAAAGGCCAGGAACCGTAAAAAGGGCCGCGTTGCTGGCGTTTTTCCATAG
GCTCCGCCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGTGGC
GAAACCCGACAGACTATAAGGATACCAGGCGTTTTCCCCCTGGAAAGCTCCCT
CGTGCGCCTCTCCTGTACCGAACCCCTGGCCGGCTTTACCCGGGAATACTGGTC
GTCATTTCTCCCCTTCAGGAAGCTGCGCTATTCTCATAGCTTCCAACGG

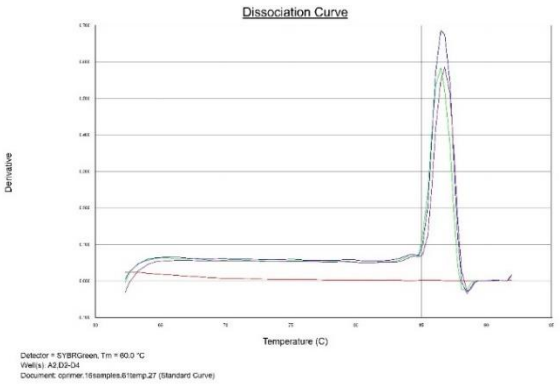
>seq3 ppk1 gene (Primer pairs C

CTCTTGTTAAGTCGCGCGGGAAGGGATGCGATGGATCAAGACCGATCGGGGTCA
GGACCGGCATGACTTCGAGGAAAAAATAGTCCTTGATCCACTCGCGTTGTTCGAC
GGTCCATGCCGAGCGCCGCAGGAAGACAATGCCTTCCTCGGCCATGTTTGGGAAG
GATCTCGTCGTTGAGCAGGGCGTACTTTTCCTCAACCAGCGAATGTGTTTCGTCAC
TGACCAGACGGTAG

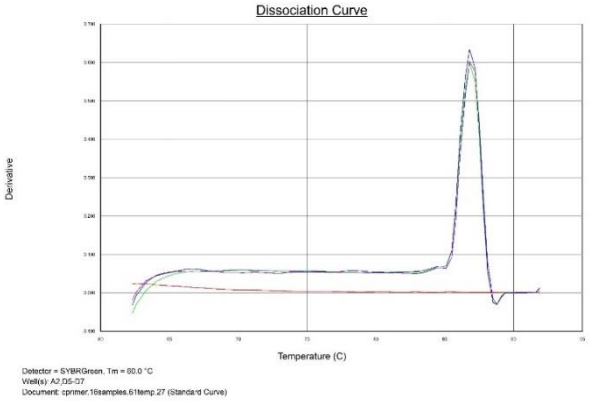
Appendix G

Ct values, RT-qPCR. Primer pairs: C, Microorganism: PAO Bacteria, *Ca.Acuumulibacter Phosphatis*, Functional Gene: *ppk1*

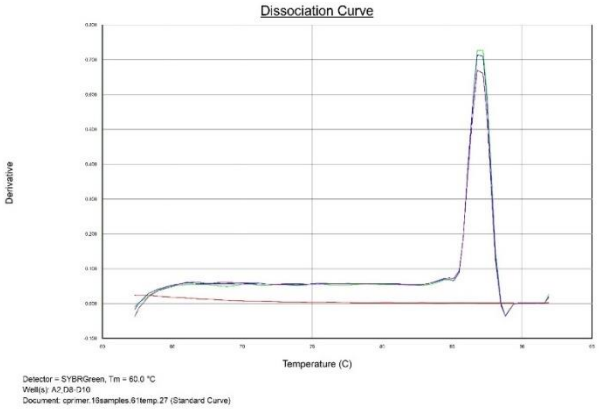
Anaerobic zone 1



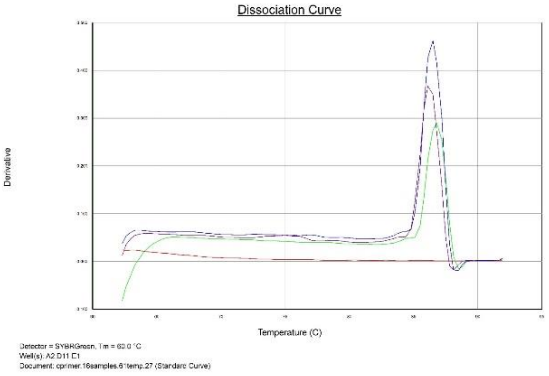
Anaerobic zone 1.1



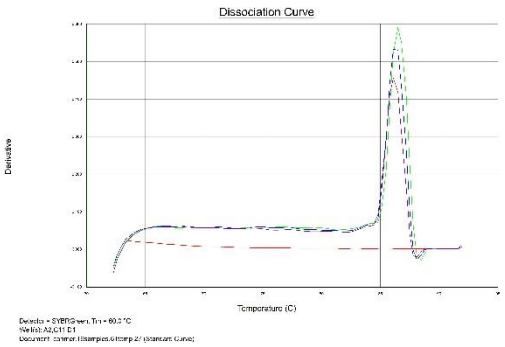
Anaerobic zone 2

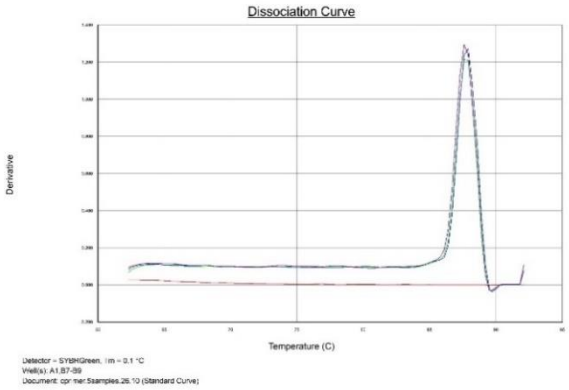


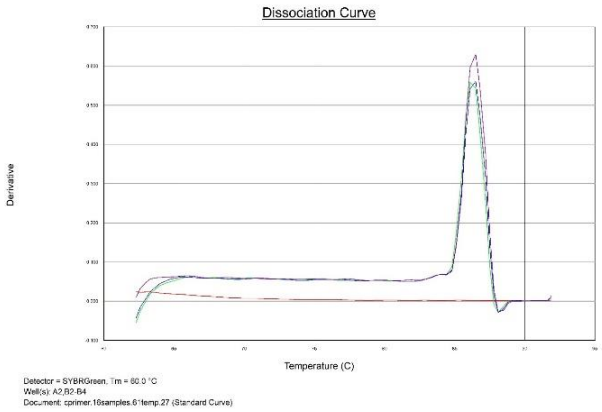
Anaerobic zone 3

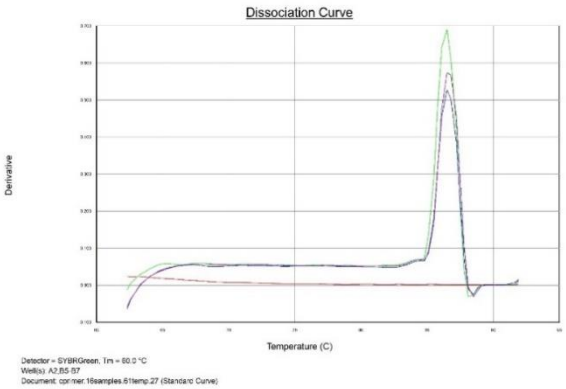


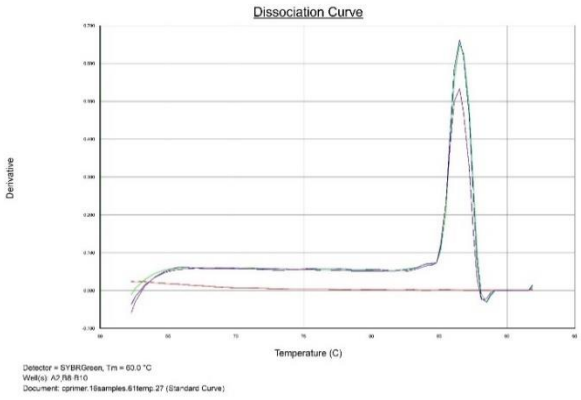
ALL Aerobic zone Samples

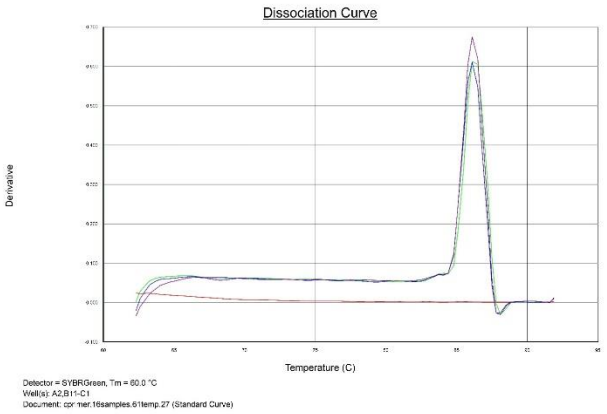


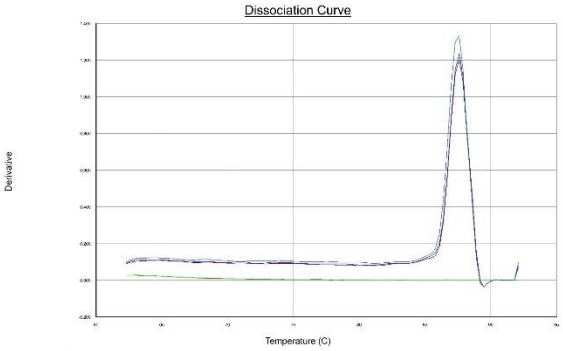


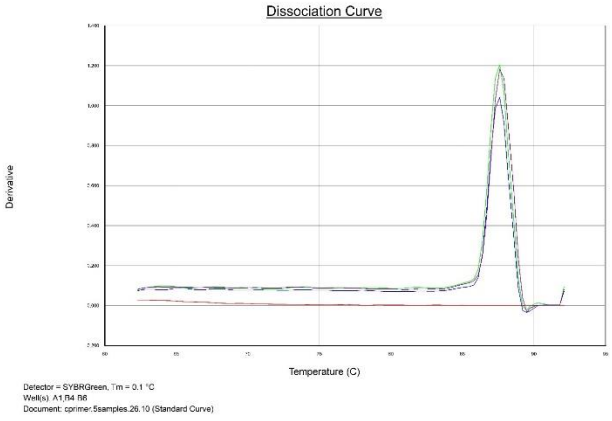


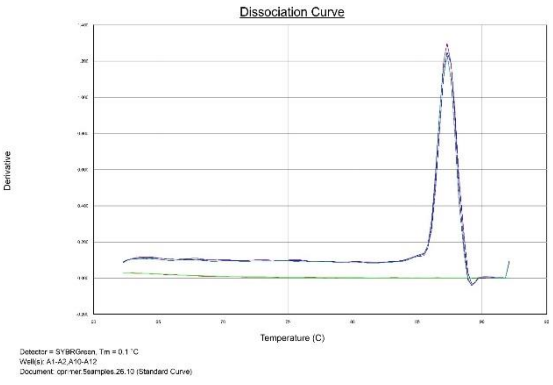


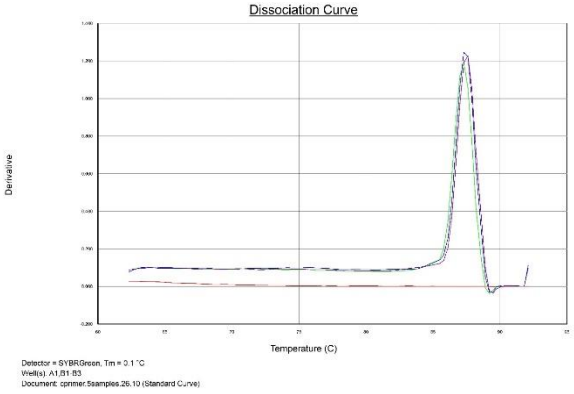


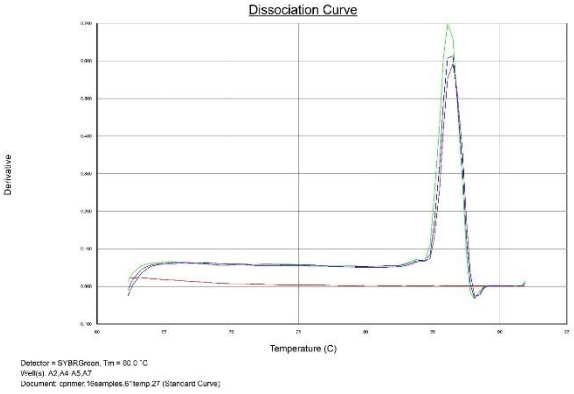

Dissociation Curve
Derivative
Temperature (C)
Detector = SYBRGreen, Tm = 60.0 °C
Well(s): A2,B11-C1
Document: cpr mer.16samples.61temp.27 (Standard Curve)

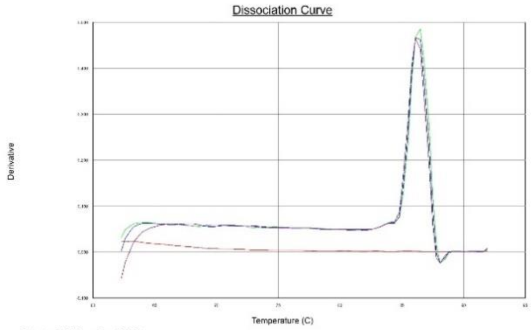

Dissociation Curve
Derivative
Temperature (C)
Detector = SYBRGreen, Tm = 0.1 °C
Well(s): A1-A2,A7-A9
Document: cprimer.5samples.26.10 (Standard Curve)


Dissociation Curve
Derivative
Temperature (C)
Detector = SYBRGreen, Tm = 0.1 °C
Well(s): A1,B4 B6
Document: cprimer.5samples.26.10 (Standard Curve)

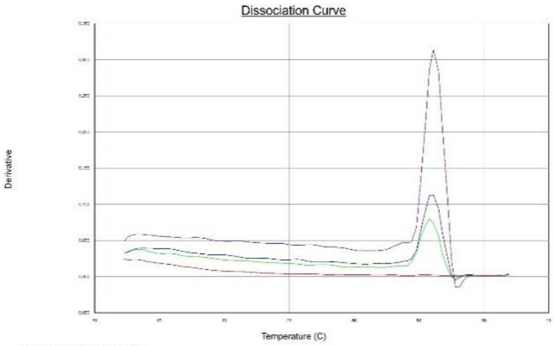

Dissociation Curve
Derivative
Temperature (C)
Detector = SYBRGreen, Tm = 0.1 °C
Well(s): A1-A2,A10-A12
Document: cprimer.5samples.26.10 (Standard Curve)


Dissociation Curve
Derivative
Temperature (C)
Detector = SYBRGreen, Tm = 0.1 °C
Well(s): A1,B1-B3
Document: cprimer.5samples.26.10 (Standard Curve)

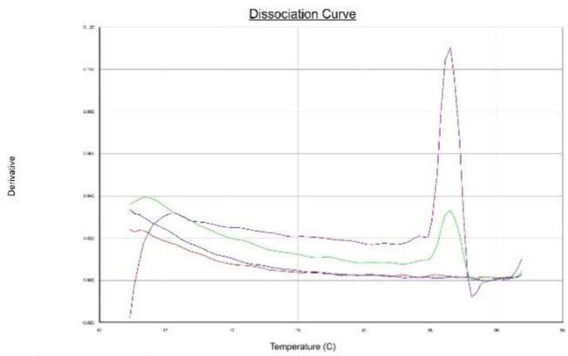

Dissociation Curve
Derivative
Temperature (C)
Detector = SYBRGreen, Tm = 60.0 °C
Well(s): A2,A4 A5,A7
Document: cprimer.16samples.61temp.27 (Standard Curve)



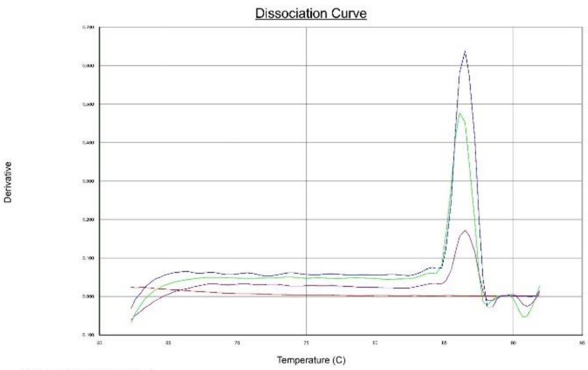
Detector = SYBRGreen, Tm = 60.0 °C
Well(s): A2,A8-A10
Document: cprimer.16samples.61temp.27 (Standard Curve)



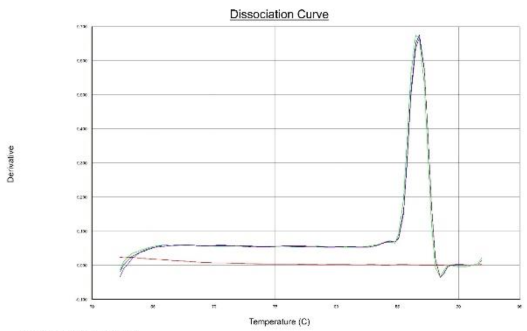
Detector = SYBRGreen, Tm = 60.0 °C
Well(s): A2,A11-B1
Document: cprimer.16samples.61temp.27 (Standard Curve)



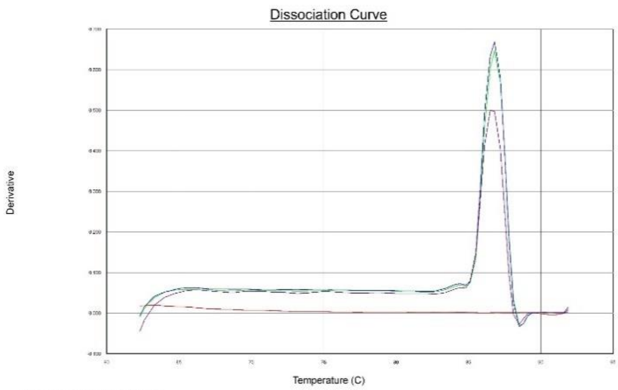
Detector = SYBRGreen, Tm = 60.0 °C
Well(s): A2,F6-F8
Document: cprimer.16samples.61temp.27 (Standard Curve)



Detector = SYBRGreen, Tm = 60.0 °C
Well(s): A2,C2-C4
Document: cprimer.16samples.61temp.27 (Standard Curve)



Detector = SYBRGreen, Tm = 60.0 °C
Well(s): A2,C5-C7
Document: cprimer.16samples.61temp.27 (Standard Curve)



Detector = SYBRGreen, Tm = 60.0 °C
Well(s): A3,C8-C10
Document: cprimer.16samples.61temp.27 (Standard Curve)