THESIS

DETERMINING THE EFFECT OF PRIMER MISMATCHES ON QUANTITATIVE PCR ACCURACY AND DEVELOPING GUIDANCE FOR DESIGN OF PRIMERS TARGETING GENES WITH SEQUENCE VARIATIONS

Submitted by Brett Michael Ledeker

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2012

Master's Committee

Advisor: Susan K. De Long

Pinar Omur-Ozbek
Kenneth F. Reardon

ABSTRACT

DETERMINING THE EFFECT OF PRIMER MISMATCHES ON QUANTITATIVE PCR ACCURACY AND DEVELOPING GUIDANCE FOR DESIGN OF PRIMERS TARGETING GENES WITH SEQUENCE VARIATIONS

Although quantitative PCR (qPCR) is a powerful tool for investigating environmental systems, target gene sequences for organisms of interest often are not well known, which has resulted in few reliable primers for many applications. Additionally, the sequences of target genes found in diverse strains often contain sequence variations, and therefore, primer sets containing single or multiple primer-template mismatches are common. However, the detrimental impact of these mismatches on quantification accuracy and amplification efficiency has not been investigated thoroughly. Thus, the research objectives of this study were to elucidate the relationships between primer mismatches and the accuracy of qPCR assays and to develop guidance for designing primers targeting genes displaying sequence variations. The *pcrA* gene (encoding perchlorate reductase) from *Dechloromonas agitata* was used as a model system for this study, and a linearized plasmid containing the cloned *pcrA* gene was used as the qPCR template. A large number of *pcrA* primers (16 forward and 16 reverse) were designed containing from zero to three mismatches at various locations. Combinations of primers were tested to determine the impact of mismatches on the amplification efficiency, the threshold cycle (C_T), and the quantification accuracy. Quantification accuracy was calculated as the percent detected by dividing the quantity measured with mismatch

primers by the quantity measured with perfect match primers and multiplying by 100. Single mismatches at the 3' end resulted in quantification accuracies as low as ~3%, and single mismatches at the 5' end resulted in quantification accuracies as low as ~33%. Double and triple mismatches at the 5' resulted in quantification accuracies as low as ~17% and ~2%, respectively. Reductions in quantification accuracy correlated with increases in C_T induced by mismatches but not with changes in amplification efficiency. Combining mismatched forward and reverse primers had an impact equivalent to the combined effect of the individual mismatch primers. Analogous qPCR tests were run with three other model genes: *celS* (encoding family 48 cellulase), *C23O* (encoding catechol dioxygenase, involved in toluene degradation), and *hydA* (encoding periplasmic hydrogenase, involved in fermentation). Primers were artificially designed to contain mismatches with these target genes, and results demonstrated that single or double mismatches can have a substantial detrimental impact on quantification accuracy in a broad range of systems. The results of this study indicate that caution must be taken to avoid mismatches when designing qPCR primers targeting genes with sequence variations and the findings serve to guide future design of primers for accurately quantifying genes in environmentally relevant systems.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Susan De Long for the insight, enthusiasm, and dedication that she put into this project. Not only did she direct my work, she taught me nearly all the basics of understanding molecular and microbiology that gave meaning to every experiment, as well as the project as a whole. I would also like to thank the other members of my committee, Dr. Pinar Omur-Ozbek and Dr. Kenneth Reardon for their suggestions and patience.

Furthermore, I would like to thank Paige Griffin, Emilie Lefèvre, Natalie Zeman, and others in my lab group who guided me in learning new techniques and made all the work more enjoyable. I would also like to thank Dr. Ann Hess for her recommendations regarding the statistical analysis, as well as Dr. Darrell Fontane for his insight at the inception of this project and for developing a computer program that made much of the analysis less difficult.

Finally, I would like to thank my wife Megan and my family for their constant love and encouragement. It was only with their support that I was able to take on this challenge and succeed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

1.1 Research Motivation

Quantitative polymerase chain reaction (qPCR) assays have been recognized as powerful tools for detecting and quantifying microorganisms, specific functional genes, and gene expression (Ginzinger, 2002; Rittmann, 2010). One major advantage of qPCR over other molecular methods, such as endpoint PCR and hybridization-based methods is that qPCR can achieve absolute quantification instead of only detection or relative quantification (Schneegurt and Kulpa, 1998; Zhang and Fang, 2006). However, successful gene detection and quantification accuracy are dependent on a large number of variables including optimal annealing temperature, amplicon length, and optimal design of oligonucleotide primers targeting the gene sequences of interest. However, genes encoding enzymes with the same function may occur in phylogenetically diverse microorganisms (Coates et al., 1999; Komon-Zelazowska et al., 2007), and substantial variations in the sequences of these functional genes often occur (Braker et al., 2000; Wu et al., 2001). Additionally, primers are typically designed based on available complete gene sequences that have been obtained from isolated pure cultures, and these pure culture strains may not represent the strains that dominate in real environmental systems (Zengler et al., 2002). Annotated functional gene sequences for environmentally relevant microorganisms remain limited despite increased genomic and metagenomic sequencing efforts, and thus target gene sequences for microorganisms that are functionally relevant in environmental systems are often not well known (Pereyra et al., 2010). Therefore, reliable primer sets are lacking for many applications. Moreover, the detection accuracy

can be difficult to determine for published primer sets when these primers are applied to environmental samples containing mixed microbial communities that have not been sequenced. One strategy to compensate for unknown target sequences and to simultaneously target genes with sequence variations is to design degenerate primers, which are sets of primers with multiple sequence variations. These primers are often designed based on deriving nucleotide sequences from protein sequences; however, multiple three-nucleotide codons can code for the same amino acids. Thus, nucleotide sequences are predicted using the most probable sequences based on related species. Despite the inclusion of degenerate bases at the location of nucleotides that may differ across related species, primers may still contain single or multiple primer-template mismatches with genes in target organisms (De Long et al., 2010). Although a few studies have shown that mismatches can reduce qPCR accuracy (Boyle et al., 2009; Bru et al., 2008), comprehensive studies quantifying the impact of primer-template mismatches in both forward and reverse primers on the amplification efficiency, threshold cycle (C_T), and quantification accuracy are lacking. Also, while some existing studies indicate that single mismatches near the 5' end of primers have no significant effect on quantification (Klein et al., 1999; Kwok et al., 1990), others indicate that these mismatches can significantly reduce detected quantities (Boyle et al., 2009; Bru et al., 2008). Moreover, additional studies have stated that single mismatches have no effect regardless of their location (Smith et al., 2002; Whiley and Sloots, 2005). Other studies involving multiple mismatches generally agree that as the number of mismatches in a primer increase, qPCR efficiency and accuracy tend to decrease (Guy et al., 2004; Smith et al., 2002), but these studies have not made efforts to separate the effect of multiple

mismatches from the effect of mismatches occurring at the 3' end of the primer. Studies elucidating how the number and location of primer mismatches affect qPCR results are needed to guide primer design when ideal primer designs are not achievable.

1.2 Research Objective

Available guidelines to optimize qPCR primer design have been developed for pure culture applications that target single gene sequences (Burpo, 2001; Lowe et al., 1990). Guidelines commonly used for simultaneously targeting genes with sequence variations in mixed cultures, such as the consensus degenerate hybrid oligonucleotide primer (CODEHOP) method, were originally developed to design primers for endpoint PCR, and thus, were meant primarily to ensure amplification rather than accurate quantification (Rose et al., 1998). Despite the intended use of such methods, a number of environmental qPCR studies have been conducted using CODEHOP-designed primers (Nozawa-Inoue et al., 2008; Pereyra et al., 2010; Quéméneur et al., 2010). Thus, reliable guidelines for designing PCR primers for quantification of genes with sequence variations are needed. Furthermore, acceptance of mismatches in primers is common for environmental applications, where lack of target gene sequences or high variability of target gene sequences leads to sub-optimal primer design. Thus, the objectives of this research are as follows:

1) Elucidate the relationship between primer mismatches and the performance of qPCR assays with respect to amplification efficiency, threshold cycle (C_T) and quantification accuracy.

a) Investigate the effect of single mismatches at a range of locations within primer sequences.
b) Investigate the effect of multiple mismatches within primer sequences.
c) Investigate the effect of mismatches occurring in both the forward and reverse primers.

2) Provide guidance for developing and applying qPCR assays to mixed cultures or environmental samples where target gene sequences may be diverse.

3) Develop a multiple primer set strategy to target a model gene (*pcrA* encoding perchlorate reductase) with sequence variations across related strains.

2.0 BACKGROUND AND LITERATURE REVIEW

2.1 Molecular Biology Tools for Quantification

Recent advances in molecular biological technology have provided an array of new quantitative tools, including hybridization-based and PCR-based techniques, which hold the potential to be significantly more accurate and informative than traditional microbiological techniques such as plate counting and optical density measurements (Higuchi et al., 1993). These tools also provide new opportunities to quantify targeted microorganisms and to predict the activity of these microorganisms via quantifying genetic markers (genes and gene transcripts, respectively). One major advantage of molecular tools is that they do not require culturing strains in the laboratory; it is estimated that only one percent of all existing microorganisms can be cultured under laboratory conditions (Zengler et al., 2002), and thus cultivation-independent quantification approaches are vital. Cultivation-independent molecular techniques include both hybridization-based techniques, such as DNA microarrays, and PCR-based techniques. However, DNA microarrays can suffer from low sensitivity, are expensive to generate, and assays are expensive to perform (Draghici et al., 2006; Long et al., 2001). Additionally, microarrays generally only provide relative quantification (Forster et al., 2003). Alternatively, PCR-based techniques can provide absolute quantification and offer low detection limits; these techniques include most probable number-PCR (MPN-PCR), competitive PCR, and real-time qPCR (Philippot, 2006).

Of the PCR-based techniques, qPCR is the only quantitative tool that does not require post-PCR processing and is highly accurate and precise (Sharma et al., 2007).

Real-time qPCR works by measuring the amplification of DNA in real time as the reaction occurs using fluorescent detection chemistry (Wittwer et al., 1997). The most common fluorescence chemistries are SYBR Green dye, which fluoresces when it binds to double stranded DNA, and TaqMan, which uses an oligonucleotide probe containing a fluorophore and a quencher and fluoresces when the probe is broken due to polymerase activity (Figure 2.1).

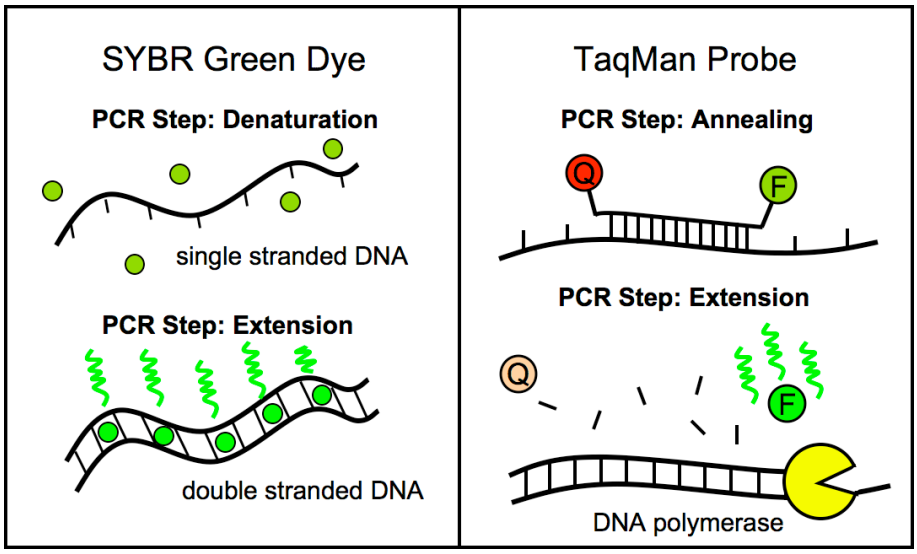

Figure 2.1. Real-time qPCR fluorescent detection chemistries. SYBR Green and TaqMan are the most common chemistries.

Fluorescence is measured at the end of each PCR cycle (denaturation, annealing, and extension) and plotted versus the cycle number on a logarithmic scale (Figure 2.2). A standard curve for calculating initial target gene quantities is established by selecting a threshold fluorescence during exponential amplification and identifying the threshold cycle (C_T) at which measured fluorescence crosses this threshold for a range of known template concentrations.

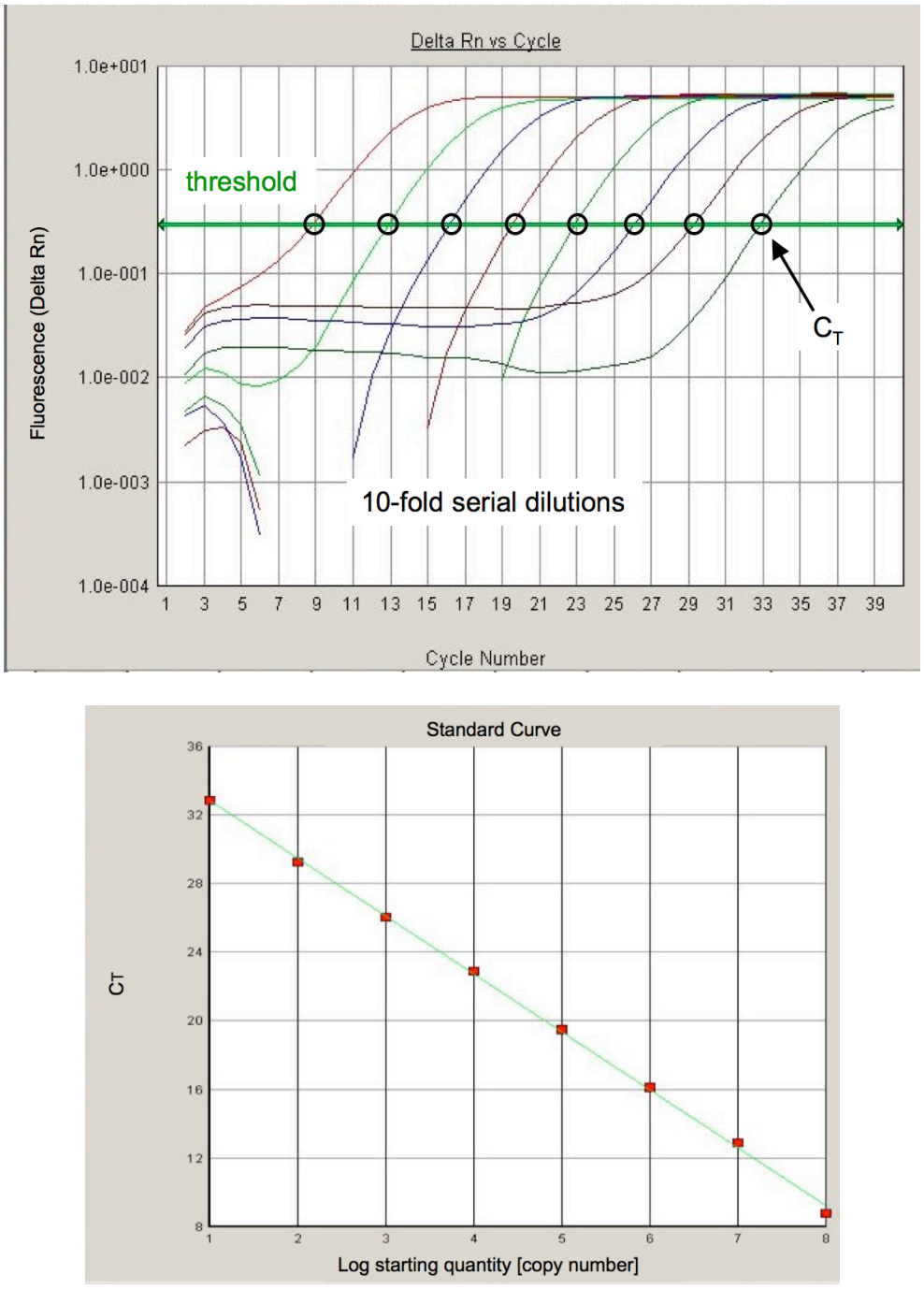

Figure 2.2. Representative plots showing qPCR fluorescence vs. C_T (top) and a standard curve (bottom). The plots show results for 10-fold serial dilutions of a calibration standard.

Due to the advantages of real-time PCR in terms of accuracy, sensitivity, high throughput, and versatility, this method has quickly gained popularity as the most

powerful quantification tool among the scientific, medical, and engineering communities (Kubista et al., 2006).

2.2 qPCR in Environmental Engineering

Quantitative real-time PCR has been applied to a wide range of biological processes relevant to environmental engineering, such as pollutant degradation, waste treatment, and pathogen inactivation (Chin et al., 2008; Ritalahti et al., 2006; Rittmann, 2006). qPCR can be used for quantification of all Bacteria or Archaea in a sample via assays that target conserved regions of 16S rRNA genes, or for quantification of bacteria capable of carrying out specific metabolic processes by targeting functional genes (genes encoding enzymes involved in these environmentally relevant processes) (Holmes et al., 2006). Gene expression also can be measured via reverse transcription-qPCR (RT-qPCR), which involves first converting isolated RNA (expressed genes) to complementary DNA via reverse transcription (Leininger et al., 2006; Li et al., 2010).

Since qPCR became widely available in the late 1990s, it has been applied to public health related aspects of environmental engineering including detection and quantification of pathogens such as *E. coli* O157:H7 and *Listeria monocytogenes* (Nogva et al., 2000; Oberst et al., 1998; Zhang and Fang, 2006). qPCR also has been used to track other potentially harmful microorganisms, such as assessing the distribution and abundance of toxic cyanobacteria in algal blooms in fresh waters (Coyne et al., 2005; Rinta-Kanto et al., 2005) and oceanic waters (Koskenniemi et al., 2007). qPCR also has been applied to the investigation of nitrogen removal processes, such as nitrification and denitrification, that are relevant to wastewater treatment and nutrient cycling in

agricultural and natural ecosystems (Hall et al., 2002; Harms et al., 2003; Henry et al., 2004; Philippot, 2006).

In environmental studies focused on reducing human impacts on natural systems, bioremediation has been gaining attention as a long-term and cost-effective method for removing toxic chemicals, such as uranium (N'Guessan et al., 2010; Wu et al., 2006) and carcinogens like *N*-nitrosodimethylamine (NDMA) (Sharp et al., 2007), released into the environment through industrial processes. qPCR has been proven as a useful tool in monitoring bacterial presence and estimating metabolic activity by measuring gene transcription in soils (Devers et al., 2004; Leininger et al., 2006) and aquatic environments (Laverick et al., 2004; Li et al., 2010). In recent studies, qPCR has been used to analyze the presence of functionally relevant bacteria in bioreactors treating acid mine drainage compounds (Pereyra et al., 2010), to detect perchlorate-reducing bacteria in environmental samples and treatment systems (De Long et al., 2010; Nozawa-Inoue et al., 2009), and to detect BTEX (benzene, toluene, ethylbenzene, and xylene) degraders under aerobic and hypoxic conditions in hydrocarbon contaminated groundwater (Schaefer et al., 2010; Tancsics et al., 2012).

qPCR also has been used to guide the design and optimization of engineered processes for water treatment and energy generation. A recently developed qPCR method for detection of live cells is propidium monoazide – qPCR (PMA-qPCR), which involves pre-treating samples with PMA to block amplification of DNA present in dead cells; this tool has been used to monitor the efficacy of various disinfection methods (Nocker et al., 2007; Wahman et al., 2009). qPCR has been used extensively to detect slow-growing methanogens in anaerobic digesters producing bioenergy in the form of methane (Song et

al., 2010; Steinberg and Regan, 2009; Yu et al., 2006). Recently published studies have tracked microorganisms in microbial fuel cells (MFCs), which use microorganism to convert organic matter into electricity through redox reactions (Jung and Regan, 2010; Ren et al., 2008; White et al., 2009).

2.3 Limitations of qPCR

Although qPCR is a powerful and versatile tool, qPCR assays are subject to the limitations and pitfalls of all PCR-based techniques (Klein, 2002). The major challenges associated with PCR-based techniques include the presence of PCR inhibitors in environmental samples, variations in DNA extraction efficiencies, and inefficiencies in primer-template annealing (Schneegurt and Kulpa, 1998). These limitations are particularly relevant to environmental analyses because environmental samples originating from a wide variety of sampling sites including soils, groundwater, surface water, and bioreactors, are often plagued by low nucleic acid yields and low quality DNA and RNA (Johnson et al., 2005; Smith et al., 2009; Stults et al., 2001). Quantifying gene expression from environmental samples is especially challenging due to the potential of nucleic acids binding to clay or organic matter, the possible co-extraction of humic substances, and the ubiquity of RNases in soil (Saleh-Lakha et al., 2011). DNA extraction efficiencies also have been shown to vary considerably depending on the extraction method and the type of environmental sample being processed (Martin-Laurent et al., 2001). However, methods exist to quantify losses occurring during extraction and the effect of inhibitors, and thus these challenges can be overcome (Johnson et al., 2005; Smith et al., 2006).

Primer-template annealing issues, however, cannot be addressed as readily due to the variety of sequences for target genes in mixed microbial communities. Recent studies have shown that published primers can have multiple primer-template mismatches with strains they were designed to target, and these mismatches have a significant effect on quantification accuracy (De Long et al., 2010; Guy et al., 2004; Sipos et al., 2007). Even single mismatches in primers have been shown to potentially have a major effect on quantification accuracy (Boyle et al., 2009; Bru et al., 2008). Moreover, when primers are not sufficiently complementary to gene sequences found in a sample, qPCR assays may fail to detect targeted genes entirely, yielding false negative results. Failures of this nature in qPCR can be difficult to identify and often are not documented in literature.

2.4 Primer Design and CODEHOP

Primer design programs, such as Primer3 and IDT SciTools, are abundant and widely used, but these programs are intended only for designing primers targeting genes in pure cultures where it is possible to design perfectly matching primers. Furthermore, for pure cultures generally primers can be developed to meet optimal design criteria including ideal primer length, GC base content, lack of secondary structures (e.g., hairpin loops and primer dimers), and optimal amplicon length (Burpo, 2001; Lowe et al., 1990). These programs cannot be used to design primers targeting multiple versions of a gene. The most widely used method for design of PCR primers targeting genes with sequence variations is CODEHOP. CODEHOP combines the degenerate primer method, where divergent sequences are targeted by introducing degenerate bases at select locations, and the consensus primer method, where divergent sequences are targeted by one primer set

designed based on the most common nucleotide at each location. These two design methods are combined to form a hybrid primer with degeneracies near the 3' end (last 11-12 bases) of the primer and no degeneracies near the 5' end as determined by the consensus method. CODEHOP begins with the input of protein sequences for target enzymes from multiple strains. These protein sequences are aligned and highly conserved regions are identified; then conserved sequences are converted into nucleotide sequences according to codon usage tables. Due to the fact that multiple gene codons can code for the same amino acid, nucleotide sequences cannot be definitively derived from protein sequences, and thus degeneracies are necessary for this translation. According to the theory proposed by Rose et al. (1998), CODEHOP operates by the mechanism that during PCR amplification, the degenerate core region ensures efficient amplification during the initial cycles by providing an exact match to the template at the 3' end where extension occurs. Degeneracies are avoided in the 5' end to increase primer specificity and because primer mismatches are thought to be more tolerable in this region.

CODEHOP represents an improved strategy for detecting related functional genes compared to previous ad hoc primer design efforts because CODEHOP uses rigorous computational methods, including a position specific scoring matrix and organism specific codon usage tables, to determine the optimal primer sequences (Braker et al., 1998; Hallin and Lindgren, 1999). However, CODEHOP does not allow for degeneracies in the 5' end, and thus, mismatches are likely in this region; in some cases these mismatches have been shown to drastically reduce quantification accuracy (De Long et al., 2010). CODEHOP also does not take other established primer design guidelines (e.g., recommendations for amplicon size and secondary structure avoidance) into account

automatically, leaving the user to account for these manually. The CODEHOP literature and documents associated with the web-based program do not promote use for quantitative assays, and thus, caution should be taken when applying this primer design strategy for development of qPCR primers for environmental applications.

2.5 Model Systems for Evaluating the Impact of Primer-Template Mismatches

The *pcrA* gene present in perchlorate-reducing bacteria (PRB) was used as a model system in this study to test the effect of primer mismatches. Perchlorate is used primarily by the defense industry as a component of explosives and rocket fuels and is toxic to human health (Chaudhuri et al., 2002; Wolff, 1998). Due to its high solubility, stability in environmental conditions, and presence in groundwater aquifers, microbial reduction has been identified as a promising remediation technique (Urbansky, 1998). Microbial degradation of perchlorate begins with the reduction of perchlorate to chlorite, which is catalyzed by perchlorate reductase (Bender et al., 2005). The *pcrA* gene from PRB (encoding the α-subunit of perchlorate reductase) was chosen as a model gene because PRB are phylogenetically diverse and sequenced *pcrA* genes display significant sequence variability (Coates et al., 1999). This model system is an illustrative example for similarly diverse microorganisms possessing the same functional gene, such as genes encoding enzymes for anaerobic respiration (Coates et al., 2002) and denitrification (Braker et al., 2000).

Three additional model genes were included in this study and published primer sets targeting functional genes in environmentally relevant microorganisms were tested. The effect of mismatch primers was evaluated for the *C23O* gene from *Pseudomonas*

putida mt-2, the *hydA* gene from *Shewanella oneidensis* MR-1, and the *celS* gene from *Clostridium thermocellum*. The *C23O* gene encodes catechol 2,3-dioxygenase, an enzyme that is involved in aerobic toluene degradation (Higashioka et al., 2009). The *hydA* gene encodes periplasmic [Fe-Fe] hydrogenase, an enzyme involved in hydrogen formation processes during fermentation (Meshulam-Simon et al., 2007). The *celS* gene encodes the enzyme glycoside hydrolase and belongs to the family 48 cellulases (Pereyra et al., 2010).

3.0 METHODOLOGY

3.1 Bacterial Strains and Culturing Conditions

PRB strains used included *Dechloromonas agitata* (ATCC 700666), *Dechloromonas aromatica* RCB (ATCC BAA-1848), *Dechloromonas* sp. PC1, and *Dechlorosoma* sp. KJ. Cultures were grown on R2A agar plates, which were incubated for one week at 30˚C.

Pseudomonas putida mt-2 (ATCC 39213) cultures were grown in tubes containing 6 mL of M9 medium (Maniatis et al., 1982). The medium was supplemented with 6 µL Stock Salt Solution (Bauchop and Eldsen, 1960), and 5 mM *m*-toluate was added as a carbon source. Cultures were grown overnight at 30˚C on a shaker at 140 oscillations per minute.

Shewanella oneidensis MR-1 (ATCC 700550) cultures were grown in tubes containing 6 mL of Luria-Bertani (LB) liquid medium. Cultures were incubated overnight at 30˚C on a shaker at 140 oscillations per minute.

Clostridium thermocellum genomic DNA (ATCC 27405D-5) was obtained from ATCC and resuspended in water. No culturing was necessary.

All cultures were grown from frozen stocks directly prior to DNA extraction.

3.2 Generating DNA Template for Absolute Quantification

pcrA gene

Plasmid template for quantification containing the cloned *D. agitata pcrA* gene was generated as shown in Figure 3.1.

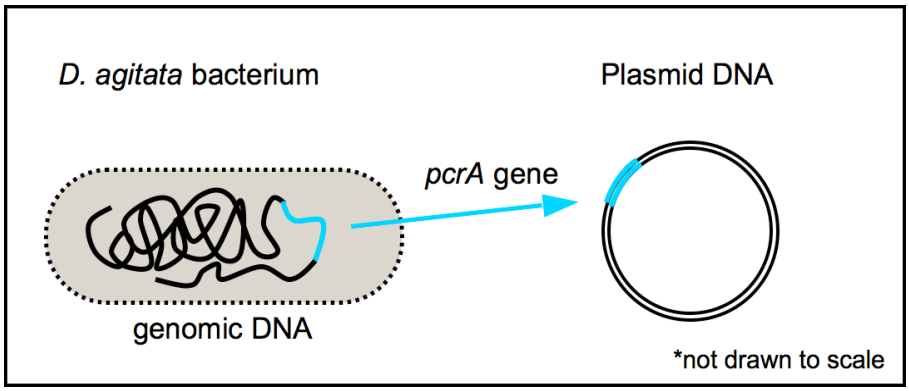

Figure 3.1. Diagram of *pcrA* cloning process. The *pcrA* gene from *D. agitata* was ligated into a plasmid vector.

The *pcrA* gene was PCR-amplified using primers designed based on a primer set (pcrA320F and pcrA598R) originally derived from CODEHOP (Nozawa-Inoue et al., 2008); however, these primers were modified to remove degeneracies and mismatches between the primers and the *pcrA* gene found in *D. agitata*. The modified primers are shown in Figure 3.2.

Forward primer sequence: 5'-GCGCACACCACTACATGTATGGTCC-3'

Reverse primer sequence: 5'-GATGGTCACTGTACCAGTCAAA-3'

Figure 3.2. Sequences of *pcrA* primers used to generate the plasmid template. Bases highlighted in black were changed to remove mismatches, and bases highlighted in gray were changed to remove degeneracies from the original published primers. The primer sequences shown are for the primers used in the present study after modifications.

PCR was run on a Bio-Rad S1000 thermal cycler (Bio-Rad, Hercules, CA) with a 50-µL reaction volume consisting of 36.75 µL of sterile water, 5 µL of 10X PCR reaction buffer (Clontech, Madison, WI), 4 µL of dNTP mix (10 mM of each nucleotide) (Clontech, Madison, WI), 1 µL of each primer (10 µM), and 0.25 µL of TaKaRa Ex Taq™ DNA Polymerase (Clontech, Madison, WI). DNA template was added to the reaction by scraping cells from the *D. agitata* culture plate and adding the cells directly to the reaction mix. The PCR reactions were run using the following thermal cycling program: 95˚C for 5 min; 35 cycles of 95˚C for 30 s, 60˚C for 30 s, 72˚C for 1 min; with a final extension at 72˚C for 7 min.

The PCR product was visualized on a 1% agarose gel, which was run at 100V for 30 minutes, stained with GelRed™ (Phenix, Candler, NC), and viewed using a BioChemi System (UVP, Upland, CA). Then the PCR amplicon was purified using a NucleoSpin® Extract II PCR clean-up kit (Macherey-Nagel, Bethlehem, PA). The purified PCR product was cloned into a plasmid vector using the TOPO TA Cloning® Kit for Sequencing, with chemically competent cells (Invitrogen, Carlsbad, CA). Amplified *pcrA* fragments were ligated into pCR®4-TOPO® plasmid per the kit protocol, as detailed below. For the ligation step, 0.5 µL of PCR amplicon were added to 1 µL of the salt solution, 3.5 µL of DNA-free water, and 1 µL of pCR®4-TOPO® vector. This ligation mixture was incubated at room temperature for 5 min and then put on ice. For transformation, 2 µL of the ligation reaction were added to a vial of chemically *E. coli.* This mixture was incubated on ice for 20 minutes, heat shocked for 30 seconds in a water bath at 42˚C, and then placed back on ice. 250 µL of provided SOC medium were added, and the cells were incubated at 37˚C for 1 hour on a shaker at 130 rpm. 50 µL, 100 µL,

and 150 µL of the transformed cells were added to three different pre-warmed LB agar plates with 50 µg/mL ampicillin; cells were spread with sterilized glass beads. Cultures were grown overnight at 37˚C, and single colonies were picked from the 50-µL spread plate and cultured overnight in LB liquid broth with 50 µg/mL ampicillin. To extract the plasmid DNA from the total cellular DNA, the liquid cultures were processed using the FastPlasmid™ Mini kit (5 Prime, Gaithersburg, MD). The concentration of cloned plasmid DNA was quantified using the Quant-iT™ kit (Invitrogen, Carlsbad, CA). Two selected clones were sequenced twice in both the forward and reverse direction using universal M13 forward and reverse primers to verify that they contained the correct *pcrA* sequence (i.e., the sequence was identical with the sequence of the *pcrA* gene in *D. agitata*). One clone with a *pcrA*-containing plasmid was selected to serve as the plasmid template for quantification in all of the qPCR tests.

Plasmid linearization has been shown to be an important step in preparing qPCR standards because plasmid DNA supercoiling can create secondary structure in the template that can suppress real-time PCR amplification (Bru et al., 2008; Chen et al., 2007; Johnson et al., 2005). Thus, to create a linear DNA template, the plasmids were digested using the XmnI restriction endonuclease (New England Biolabs, Beverly, MA) to cut the plasmid at one location as shown in Figure 3.3.

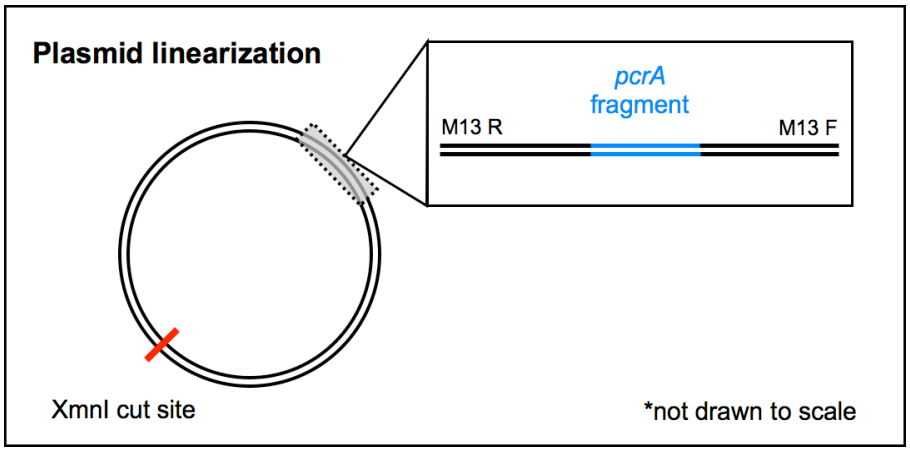

Figure 3.3. Plasmid map of *pcrA* template. Diagram shows the relative locations of the linearization cut site and cloned *pcrA* gene. There are 2093 bp between the cloned fragment and the cut site on one side and 1908 bp on the other side.

Restriction digest reactions were set up containing 150 µL of plasmid (approximately 10 µg of plasmid DNA), 2 µL of XmnI enzyme, 2 µL of bovine serum albumin (BSA, provided with enzyme), 20 µL of NEBuffer 4, and 26 µL of nuclease-free water for a total reaction volume of 200 µL. Digestion reactions were incubated for an incubation time of 4 hours at 37˚C and then the enzyme was heat-inactivated at 65˚C for 20 minutes. The restriction digest product was purified using the NucleoSpin® Extract II kit (Macherey-Nagel), and re-quantified.

DNA template for quantification of the *pcrA* gene in *D. aromatica* RCB, *Dechloromonas* sp. PC1, and *Dechlorosoma* sp. KJ was generated by extracting genomic DNA. Cells were scraped from R2A agar plates and processed using the UltraClean® Microbial DNA Isolation Kit (Mo Bio, Carlsbad, CA) per the recommended protocol.

C230 gene

DNA template for quantification of the *C23O* gene was obtained by isolating the TOL plasmid from strain *P. putida* mt-2. Approximately 6 mL of *P. putida mt-2* liquid culture were directly processed using the FastPlasmid™ Mini kit (5 Prime, Gaithersburg, MD) per the manufacturer's instructions to extract plasmid DNA.

hydA gene

DNA template for quantification of the *hydA* gene was obtained by extracting genomic DNA from *S. oneidensis* MR-1. Approximately 6 mL of *S. oneidensis* MR-1 liquid culture was processed using the UltraClean® Microbial DNA Isolation Kit (Mo Bio, Carlsbad, CA) per the manufacturer's protocol.

celS gene

Commercially available genomic DNA from *C. thermocellum* (ATCC 27405D-5) was used as the template for quantification of the *celS* (*cel48*).

Extracted DNA was quantified using a Quant-iT™ kit (Invitrogen, Carlsbad, CA) for all templates.

3.3 Primer Design

pcrA gene

In order to thoroughly investigate the effect that primer-template mismatches have on qPCR detection and quantification accuracy, forward and reverse *pcrA* primers were

designed containing mismatches at a range of locations. The effects of the number of mismatches and mismatch locations were tested by running qPCR assays with these primers.

First, primers were designed that contained zero mismatches with the cloned fragment of *pcrA* derived from *D. agitata*. Primers were designed based on guidelines specifically geared for quantitative PCR performance. The perfect match primers were designed using SciTools from Integrated DNA Technologies (IDT, Coralville, IA). The sequence of the cloned fragment of the *pcrA* gene was used as the program input to generate the perfect match primers *pcrA*. The program produced one primer set for the given fragment. Then the perfect match primer set produced was manually evaluated to confirm the primers adhered to qPCR primer design guidelines (Burpo, 2001; Lowe et al., 1990):

- Primers sequence length is within the optimal range of 18-22 base pairs.
- The G/C content is between 40% and 60%.
- The primers have no more than 3 G or C bases within the last 5 bases at the 3' end.
- The primers have a melting temperature (T_m) difference between the forward and reverse primers of less than 5˚C.
- The primers are free of secondary structure that includes hairpin loops, self dimers, and cross dimers that would be likely to inhibit amplification.
- The primers have no long base repeats (5 dinucleotides) or runs (5 bases).
- The primers have a G or C base at the 3' end.

- The primers result in an amplicon length that is between the 100-150 bp optimal range for qPCR.

The annealing sites of the perfect match primers were located within the cloned *pcrA* fragment as shown in Figure 3.4 below.

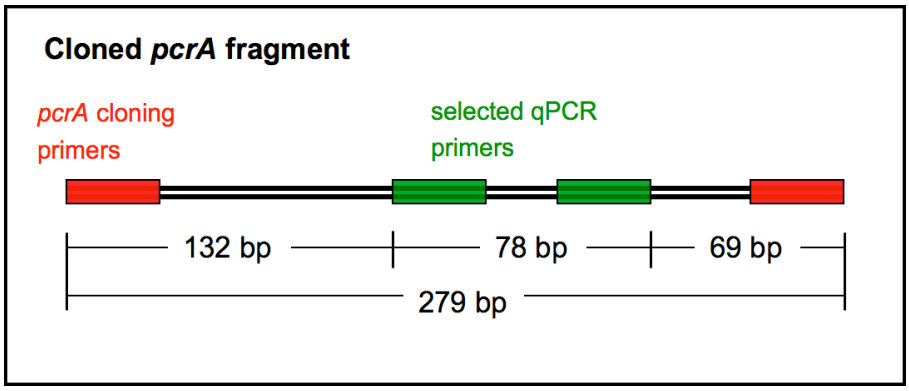

Figure 3.4. Location of the selected primer set within the cloned *pcrA* gene fragment.

For testing single mismatch primers, mismatch locations were chosen at the 5' end (bases 1 through 7), the middle (8-15), and at the 3' end (16-22) of the oligonucleotide sequence. Primers with double and triple mismatches were created by combining single mismatches from the beginning and middle of the oligonucleotide sequence; 3'-end mismatches were avoided in the multiple mismatch primers because the effect of single 3'-end mismatches had previously been shown to result in reductions of quantification accuracy as high as three orders of magnitude (Bru et al., 2008; Boyle et al., 2009). All mismatch primers were evaluated with the OligoAnalyzer (3.1) program from IDT SciTools to ensure that the base changes made to create mismatches did not lead to any secondary structures that would be detrimental to qPCR detection or accuracy. The criteria for classifying a primer as having detrimental secondary structures

were the following: 1) Gibbs free energy (ΔG) of the structure was less than -3 kcal/mol for hairpin loops, 2) ΔG was less than -6 kcal/mol for self dimers, or 3) ΔG was less than -6 kcal/mol for cross dimers. Mismatches were introduced by base conversion (i.e., guanidine was changed to cytosine, and adenine was changed to thymine or vice versa), in order to avoid changing the annealing temperature of the primers, as G/C content significantly affects melting temperature calculations (Bru et al., 2008). The perfect match primers and mismatch primers resulting from the design strategies described above are listed in Table 3.1. Primer names indicate whether they are a forward primer (F) or reverse primer (R) and where mismatch locations occur in the primer sequence, with multiple mismatch locations separated by hyphens.

Table 3.1. Perfect match and mismatch primers designed for the *pcrA* gene

Forward Primer	Mismatch Location	Sequence (5' to 3')
PF	none	ACACGATCAAGAATCACTCTCC
F2	2	AGACGATCAAGAATCACTCTCC
F4	4	ACAGGATCAAGAATCACTCTCC
F10	10	ACACGATCATGAATCACTCTCC
F12	12	ACACGATCAAGTATCACTCTCC
F19	19	ACACGATCAAGAATCACTGTCC
F21	21	ACACGATCAAGAATCACTCTGC
F2-4	2, 4	AGAGGATCAAGAATCACTCTCC
F2-10	2, 10	AGACGATCATGAATCACTCTCC
F2-12	2, 12	AGACGATCAAGTATCACTCTCC
F4-10	4, 10	ACAGGATCATGAATCACTCTCC
F4-12	4, 12	ACAGGATCAAGTATCACTCTCC
F10-12	10, 12	ACACGATCATGTATCACTCTCC
F2-4-10	2, 4, 10	AGAGGATCATGAATCACTCTCC
F2-4-12	2, 4, 12	AGAGGATCAAGTATCACTCTCC
F4-10-12	4, 10, 12	ACAGGATCATGTATCACTCTCC

* Base conversions are highlighted in gray.

Reverse Primer	Mismatch Location	Sequence (5' to 3')
PR	none	CAGAGAACGATACTGGTGCTG
R3	3	CACAGAACGATACTGGTGCTG
R5	5	CAGACAACGATACTGGTGCTG
R11	11	CAGAGAACGAAACTGGTGCTG
R13	12	CAGAGAACGATAGTGGTGCTG
R18	13	CAGAGAACGATACTGGTCCTG
R20	20	CAGAGAACGATACTGGTGCAG
R3-5	3, 5	CACACAACGATACTGGTGCTG
R3-11	3, 11	CACAGAACGAAACTGGTGCTG
R3-13	3, 13	CACAGAACGATAGTGGTGCTG
R5-11	5, 11	CAGACAACGAAACTGGTGCTG
R5-13	5, 13	CAGACAACGATAGTGGTGCTG
R11-13	11, 13	CAGAGAACGAAAGTGGTGCTG
R3-5-11	3, 5, 11	CACACAACGAAACTGGTGCTG
R3-5-13	3, 5, 13	CACACAACGATAGTGGTGCTG
R5-11-13	5, 11, 13	CAGACAACGAAAGTGGTGCTG

* Base conversions are highlighted in gray.

C23O gene

Primers targeting the *C23O* gene of *P. putida* mt-2 were developed by adapting primers from a previous study (Higashioka, et al., 2009). Theses primers (E3 [forward] and E2 [reverse]) were originally designed for mixed culture environmental assays. The E2 and E3 primers were used as the perfect match primers for the *C23O* gene and were modified to introduce mismatches. The E2 and E3 primers were evaluated against the design guidelines detailed above and were found to deviate from these guidelines as follows. The E3 primer sequence length was 26 bp, and this primer did not have a G or C at the 3' end; the amplicon length is 356 bp. The primer design approach was the same as for *pcrA,* and primers were checked as above to ensure that detrimental secondary structures were not introduced in the design. Single mismatches were chosen at locations

at the 5' end, internally (two locations selected), and 3' end of the primer sequence. Double and triple mismatch primers were designed by combining two and three of the single mismatches located internally and at the 5' end. The primers designed for the *C23O* gene are listed in Table 3.2.

Table 3.2. Perfect match and mismatch primers designed for the *C23O* gene

Forward Primer	Mismatch Location	Sequence (5' to 3')
E3	none	GGTATGGCGGCTGTGCGTTTCGACCA
C23O_F3	3	GGAATGGCGGCTGTGCGTTTCGACCA
C23O_F12	12	GGTATGGCGGCAGTGCGTTTCGACCA
C23O_F14	14	GGTATGGCGGCTGAGCGTTTCGACCA
C23O_F23	23	GGTATGGCGGCTGTGCGTTTCGTCCA
C23O_F3-14	3, 14	GGAATGGCGGCTGAGCGTTTCGACCA
C23O_F3-12-14	3, 12, 14	GGAATGGCGGCAGAGCGTTTCGACCA
Reverse Primer	**Mismatch Location**	**Sequence (5' to 3')**
E2	none	CAGAGAACGATACTGGTGCTG
C23O_R3	3	AGTACACTTCGTTGCGGTTACC
C23O_R7	7	AGAACAGTTCGTTGCGGTTACC
C23O_R12	12	AGAACACTTCGATGCGGTTACC
C23O_R20	20	AGAACACTTCGTTGCGGTTTCC
C23O_R3-12	3, 12	AGTACACTTCGATGCGGTTACC
C23O_R3-7-12	3, 7, 12	AGTACAGTTCGATGCGGTTACC

* Base conversions are highlighted in gray.

hydA gene

Primers targeting the *hydA* gene of *S. oneidensis* MR-1 were developed by adapting published primers hydA_hydB_1002F (forward) and hydA_hydB_1489RC (reverse) (Meshulam-Simon, et al., 2006). These primers were originally designed for application to *S. oneidensis* MR-1 in pure culture, and thus do not contain any degeneracies or mismatches. The primers were found to deviate from the design

guidelines detailed above as follows. The hydA_hydB_1489RC primer contains more than 3 G or C bases within the last 5 bases at the 3' end, this primer can form a self dimer (ΔG = -8.16 kcal/mol), and the amplicon length is 487 bp. The primers containing single and multiple mismatches were designed and checked as described above. Additionally, the mismatch primers were analyzed using OligoAnalyzer to ensure that base modifications did not change the ΔG value of the self dimer in the reverse primer so that perfect match and mismatch primers could be compared. The primers designed for the *hydA* gene are listed below in Table 3.3.

Table 3.3. Perfect match and mismatch primers designed for the *hydA* gene

Forward Primer	Mismatch Location	Sequence (5' to 3')
hydA_hydB_1002F	none	CGTGAAATCAGCCTCTGTC
hydA_F3	3	CGAGAAATCAGCCTCTGTC
hydA_F8	8	CGTGAAAACAGCCTCTGTC
hydA_F11	11	CGTGAAATCACCCTCTGTC
hydA_F17	17	CGTGAAATCAGCCTCTCTC
hydA_F3-11	3, 11	CGAGAAATCACCCTCTGTC
hydA_F3-8-11	3, 8, 11	CGAGAAAACACCCTCTGTC

Reverse Primer	Mismatch Location	Sequence (5' to 3')
hydA_hydB_1489RC	none	TCCTAATGGCTCGCCACC
hydA_R3	3	TCGTAATGGCTCGCCACC
hydA_R7	7	TCCTAAAGGCTCGCCACC
hydA_R11	11	TCCTAATGGCACGCCACC
hydA_R17	17	TCCTAATGGCTCGCCAGC
hydA_R3-11	3, 11	TCGTAATGGCACGCCACC
hydA_R3-7-11	3, 7, 11	TCGTAAAGGCACGCCACC

* Base conversions are highlighted in gray.

celS gene

Primers targeting the *celS* gene of *C. thermocellum* were developed by modifying published primers *cel48*_880F (forward) and *cel48*_980R (reverse) (Pereyra, et al., 2010). These primers were originally designed using CODEHOP for application to environmental samples containing mixed microbial communities, and thus contain degenerate bases. The published primers were modified to produce a perfect match with the *celS* gene found in *C. thermocellum*. The perfect match primers were found to deviate from the design guidelines as follows. The primers were longer than recommended (29 bp and 25 bp for the forward and reverse primers, respectively), both primers did not contain a G or C at the 3' end, and the reverse primer can form a self dimer (ΔG = -7.96 kcal/mol). The primers containing single and multiple mismatches were designed and checked as described above; the base modifications did not change the ΔG value of the self dimer in the reverse primer. The primers designed for *C. thermocellum* and the *celS* (*cel48*) gene are listed below in Table 3.4.

Table 3.4. Perfect match and mismatch primers designed for the *celS* gene

Forward Primer	Mismatch Location	Sequence (5' to 3')
cel48_PF	none	CACTGGTTGATGGACGTTGACAACTGGTA
cel48_F3	3	CAGTGGTTGATGGACGTTGACAACTGGTA
cel48_F7	7	CACTGGATGATGGACGTTGACAACTGGTA
cel48_F13	13	CACTGGTTGATGCACGTTGACAACTGGTA
cel48_F27	27	CACTGGTTGATGGACGTTGACAACTGCTA
cel48_F3-13	3, 13	CAGTGGTTGATGCACGTTGACAACTGGTA
cel48_F3-7-13	3, 7, 13	CAGTGGATGATGCACGTTGACAACTGGTA

Reverse Primer	Mismatch Location	Sequence (5' to 3')
cel48_PR	none	CCTGTTCACCTCTTTGGAAGGTGTT
cel48_R4	4	CCTCTTCACCTCTTTGGAAGGTGTT
cel48_R12	12	CCTGTTCACCTGTTTGGAAGGTGTT
cel48_R15	15	CCTGTTCACCTCTTAGGAAGGTGTT
cel48_R25	25	CCTGTTCACCTCTTTGGAAGGTGTA
cel48_R4-15	4, 15	CCTCTTCACCTCTTAGGAAGGTGTT
cel48_R4-12-15	4, 12, 15	CCTCTTCACCTGTTAGGAAGGTGTT

* Base conversions are highlighted in gray.

Multi-primer set quantification strategy for the pcrA gene

In order to investigate the feasibility of using a multi-primer set approach to achieve accurate quantification, primers were developed targeting the *pcrA* genes found in three PRB strains for which the *pcrA* gene has been fully sequenced: *D. agitata* (GenBank accession no. AY124796), *D. aromatica* RCB (AAZ47315), and *Dechloromonas* sp. PC1 (EU022026). These primers were designed by modifying published primers (Nozawa-Inoue et al., 2008). These primers (pcrA320F and pcrA598R) were originally designed by CODEHOP for mixed culture environmental assays targeting PRB. The primers were modified to remove degenerate bases to produce perfect match primers targeting the *pcrA* genes in each respective strain and then evaluated against the design guidelines detailed above in the section for *D. agitata*. Deviations from these guidelines were a longer primer length than recommended (25 bp for the forward primer), a relatively large amplicon (278 bases), the lack of a G or C 3' base in the reverse primer, and self dimers and cross dimers in nearly all of the primers. Primers sequences are listed below in Table 3.5. Primer names indicate the target gene, the PRB strain, and the primer direction.

Table 3.5. Primers sets designed to target *pcrA* genes with sequence variations in three PRB strains

Forward Primer	Mismatch Location	Sequence (5' to 3')
pcrA_agitata_F	none	GCGCACACCACTACATGTATGGTCC
pcrA_aromatica_F	none	GTGCCCACGACTACATGTATGGCCC
pcrA_PC1_F	none	GTGGTCACGACTACATGTATGGGCC

Reverse Primer	Mismatch Location	Sequence (5' to 3')
pcrA_agitata_R	none	GATGGTCACTGTACCAGTCAAA
pcrA_aromatica_R	none	GGTGATCGCCATACCAGTCGAA
pcrA_PC1_R	none	GATGATCACCGTACCAGTCGAA

*Bases with differences between strains are highlighted.

After the feasibility of this multi-primer set approach was demonstrated, two additional PRB strains were taken into consideration (*Dechlorosoma* sp. PCC [GenBank accession no. EU022027] and *Dechlorosoma* sp. KJ [EU571095]). Thus, all PRB strains for which the *pcrA* gene has been fully sequenced were considered in designing two sets of degenerate primers without mismatches for quantifying *pcrA* in mixed microbial communities. *Dechloromonas* sp. PC1 and *Dechlorosoma* sp. PCC have identical *pcrA* sequences and thus will be referred to as PC1/PCC hereafter. Alignments of the primer annealing locations for all PRB strains with available *pcrA* sequences are shown in Figure 3.5.

Forward Primer	
D. agitata	: GCGCACACCACTACATGTATGGTCC
D. aromatica	: GTGCCCACGACTACATGTATGGCCC
PC1/PCC	: GTGGTCACGACTACATGTATGGGCC
KJ	: GTGGTCACGACTATATGTATGGACC

Reverse Primer	
D. agitata	: GATGGTCACTGTACCAGTCAAA
D. aromatica	: GGTGATCGCCATACCAGTCGAA
PC1/PCC	: GATGATCACCGTACCAGTCGAA
KJ	: GGTGATCGCCATACCAATCGCA

Figure 3.5. Gene sequences for primer annealing sites for PRB strains with available *pcrA* sequences. Highlighted bases indicate locations where mismatches exist between strains.

D. aromatica RCB, *Dechloromonas* sp. PC1, *Dechlorosoma* sp. PCC, and *Dechlorosoma* sp. KJ *pcrA* genes have more similar sequences than the *D. agitata pcrA* gene. Thus, one set of degenerate primers was designed to target *pcrA* genes with sequence similarity to *D. aromatica*, and strains PC1, PCC, and KJ (hereafter referred to as *aromatica*-type *pcrA* genes). To avoid mismatches, degenerate bases were included at the locations of divergence. A second primer set was designed to target *agitata*-type *pcrA* genes. Figure 3.6 shows these primer sequences.

agitata-type *pcrA* primer set	
agitata-type F :	GCGCACACCACTACATGTATGGTCC
agitata-type R :	GATGGTCACTGTACCAGTCAAA

aromatica-type *pcrA* primer set	
aromatica-type F:	GTGSYCACGACTAYATGTATGGVCC
aromatica-type R:	GRTGATCRCCRTACCARTCGMA

Figure 3.6 Primer sets targeting all known *pcrA* gene sequences. The *agitata*-type primer set is a perfect match to the *pcrA* gene from *D. agitata*. The *aromatica*-type primer set contains degenerate bases to target *pcrA* in four PRB strains.

3.4 Quantitative PCR

pcrA gene

An annealing temperature of 53˚C was selected by performing endpoint gradient PCR on the plasmid template using the *pcrA* perfect match primers. PCR was run on a Bio-Rad S1000 thermal cycler (Bio-Rad, Hercules, CA) with a 50 µL reaction volume consisting of 34.75 µL of sterile water, 5 µL of 10X PCR reaction buffer (Clontech, Madison, WI), 4 µL of dNTP mix (10 mM of each nucleotide) (Clontech, Madison, WI), 1 µL of each primer (10 µM), 0.25 µL of TaKaRa Ex Taq™ DNA Polymerase (Clontech, Madison, WI), and 4 µL of the *pcrA* containing plasmid (10^8 copies). The PCR reactions were run at 95˚C for 5 min; 35 cycles of 95˚C for 30 s, annealing temperatures from 51˚C to 60˚C across 8 reactions for 30 s, 72˚C for 1 min; with a final extension at 72˚C for 7 min. The resulting PCR products were each run on an 1% agarose gel for 30 minutes at 100V, stained with GelRed™ (Phenix, Candler, NC) and viewed using a BioChemi System (UVP, Upland, CA) to identify the annealing temperature that produced the maximum band intensity.

Based on the DNA quantification results and the size of the pCR®4-TOPO® plasmid containing the *pcrA* fragment (4234 bp), the template concentration in terms of number of copies of the *pcrA* gene per µL was calculated. Then, template for qPCR standard curves was prepared by diluting plasmid DNA such that standard reactions contained 10 to 10^8 copies per reaction. The equation for calculating target gene concentration is given below.

$$\text{Copies/}\mu\text{L} = L \cdot \left(\frac{C}{m \cdot N} \right)$$

where L = Avogadro's constant (6.022 x 10^{23} molecules/mol)
C = concentration of DNA (g/µL)
m = molecular weight of one bp of DNA (650 g/mol)
N = length of plasmid or genomic DNA (bp)

qPCR was performed using an ABI 7300 real-time PCR system (Applied Biosystems, Foster City, CA). 25-µL qPCR reaction volumes contained 1X ABsolute™ Blue QPCR SYBR Green mix (Thermo Fisher Scientific, Waltham, MA), 70 nM of each primer, and 5 µL of nuclease-free water. 4 µL of plasmid template were added such that the final template concentrations were 10 to 10^8 target gene copies per reaction. Thermal cycling consisted of 95˚C for 15 min, followed by 40 cycles of 95˚C for 15 s, 53˚C for 30 s, and 72˚C for 30 s. The qPCR machine was set to read fluorescence directly following the 72˚C extension period of each cycle.

Each mismatch primer was tested by running duplicate qPCR reactions containing the mismatch primer and its perfect match primer (e.g., a mismatch forward primer with a perfect match reverse primer). Assays were conducted in 96-well plates, and reactions were run containing a broad range of template concentrations (10 to 10^8 gene copies per reaction) to assess the affect of initial template concentration on detection accuracies. All plates also included duplicate reactions with perfect match primers as a reference for interpreting the mismatch reaction data. No template controls were included for all experiments to monitor for contamination, and the absence of non-specific amplification was confirmed by analyzing dissociation curves.

Subsequently tests were run to assess the affect of combinations of forward and reverse mismatch primers. Nine representative primer combinations (F2:R3, F4:R20, F12:R11, F19:R18, F2-10:R5-11, F10-12:R11-13, F2-4-10:R20, F2-4-12:R5, and F4-10-12:R3-5-13) were chosen to investigate whether the effect of primer combinations could be calculated from the effects of individual mismatch primers tested with their complementary perfect match primer. qPCR was conducted as described above.

C23O gene

The number of copies of *C230* per µL in extracted TOL plasmid DNA was calculated using the equation above (N = 128,921 bp). qPCR reactions were set up as described above. Reactions with the perfect match primers contained 10 to 10^6 gene copies per reaction. Each mismatch primer was tested with the appropriate forward or reverse perfect match primer for 10^2 and 10^5 gene copies per reaction. Thermal cycling consisted of 95˚C for 15 min, followed by 40 cycles of 95˚C for 15 s, 60˚C for 30 s, and 72˚C for 30 s. The annealing temperature was the temperature listed for the *C23O* primer set selected from Higashioka, et al. (2009).

hydA gene

The number of *hydA* gene copies per µL in extracted *S. oneidensis* MR-1 genomic DNA was calculated using the equation above (N = 5,130 Kb). qPCR reactions were set up as described above with the exception that reactions with the perfect match primers contained 10^3 to 10^8 copies per reaction, and mismatch primers were tested for 10^4 and 10^7 gene copies per reaction. Thermal cycling consisted of 95˚C for 15 min, followed by

40 cycles of 95˚C for 15 s, 50˚C for 30 s, and 72˚C for 30 s. The annealing temperature was based on Meshulam-Simon et al. (2006).

celS gene

The number of *celS* gene copies in *C. thermocellum* genomic DNA was calculated using the equation above (N = 3,800 Kb). qPCR reactions were set up as described above; perfect match primers contained 10^3 to 10^8 copies per reaction, and mismatch primers were tested for 10^4 and 10^7 gene copies per reaction. Thermal cycling consisted of 95˚C for 15 min, followed by 40 cycles of 95˚C for 15 s, 56˚C for 30 s, and 72˚C for 30 s. The annealing temperature was based on Pereyra et al. (2010).

Multi-primer set quantification strategy for the pcrA gene

The number of *pcrA* gene copies per µL in genomic DNA extracts from each PRB was calculated using the equation above (N= 4,500 Kb for *D. aromatica* RCB). For *Dechloromonas* sp. PC1, the genome size was unknown and was estimated to be 4,500 Kb as is the case for *D. aromatica*. Also, *D. aromatica* RCB and *Dechloromonas* sp. PC1 were assumed to have one copy of *pcrA* per genome as is the case for *D. aromatica* (Bender et al., 2005). qPCR was run as described above. Standard curves for each strain-specific primer set were performed using template dilutions of 10^3 to 10^6 gene copies per reaction. For each primer set, the reactions containing 100% template of the strain targeted by that primer set was set at 100% quantification accuracy and used to do determine quantification accuracy for the other template mixtures.

Template mixtures containing purified genomic DNA from three PRB strains (*D. agitata*, *D. aromatica* RCB, and strain PC1) were tested at 10^3 and 10^6 gene copies per reaction. All tests were conducted in duplicate. The annealing temperature (55˚C) was selected by conducting endpoint gradient PCR for all three primer sets with annealing temperatures from 51˚C to 60˚C. Thermal cycling consisted of 95˚C for 15 min, followed by 40 cycles of 95˚C for 15 s, 55˚C for 30 s, and 72˚C for 30 s.

For the *agitata*-type and *aromatica*-type primer sets, qPCR template mixtures also included genomic DNA from *Dechlorosoma* sp. KJ. For strain KJ, the genome size was unknown and was estimated to be 4,500 Kb, and it was assumed that the genome contained one copy of *pcrA*. For the *aromatica*-type primer set, *D. aromatica* RCB DNA was used for the standard curve; for the *agitata*-type primer set *D. agitata* DNA was used for the standard curve. Also, due to the fact that the *aromatica*-type primer set contains multiple degeneracies, a range of primer concentrations (70 nM, 100 nM, 120 nM, and 140 nM) were tested for qPCR; a primer concentration of 140 nM was selected because it resulted in the most accurate quantification.

3.5 Analysis of Quantification Accuracy

The impact of primer mismatches on amplification efficiency, C_T lag, and quantification accuracy was determined for each primer combination. To calculate amplification efficiency, two different methods were used. First, C_T was plotted versus the log of the number of gene copies per reaction and the slope was calculated (Dorak, 2006). Efficiency was calculated from the slope according to the following equation.

$$\text{Amplification efficiency} = \left(\left(10^{\frac{-1}{slope}}\right) - 1\right) x\, 100\%$$

A slope of -3.322 equals 100% efficiency.

The same fluorescence threshold value and baseline was used for all tests conducted. Second, amplification efficiency was calculated for each individual reaction using a linear regression program (LinRegPCR) developed by Ramakers, et al. (2003) and efficiency values were averaged across dilutions using the program's included grouping function. The efficiency values from LinReg range from 1.0-2.0 with 2.0 representing 100% amplification efficiency. These numbers were converted to amplification efficiencies by exponentiating (base 10) the LinReg efficiency value of each mismatch primer reaction. The LinReg values were used to represent amplification efficiency for further calculations as the standard curve method occasionally produced values exceeding 100%.

Mismatch primer C_T lag and quantification accuracy were calculated based on results for perfect match primers run on the same plate to account for any plate to plate variability. Average C_T lag was calculated by averaging the differences between the C_T values of the mismatch primer reactions and the perfect match primer reactions for each template concentration. Although the *pcrA* gene from *D. agitata* was tested from 10 to 10^8 copies per reaction, reactions with 10 copies generally diverged from the trend line, and thus, were omitted in downstream analysis. To calculate the quantification accuracies, the measured quantity was divided by the quantity of template added to each reaction (gene copies per reaction) and multiplied by 100 to produce a percentage value using the equation below.

Quantification accuracy = $\frac{Q_m}{N}$ x 100

where Q_m = measured concentration using mismatch primers (gene copies per reaction)
N = concentration of template added to each reaction (gene copies per reaction)

Average quantification accuracies were calculated for the range of template concentrations used. Standard deviations for quantification accuracy were also calculated from all replicates and template concentrations for each primer set. Replicates and dilution points were treated the same in contributing to error. For combined quantification accuracies where the resulting accuracy is the summation of individual percentages detected from multiple data sets, such as the multi-primer quantification approach, error was propagated using the following equation, $\sigma_x = \sqrt{\sigma_a{}^2 + \sigma_b{}^2}$, where $x = a + b$ and σ represent the standard deviation for a data set (Ku, 1966). For combined quantification accuracies where the resulting accuracy is the product of individual results, such as the predicted effect of combining forward and reverse mismatch primers, error was propagated using the following equation, $\frac{\sigma_x}{x} = \sqrt{\left(\frac{\sigma_a}{a}\right)^2 + \left(\frac{\sigma_b}{b}\right)^2}$, where $x = a * b$.

4.0 RESULTS AND DISCUSSION

4.1 Single Mismatch Primers

Figure 4.1 shows quantification accuracy for single mismatch primers.

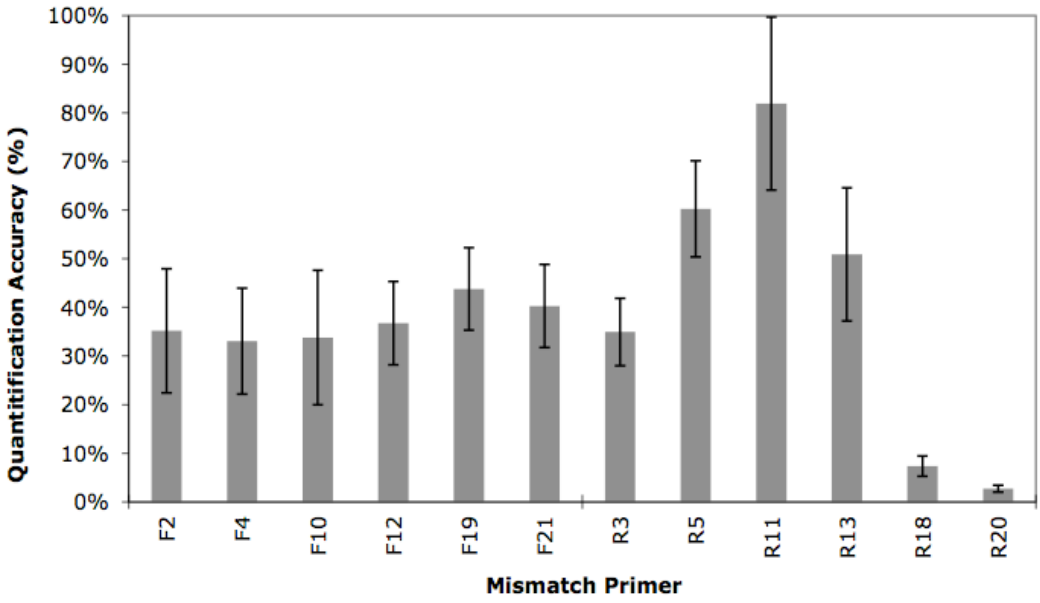

Figure 4.1. Quantification accuracy for single mismatch primers targeting the *pcrA* gene. Error bars represent standard deviation of quantification accuracy across all replicates and dilution points.

All single mismatches had a deleterious effect on quantification accuracy; calculated average accuracies ranged from 81.9% to 2.73%. Although there was typically some variation in accuracy across dilution points (e.g., ranging from 21.1-68.9% for F2, and 1.7-4.2% for R20), no consistent effect on accuracy of starting DNA concentration was identified for mismatch primers. Forward primer mismatches resulted in accuracies of 33.1-43.8%. The location of the mismatch within the forward primer, however, did not have a significant effect on quantification accuracy. In the reverse primer, results were

more varied. Quantification accuracy was highest (81.9%) for mismatches in the middle of the primer (e.g., R11). Mismatches near the 3' end of the reverse primer (R18 and R20) reduced the quantification accuracy by the greatest amount, resulting in measured quantities that were less than 10%. These results are consistent with previous studies of single primer mismatches. Bru et al. (2008) observed decreases in quantification accuracy of up to three orders of magnitude for primers with single mismatches near the 3' end of the primer and also observed different trends for forward and reverse primers. Bru et al. also showed that the same mismatches introduced into the template DNA, via PCR amplifying with each mismatch primer set, and tested with perfect match primers (i.e., primers that matched the original unaltered template) resulted in the same decreases in accuracy. Boyle et al. (2009) examined mismatches resulting from single nucleotide polymorphisms (SNPs) and observed quantification accuracies of 50% and 10% for mismatches introduced in the middle and 3' end of the primer, respectively. Süss et al. (2009) investigated single mismatch scenarios that showed a reduction in quantification accuracy to as low as 63.1% when mismatches were introduced in the last base at the 3' end of the primer.

In all three of the above studies, however, mismatches did not invariably reduce quantification accuracy. Mismatches near the 5' end and in the middle of the primer occasionally resulted in a measured quantity that was equivalent to or greater than the quantity measured with perfect match primers. While our results showed similar losses in accuracy to those of previous studies for mismatches occurring at the 3' end, it was shown that mismatches occurring in the middle of the primer and the 5' end could produce reductions in accuracy much greater than those previously reported.

4.2 Multiple Mismatch Primers

Figure 4.2 shows quantification accuracy for primers containing multiple mismatches.

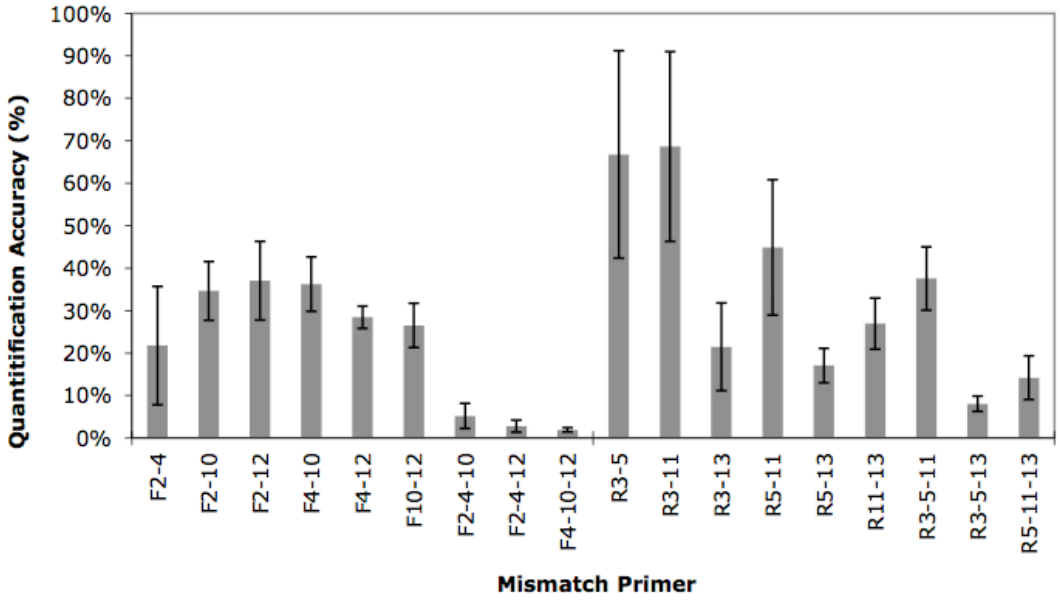

Figure 4.2. Quantification accuracy for multiple mismatch primers targeting the *pcrA* gene. Error bars represent standard deviation of quantification accuracy for all replicates and dilution points.

For the forward primer, in all cases multiple mismatches located only in the 5'-end and middle of the primers, which were shown to be less severe than 3'-end mismatches, resulted in accuracies below 37.0%. Double mismatch primers resulted in accuracies of 21.8-37.0% for the forward primer. Triple mismatches in the forward primer resulted in the greatest reduction of accuracy (0.8-2.3%). Multiple mismatches in the reverse primer had more varied effects than multiple mismatches in the forward primer with double mismatches resulted in accuracy ranging from 17.1-68.7%. However, triple mismatches again resulted in the greatest reduction of quantification accuracy (4.2-26.5%). Similarly, Sipos et al. (2007) observed that when universal bacterial 16S rRNA

gene targeted primers were used, 16S rRNA genes that perfectly matched the primers were preferentially amplified compared to 16S rRNA genes that contained three mismatches near the 5' end of one primer, although the quantitative effect of the mismatches was otherwise not reported. Guy et al. (2004) showed that multiple mismatches in primers (one 3' end mismatch in the forward primer and two 5' end mismatches in the reverse primer) targeting *Giardia Lamblia* resulted in increased C_T values and reductions in quantification accuracy of up to 2-4 orders of magnitude. These substantial reductions in accuracy are likely due to the assay containing mismatches in both primer directions as well as having mismatches in the probe (TaqMan). Smith et al. (2002) investigated the 5' exonuclease assay and showed that multiple mismatches had a greater detrimental effect on qPCR efficiency than single mismatches, and that as the number of mismatches increased, the effect on amplification efficiency became more pronounced, with 2 or more mismatches reducing amplification efficiency up to half the original value.

4.3 Accuracy Correlation to Amplification Efficiency and C_T Lag

The literature regarding the impact of primer mismatches on qPCR shows a general lack of consensus regarding how to measure and report detrimental effects of primer mismatches, and the mechanisms responsible for the reductions in quantification accuracy are not well understood. In some cases, the effect of primer mismatches has been evaluated by the calculated amplification efficiency (Klein et al., 1999; Smith et al., 2002). In other studies, threshold cycle (C_T) values or product yields are compared (Bru et al., 2008; Guy et al., 2004; Kwok et al., 1990; Whiley and Sloots, 2005). Thus, herein

statistical analysis was conducted to examine if amplification efficiency or average C_T lag are correlated with quantification accuracy. Amplification efficiencies determined via LinReg were used because previous studies have shown higher reproducibility with this method (Karlen et al., 2007). Average C_T lag was found to be inversely correlated with the quantification accuracy (Figure 4.3).

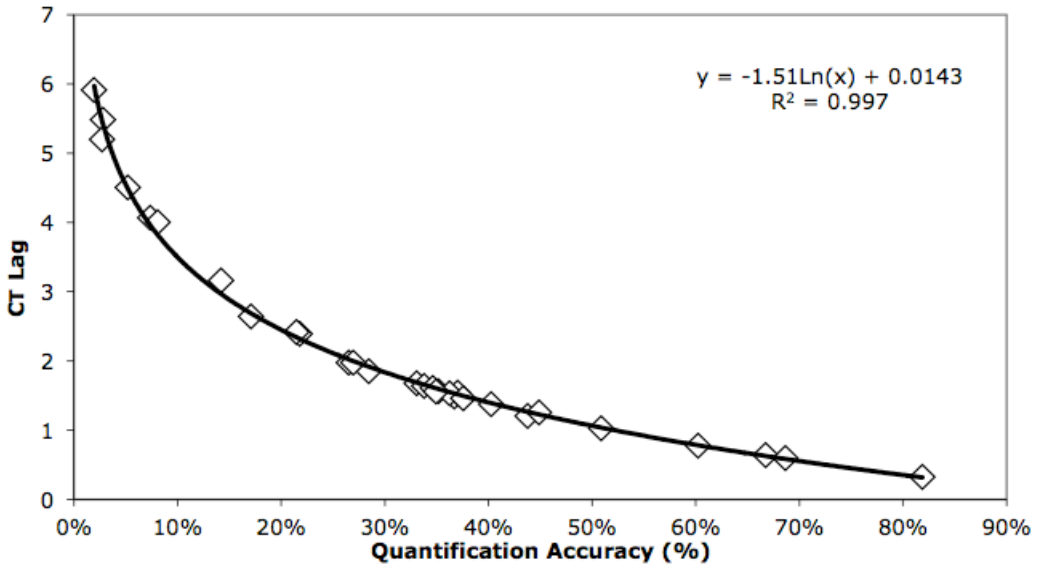

Figure 4.3. Quantification accuracy vs. C_T lag. The data fit a logarithmic curve and are highly correlated (R^2=0.997).

Amplification efficiency did not correlate with quantitative accuracy nor did it appear to follow any kind of trend based on the location or number of mismatches (Figure 4.4).

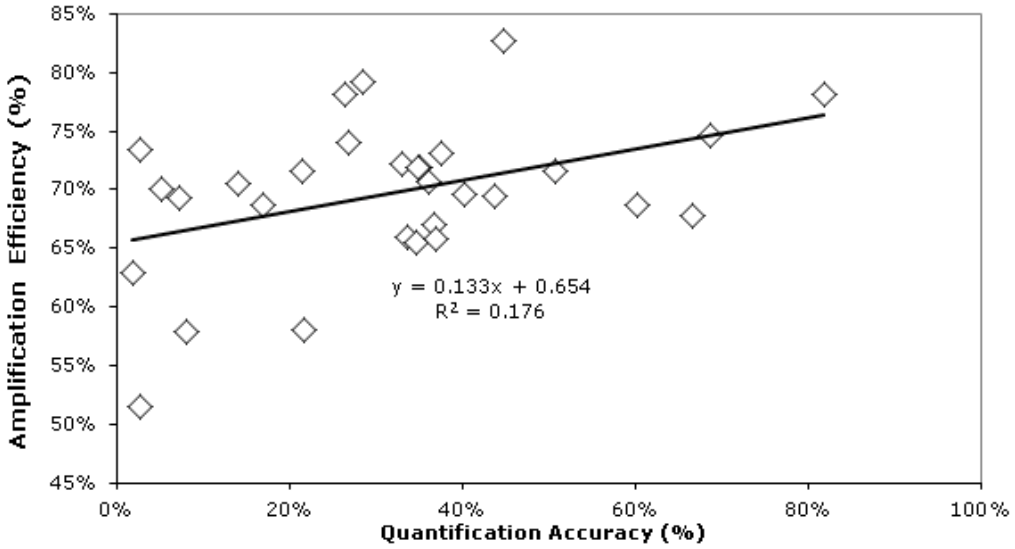

Figure 4.4. Quantification accuracy vs. amplification efficiency. The data shows poor correlation (R^2=0.176).

Although the correlation between C_T lag and quantification accuracy was anticipated as C_T values are used to determine quantity, the lack of correlation between amplification efficiency and accuracy was not expected. Based on the LinReg results for each dilution point, the efficiency also showed no consistent trend due to the starting copy number of template, and adjusting the threshold value (± 50%) did not change the poor correlation. This result suggests that the dominant mechanism by which mismatches reduce quantification accuracy is inefficiencies in primer binding during the initial rounds of qPCR, leading to a delay of the exponential phase of amplification (i.e., increased C_T lag). This finding is consistent with research conducted on PCR amplification (Nogva and Rudi, 2004), which has suggested that the initial cycles of PCR are dominated by the original DNA template. However, by the sixth cycle, the amplicon already outnumbers the original template by a ratio of 26:1 according to the theoretical doubling rate. Due to the fact that mismatches in the primer sequence are incorporated into the amplicon during

PCR, the primers perfectly match the dominant template after the first few rounds of amplification and should essentially amplify with 100% efficiency from that point.

4.4 Combining Mismatch Forward and Reverse Primers

Primer sets often contain mismatches in both the forward and reverse primers. A recent study found that for primers containing single mismatches, the effect of mismatches in both primers is the product of the effect of the individual mismatch primers (Boyle et al., 2009). For example, if mismatches in the forward primer result in a quantification accuracy of 80% (20% reduction), and mismatches in the reverse primer also result in a quantification accuracy of 80% (20% reduction), when these mismatch primers are used together, the resulting quantification accuracy will be 64%. Thus, herein this theory was tested for application to primers with double or triple mismatches using a subset of the *pcrA* mismatch primers. Experimental measured quantification accuracies for combinations of mismatch primers were found to be highly correlated with predicted quantification accuracies calculated according to Boyle et al. (2009) (Figure 4.5).

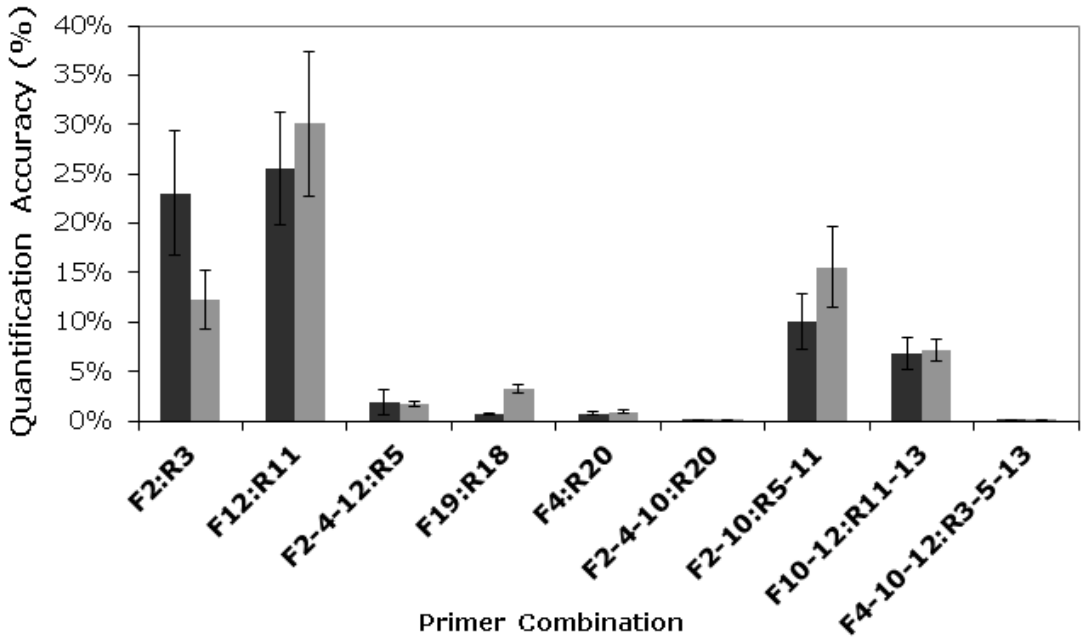

Figure 4.5. Experimentally measured and predicted quantification accuracies for mismatch primer combinations. Black bars represent experimentally measured accuracy, and gray bars represent predicted accuracy based on the effect of the individual mismatch primers. Error bars represent the standard deviation for the measured accuracies and the combined standard deviations of the individual mismatch primers propagated by multiplication.

Measured quantification accuracy and predicted accuracy were found to be linearly correlated (R^2=0.798). C_T lags also were predicted (labeled ΔC_q in Boyle et al., 2009) by adding the average C_T lag values of the individual mismatch primer reactions; and measured and predicted C_T lag was found to be linearly correlated (R^2=0.949). Thus, results suggest that quantification accuracy for combinations of double and triple mismatch primers can be predicted based on the results for mismatch primers tested with their respective perfect match primers.

4.5 Additional Model Genes: *C23O*, *hydA*, and *celS* genes

The effect of mismatch primers was evaluated for three additional model genes (the *C23O* gene, the *hydA* gene, and the *celS* gene) by modifying published primers targeting these genes. In all cases, the primers targeting the additional model genes did not meet optimal qPCR design criteria (e.g. secondary structures were present) presumably because these suboptimal features were unavoidable. Therefore, the results for these additional model genes are more representative of assays that are likely to be applied in environmental systems.

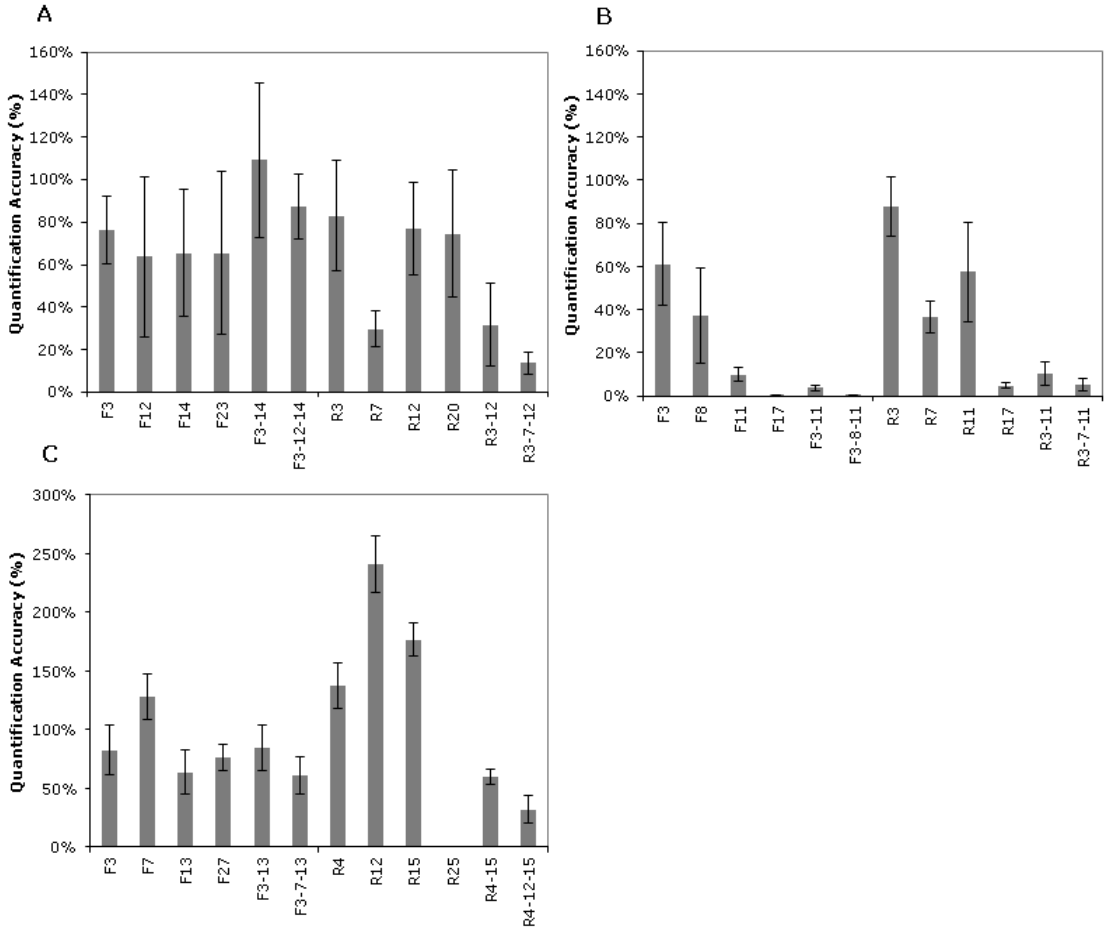

Figure 4.6. Quantification accuracy of mismatch primers targeting *C23O* in *P. putida* mt-2 (A), *hydA* in *S. oneidensis* MR-1 (B), and *celS* in *C. thermocellum* (C). Error bars represent standard deviation for quantification accuracy across all replicates and dilution points.

For the three additional model genes tested, single 5' end mismatches resulted in quantification accuracies ranging from 61.0 % to 137% (median of 82.8%) as compared to 33.1% to 60.2% (median of 35.1%) for *pcrA*. Single mismatches in the middle of the primers for the three additional model genes resulted in quantification accuracies ranging from 9.7% to 241% (median of 63.8%) as compared to 36.7% to 81.9% (median of 43.8%) for *pcrA*. For the three additional model genes tested, single 3' end mismatches resulted in quantification accuracies ranging from 0% to 76.0% (median of 35.0%) as compared to 2.7% to 43.8% (median of 23.8%) for *pcrA*. Double mismatches in the 5' end or middle of the primers for the three additional model genes resulted in quantification accuracies ranging from 3.8% to 109% (median of 45.9%) as compared to 17.1% to 68.7% (median of 31.6%) for *pcrA*. Triple mismatches for the three additional model genes tested resulted in quantification accuracies ranging from 0.1% to 87% (median of 22.8%) as compared to 2.0% to 37.6% (median of 6.7%) for *pcrA*. Although these additional model systems displayed considerable variation in their results, mismatches near the 3' end and triple mismatches ultimately showed the lowest median accuracies in both the additional model genes and *pcrA*, while single and double mismatches at the 5' end/middle of the primer resulted in smaller median accuracy losses, but still showed a significant effect.

Variability in the results across the three additional model genes might be explained by the fact that these primer sets were designed for targeting different types of real gene systems (pure vs. mixed culture) and were not explicitly designed with qPCR assay optimization guidelines in mind. Some mismatches resulted in accuracies higher than 100% most likely due to the fact that the nearly all the perfect match primers for

these systems showed deviations from optimal primer design guidelines and therefore expressed suboptimal behavior. Despite our efforts to introduce mismatches without modifying secondary structure or other deviations, some mismatches may have resulted in improved performance. The observations derived from these experiments indicate that the results achieved with the mismatch primers targeting *pcrA* from *D. agitata* are applicable beyond that strain, gene, and primer set. Although individual results varied widely depending on the gene system and primer set, mismatches resulted in a maximum median accuracy of 82.8% for single 5' end mismatches and a minimum median accuracy of 22.8% for triple 5' end/middle mismatches, indicating that caution should be taken when accepting mismatches in any quantity or location.

4.6 Multi-Primer Set Quantification Strategy for *pcrA*

Due to the fact that environmentally relevant functional genes often have diverse sequences, designing a single set of degenerate primers without mismatches is challenging; this is the case for *pcrA* genes found in PRB. Since the results of this study have indicated that even single mismatches in the 5'end or middle of the primer can reduce quantification accuracy, the *pcrA* gene was used as a model system to develop an alternative quantification strategy involving the use of multiple primer sets that can be used in combination. To determine if this multi-primer set approach was a feasible strategy for improving quantification accuracy, template mixtures containing *pcrA* genes from three PRB strains (*D. agitata*, *D. aromatica* RCB, *Dechloromonas* sp. PC1) possessing *pcrA* genes with sequence variations were quantified using three distinct perfect match primer sets each targeting the *pcrA* gene in one of the strains (Figure 4.7).

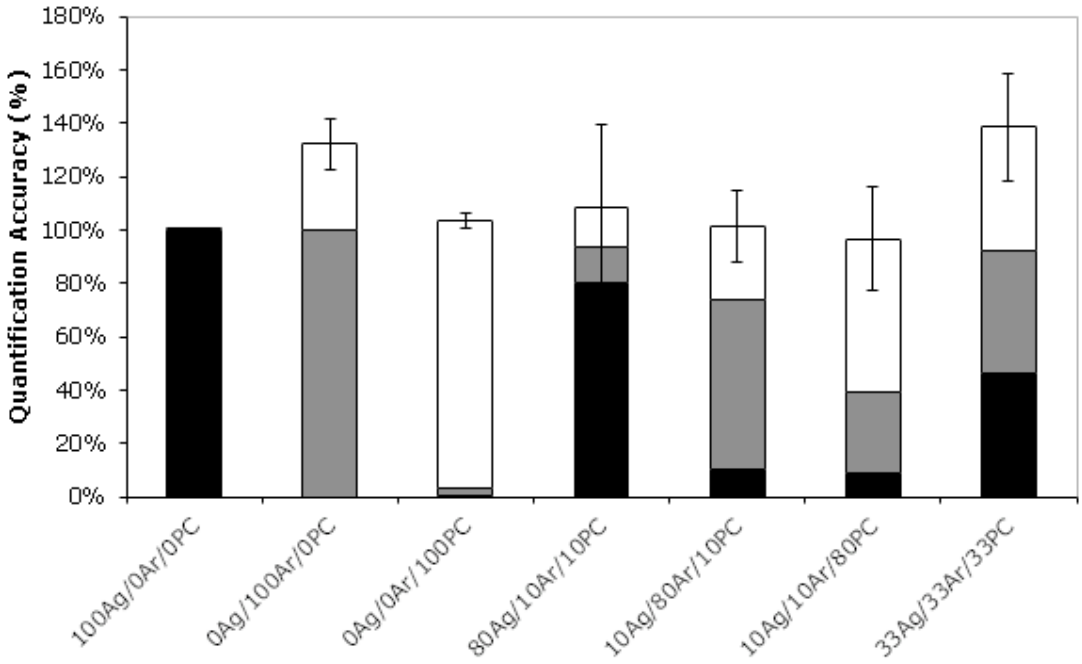

Figure 4.7. Multi-primer set quantification strategy targeting *pcrA* in artificially generated mixed microbial samples using three PRB strains. The composition of the template mixtures is indicated below each column. Ag indicates *D. agitata*, Ar indicates *D. aromatica* RCB, and PC indicates *Dechloromonas* sp. PC1. The numbers in front of each two-letter strain code indicate the percentage of the gene copies in the template attributable to that strain. For each template mixture, all three qPCR assays were run (i.e., assays were run with all perfect match primer sets) and the percentages of template quantified were added for all three assays. The columns show the contributions from testing the mixtures with perfect match primers targeting *pcrA* in *D. agitata* (black), *D. aromatica* RCB (gray), and strain PC1 (white). Error bars represent the combined propagated standard deviations from each primer set. The error bar for the 100Ag/0Ar/0PC mixture was too small to be visible on the graph.

With the exception of two template mixtures (0Ag/100Ar/0PC and 33Ag/33Ar/33PC), this multi-primer set strategy resulted in quantification accuracies that were within 10% of perfect quantification. The template mixture containing 100% *D. aromatica* RCB DNA resulted in an overestimation of the *pcrA* gene quantity (132% quantification accuracy) due to cross-detection by the PC1-targeted primer set. It is unclear why the template mixture containing equal amounts of each template (33Ag/33Ar/33PC) resulted in an overestimation of the *pcrA* gene quantity (138% quantification accuracy).

Individual contributions from each strain were all slightly higher than expected values (within 23% of expected template percentage). Results suggest that if cross-detection can be avoided, analyzing mixed culture samples with multiple primer sets can improve quantification accuracy as compared to the single primer set approach.

However, the particular PRB strains present in environmental samples are rarely known, and strains present may have *pcrA* sequences that diverge from known *pcrA* sequences. Therefore, the three sets of perfect match primers are not suitable for environmental application. Furthermore, it is desirable to minimize the number of required qPCR assays to minimize cost for environmental applications. Thus, a quantification strategy was developed using only two sets of degenerate primers designed to target all known *pcrA* genes. This two-primer set strategy was directly compared against the previously published CODEHOP-derived *pcrA* qPCR assay (Nozawa-Inoue et al., 2008). Figure 4.8 shows the results of the existing primer set applied to artificially generated mixed microbial samples compared to the total quantification accuracies determined by adding the percentages of *pcrA* genes quantified for each of the two assays (qPCR assays with *agitata*-type and *aromatica*-type *pcrA* primers).

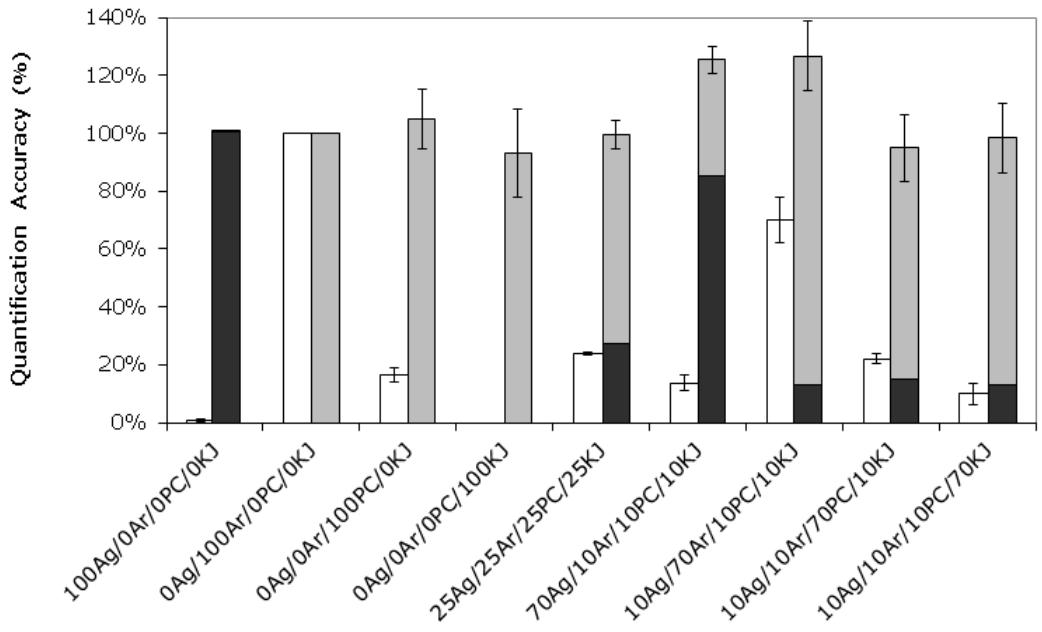

Figure 4.8. Multi-primer set quantification strategy targeting *pcrA* compared to an existing primer set targeting *pcrA* in artificially generated mixed microbial samples. The composition of the template mixtures is indicated below each column. Ag indicates *D. agitata*, Ar indicates *D. aromatica* RCB, PC indicates *Dechloromonas* sp. PC1, and KJ indicates *Dechlorosoma* sp. KJ. The numbers in front of each two-letter strain code indicate the percentage of the gene copies in the template attributable to that strain. The columns represent the results of the pcrA320 and pcrA598 primers (white) and the combined results from the *agitata*-type (black) and *aromatica*-type primer sets (gray). Error bars represent the combined propagated standard deviations from each primer set.

With the exception of two template mixtures (70Ag/10Ar/10PC/10KJ and 10Ag/70Ar/10PC/10KJ), the two-primer set strategy resulted in quantification accuracies that were within 10% of perfect quantification. Contributions from each primer set were within 10.5% of the expected values, with the exception of the two overestimated mixtures, which were within 23.7% of the expected accuracy values. For the overestimated mixtures, the primer set targeting the dominant strain (*agitata*-type for the 70Ag mixture and *aromatica*-type for the 70Ar mixture) was more prone to overestimation. The multi-primer set strategy achieved more accurate quantification than the previously available assay (i.e., the percent difference between the measured and

expected values was smaller for the multi-primer set method than for the single, degenerate primer strategy) for all template mixtures except 0Ag/100Ar/0PC/0KJ, which was quantified equally well by both approaches.

Genomic information for additional PRB possessing *pcrA* was released after primers were designed for this study. In order to verify the utility of the *agitata*-type and *aromatica*-type primers for targeting all existing variations of *pcrA*, the primer sequences designed herein were compared to the newly submitted *pcrA* sequences. The *pcrA* gene from *Dechlorosoma* sp. JD125 and *Dechloromonas* sp. JD15 (Peng, unpublished) both possessed a perfect match to the forward and reverse *agitata*-type primers. The *pcrA* gene from *Dechlorosoma suillum* PS (Melnyk et al., 2011) possessed a perfect match to both forward and reverse *aromatica*-type primers, based on degenerate bases. These sequence matches indicate the utility of the *agitata*- and *aromatica*-type primers beyond the variations of *pcrA* evaluated in this study.

Potential for application of the multi-primer set quantification strategy to real systems

The locations of the degenerate bases in the primers designed herein are the same as for the previously published CODEHOP-derived primers with the exception of one degeneracy in the reverse primer (occurs at the 21st base as opposed to the 20th base for the existing primer). CODEHOP adds degeneracies near the 3' end to account for the degeneracy of the genetic code. Because the primers designed herein include essentially the same degeneracies as those selected by CODEHOP, these new primers likewise account for probable sequence variations of *pcrA* encoding the same amino acid sequence.

The specificity of the degenerate *aromatica*-type primer set was assured by designing these primers to target two regions (forward and reverse primer binding sites) of the *pcrA* gene corresponding to regions of the PcrA protein that do not share amino acid similarity with closely related non-target genes. These genes include chlorate reductase (ClrA; GenBank accession no. CAD97447), selenate reductase (SerA; AJ007744), dimethylsulfide dehydrogenase (DdhA; AF453479), ethylbenzene dehydrogenase (EbdA; AF337952), and nitrate reductase (NarG; NP415742). These regions, identified by Nozawa-Inoue et al. (2008), contain a total of at least 9 out of 15 amino acid differences when compared to the PcrA sequence from *D. aromatica* RCB, and therefore would have sufficiently differing gene sequences to prevent amplification with the *aromatica*-type primer set.

The objectives of this study were primarily focused on systems where the locations of mismatches or sequence variations are known. However, it should be noted that in many environmental systems, gene sequence information is unavailable or incomplete. In order to identify locations of sequence variation and to allow for the improved design of qPCR primers avoiding mismatches, metagenomic techniques, such as pyrosequencing, could be employed for environmental samples.

Guidance for designing qPCR primer sets targeting genes with sequence variations

As is evident in the results for *pcrA* genes with sequence variations, the mismatch avoidance and multi-primer set strategy shows a clear advantage over the existing methods that guide qPCR primer design. Based on the study findings, we have developed some recommendations as additional guidance when designing new primer sets.

The primary recommendation is to avoid primer-template mismatches; this study has shown that mismatches can reduce the quantification accuracy of primers to as low as 0.8%, and mismatches previously thought tolerable were shown to have significant effects as well (median accuracy of 36.0% for single 5' end/middle mismatches in *pcrA*). For target genes showing a high degree of gene sequence variation as in the case of *pcrA*, it may be necessary to utilize a multi-primer set approach as developed herein. Minimize the number of primer sets required by introducing degeneracies, the locations of which can be identified by aligning gene sequences at highly conserved regions in genetic or protein sequences. There is no generally accepted maximum for primer degeneracy (total number of primer variations), although based on existing primer sets (Bender et al., 2004), degeneracy tends to not exceed a value of 500. When developing a primer set possessing a high degree of degeneracy, specificity must be checked on a case-by-case basis. If the primer sets are shown to have minimal cross-detection, quantitative results from each set can be combined to achieve improved accuracy.

5.0 CONCLUSIONS

Primer mismatches are a nearly ubiquitous challenge when targeting mixed microbial samples in real systems, particularly in environmental engineering. One major question in primer design has been which mismatch quantities and locations are acceptable or tolerable. Although mismatches may not inhibit PCR detection entirely, it is well accepted that mismatches can reduce quantification qPCR assays. However, there has been no consensus on how single and multiple mismatches impact quantitative accuracy, and some past studies have suggested that single mismatches may have no effect on quantification at all. While it is generally accepted that single mismatches near the 5' end and middle of the primer are tolerable, this study has shown that single mismatches at any location in the primer sequence can have a significant impact on quantification accuracy. Multiple 5' end/middle mismatches and mismatches near the 3' end of the primer generally resulted in the largest accuracy losses. In order to overcome the effect of mismatches when targeting a model gene with sequence variations, multiple primer sets were designed to target the *pcrA* gene in PRB using degeneracies to avoid mismatches. By combining the results for the two qPCR assays targeting distinct variants of *pcrA*, quantification accuracy near 100% was achieved for template mixtures, and this method showed a significant improvement over the previously published single primer set assay. Thus, the multi-primer set strategy shows distinct advantages over existing single primer set approaches and its application to other environmentally relevant genes holds promise for increasing the accuracy and utility of qPCR assays for environmentally relevant target genes with sequence variations.

REFERENCES

Bauchop, T. and S.R. Eldsen. 1960. The growth of micro-organisms in relation to their energy supply. J. Gen. Microbiol. **23**:457-469.

Bender, K.S., M.R. Rice, W.H. Fugate, J.D. Coates, and L.A. Achenbach. 2004. Metabolic primers for detection of (per) chlorate-reducing bacteria in the environment and phylogenetic analysis of *cld* gene sequences. Appl. Environ. Microbiol. **70**:5651-5658.

Bender, K.S., C. Shang, R. Chakraborty, S.M. Belchik, J.D. Coates, and L.A. Achenbach. 2005. Identification, characterization, and classification of genes encoding perchlorate reductase. J. Bacteriol. **187**:5090-5096.

Boyle, B., N. Dallaire, and J. MacKay. 2009. Evaluation of the impact of single nucleotide polymorphisms and primer mismatches on quantitative PCR. BMC Biotechnology **9**:75-90.

Braker, G., A. Fesefeldt, and K.P. Witzel. 1998. Development of PCR primer systems for amplification of nitrite reductase genes (*nirK* and *nirS*) to detect denitrifying bacteria in environmental samples. Appl. Environ. Microbiol. **64**:3769-3775.

Braker, G., J. Zhou, W. Liyou, A.H. Devol, and J.M. Tiedje. 2000. Nitrite reductase genes (*nirK* and *nirS*) as functional markers to investigate diversity of denitrifying bacteria in Pacific Northwest marine sediment communities. Appl. Environ. Microbiol. **66**:2096-2104.

Bru, D., F. Martin-Laurent, and L. Philippot. 2008. Quantification of the detrimental effect of a single primer-template mismatch by real-time PCR using the 16S rRNA gene as an example. Appl. Environ. Microbiol. **74**:1660-1663.

Burpo, F.J. 2001. A critical review of PCR primer design algorithms and cross-hybridization case study. Biochemistry **218**:1-12.

Chaudhuri, S.K., S.M. O'Connor, S.L. Gustavson, L.A. Achenbach, and J.D. Coates. 2002. Environmental factors that control microbial perchlorate reduction. Appl. Environ. Microbiol. **68**:4425-4430.

Chen, J., F.F. Kadlubar, and J.Z. Chen. 2007. DNA supercoiling suppresses real-time PCR: A new approach to the quantification of mitochondrial DNA damage and repair. Nucleic Acids Res. **35**:1377-1388.

Chin, K.J., M.J. Sharma, L.A. Russell, K.R. O'Neill, and D.R. Lovley. 2008. Quantifying expression of a dissimilatory (bi)sulfite reductase gene in petroleum-contaminated marine harbor sediments. Microb. Ecol. **55**:489-499.

Coates, J.D., U. Michaelidou, R.A. Bruce, S.M. O'Connor, J.N. Crespi, and L.A. Achenbach. 1999. Ubiquity and diversity of dissimilatory (per)chlorate-reducing bacteria. Appl. Environ. Microbiol. **65**:5234-5241.

Coates, J.D., K.A. Cole, R. Chakraborty, S.M. O'Connor, and L.A. Achenbach. 2002. Diversity and ubiquity of bacteria capable of utilizing humic substances as electron donors for anaerobic respiration. Appl. Environ. Microbiol. **68**:2445-2452.

Coyne, K.J., S.M. Handy, E. Demir, E.B. Whereat, D.A. Hutchins, K.J. Portune, M.A. Doblin, and S.C. Cary. 2005. Improved Quantitative Real-Time PCR Assays for Enumeration of Harmful Algal Species in Field Samples Using an Exogenous DNA Reference Standard. Limnol. Oceanogr.: Methods **3**:381-391.

De Long, S.K., K.A. Kinney, and M.J. Kirisits. 2010. qPCR assays to quantify genes and gene expression associated with microbial perchlorate reduction. J. Microbiol. Methods **83**:270-274.

Devers, M., G. Soulas, and F. Martin-Laurent. 2004. Real-time reverse transcription PCR analysis of expression of atrazine catabolism genes in two bacterial strains isolated from soil. J. Microbiol. Methods **56**:3-15.

Draghici, S., P. Khatri, A.C. Eklund, and Z. Szallasi. 2006. Reliability and reproducibility issues in DNA microarray measurements. Trends in Genetics **22**:101:109.

Dorak, M.T. (Ed.). 2006. Real-Time PCR (Advanced Method Series). Taylor and Francis, Oxford.

Forster, T., D. Roy, and P. Ghazal. 2003. Experiments using microarray technology: limitations and standard operating procedures. J. Endocrin. **178**:195-204.

Ginzinger, D.G. 2002. Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. Experimental Hematology **30**:503:512.

Guy, R.A., C. Xiao, and P.A. Horgen. 2004. Real-time PCR assay for detection and genotype differentiation of *Giardia lamblia* in stool specimens. J. Clin. Microbiol. **42**:3317-3320.

Hall, S.J., P. Hugenholtz, N. Siyambalapitiya, J. Keller, and L.L. Blackall. 2002. The development and use of real-time PCR for the quantification of nitrifiers in activated sludge. Water Sci. Technol. **46**:267-272.

Hallin, S. and P.E. Lindgren. 1999. PCR detection of genes encoding nitrite reductase in denitrifying bacteria. Appl. Environ. Microbiol. **65**:1652-1657.

Harms, G., A.C. Layton, H.M. Dionisi, I.M. Gregory, V.M. Garrett, S.A. Hawkins, K.G. Robinson, and G.S. Sayler. 2003. Real-time PCR quantification of nitrifying bacteria in a municipal wastewater treatment plant. Environ. Sci. Technol. **37**:343–351.

Henry, S., E. Baudoin, J.C. Lopez-Gutierrez, F. Martin-Laurent, A. Brauman, and L. Philippot. 2004. Quantification of denitrifying bacteria in soils by *nirK* gene targeted real-time PCR. J. Microbiol. Methods **59**:327-335.

Higashioka, Y., H. Kojima, S. Sato, and M. Fukui. 2009. Microbial community analysis at crude oil-contaminated soils targeting the 16S ribosomal RNA, *xylM*, *C23O*, and *bcr* genes. J. Appl. Microbiol. **107**:126-135.

Higuchi, R., C. Fockler, G. Dollinger, and R. Watson. 1993. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Biotechnology **11**:1026-1030.

Holmes, V.F., J. He, P.K.H. Lee, and L. Alvarez-Cohen. 2006. Discrimination of multiple *Dehalococcoides* strains in a trichloroethene enrichment by quantification of their reductive dehalogenase genes. Appl. Environ. Microbiol. **72**:5877-5883.

Johnson, D.R., P.K.H. Lee, V.F. Holmes, and L. Alvarez-Cohen. 2005. An internal reference technique for accurately quantifying specific mRNAs by real-time PCR with application to the *tceA* reductive dehalogenase gene. Appl. Environ. Microbiol. **71**:3866-3871.

Johnson, D.R., P.K.H. Lee, V.F. Holmes, A.C. Fortin, and L. Alvarez-Cohen. 2005. Transcriptional expression of the *tceA* gene in a *Dehalococcoides*-containing microbial enrichment. Appl. Environ. Microbiol. **71**:7145-7151.

Jung, S. and J.M. Regan. 2010. Influence of external resistance on electrogenesis, methanogenesis, and anode prokaryotic communities in microbial fuel cells. Appl. Environ. Microbiol. **77**:564-571.

Karlen, Y., A. McNair, S. Perseguers, C. Mazza, and N. Mermod. 2007. Statistical significance of quantitative PCR. BMC Bioinformatics **8**:131-147.

Klein, D., P. Janda, R. Steinborn, M. Müller, B. Salmons, and W.H. Günzburg. 1999. Proviral load determination of different feline immunodeficiency virus isolates using real-time polymerase chain reaction: Influence of mismatches on quantification. Electrophoresis **20**:291-299.

Klein, D. 2002. Quantification using real-time PCR technology: applications and limitations. Trends in Molec. Med. **8**:257-260.

Komon-Zelazowska, M., J. Bissett, D. Zafari, L. Hatvani, L. Manczinger, S. Woo, M. Lorito, L. Kredics, C.P. Kubicek, and I.S. Druzhinina. 2007. Genetically closely related but phenotypically divergent *Trichoderma* species cause green mold disease in oyster mushroom farms worldwide. Appl. Environ. Microbiol. **73**:7415-7426.

Koskenniemi, K., C. Lyra, P. Rajaniemi-Wacklin, J. Jokela, K. Sivonen. 2007. Quantitative real-time PCR detection of toxic *Nodularia* cyanobacteria in the baltic sea. Appl. Environ. Microbiol. **73**:2173-2179.

Ku, H. 1966. Notes on the use of the propagation of error formulas. J. Res. NBS C. Eng. Instrum. **70C**:263-273.

Kubista, M., J.M. Andrade, M. Bengtsson, A. Forootan, J. Jonák, K. Lind, R. Sindelka, R. Sjöback, B. Sjögreen, L. Strömbom, A. Ståhlberg, and N. Zoric. 2006. The real-time polymerase chain reaction. Molec. Aspects Med. **27**:95-125.

Kwok, S., D.E. Kellogg, N. McKinney, D. Spasic, L. Goda, C. Levenson, and J.J. Sninsky. 1990. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. Nucleic Acids Res. **18**:999-1005.

Laverick, M.A., A.P. Wyn-Jones, and M.J. Carter. 2004. Quantitative RT-PCR for the enumeration of noroviruses (Norwalk-like viruses) in water and sewage. Lett. Appl. Microbiol. **39**:127-136.

Leininger, S., T. Urich, M. Schloter, L. Schwark, J. Qi, G.W. Nicol, J.I. Prosser, S.C. Schuster, and C. Schleper. 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature Lett. **442**:806-809.

Li, D., A.Z. Gu, W. Yang, M. He, X. Hu, and H.C. shi. 2010. An integrated cell culture and reverse transcription quantitative PCR assay for detection of infectious rotaviruses in environmental waters. J. Microbiol. Methods **82**:59-63.

Long, A.D., H.J. Mangalam, B.Y.P. Chan, L. Tolleri, G.W. Hatfield, and P. Baldi. 2001. Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. J. Biol. Chem. **276**:19937-19944.

Lowe, T., J. Sharefkin, S.Q. Yang, and C.W. Dieffenbach. 1990. A computer program for selection of oligonucleotide primers for polymerase chain reaction. Nucleic Acids Res. **18**:1757-1761.

Maniatis T., E.F. Fritsch, J. and Sambrook. 1982. Molecular Cloning: A Labortory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Martin-Laurent, F., L. Philippot, S. Hallet, R. Chaussod, J.C. Germon, G. Soulas, and G. Catroux. 2001. DNA extraction from soils: Old bias for new microbial diversity analysis methods. Appl. Environ. Microbiol. **67**:2354–2359.

Melnyk, R.A., A. Engelbrektson, I.C. Clark, H.K. Carlson, K. Byrne-Bailey, and J.D. Coates. 2011. Identification of a perchlorate reduction genomic island with novel regulatory and metabolic genes. Appl. Environ. Microbiol. **77**:7401-7404.

Meshulam-Simon, G., S. Behrens, A.D. Choo, and A.M. Spormann. 2007. Hydrogen metabolism in *Shewanella oneidensis* MR-1. Appl. Microbiol. **73**:1153-1165.

N'Guessan, A.L., H. Elifantz, K.P. Nevin, P.J. Mouser, B. Methé, T.L. Woodard, K. Manley, K.H. Williams, M.J. Wilkins, J.T. Larsen, P.E. Long, and D.R. Lovley. 2010. Molecular analysis of phosphate limitation in *Geobacteraceae* during the bioremediation of a uranium-contaminated aquifer. ISME J. **4**:253-266.

Nocker, A., K.E. Sossa, and A.K. Camper. 2007. Molecular monitoring of disinfection efficacy using propidium monoazide in combination with quantitative PCR. J. Microbiol. Methods **70**:252-260.

Nogva, H.K., K. Rudi, K. Naterstad, A. Holck, and D. Lillehaug. 2000. Application of 5′-nuclease PCR for quantitative detection of *Listeria monocytogenes* in pure cultures, water, skim milk, and pasteurized whole milk. Appl. Environ. Microbiol. **66**:4266-4271.

Nogva, H.K. and K. Rudi. 2004. Potential influence of the first PCR cycles in real-time comparative gene quantifications. BioTechniques **37**:246-253.

Nozawa-Inoue, M., M. Jien, N.S. Hamilton, V. Stewart, K.M. Scow, and K.R. Hristova. 2008. Quantitative detection of perchlorate-reducing bacteria by real-time PCR targeting the perchlorate reductase gene. Appl. Environ. Microbiol. **74**:1941-1944.

Oberst, R.D., M.P. Hays, L.K. Bohra, R.K. Phebus, C.T. Yamashiro, C. Paszko-Kolya, S.J.A. Flood, J.M. Sargeant, and J.R. Gillespie. 1998. PCR-based DNA amplification and presumptive detection of *Escherichia coli* O157:H7 with an internal fluorogenic probe and the 5′ nuclease TaqMan assay. Appl. Environ. Microbiol. **64**:3389-3396.

Pereyra, L.P., S.R. Hiibel, M.V. Prieto Riquelme, K.F. Reardon, and A. Pruden. 2010. Detection and quantification of functional genes of cellulose-degrading, fermentative, and sulfate-reducing bacteria and methanogenic archaea. Appl. Environ. Microbiol. **76**:2192-2202.

Philippot, L. 2006. Use of functional genes to quantify denitrifiers in the environment. Biochem. Soc. Trans. **34**:101-103.

Quéméneur, M., A. Cébron, P. Billard, F. Battaglia-Brunet, F. Garrido, C. Leyval, and C. Joulian. 2010. Population structure and abundance of arsenite-oxidizing bacteria along an arsenic pollution gradient in waters of the Upper Isle River Basin, France. Appl. Environ. Microbiol. **76**:4566-4570.

Ramakers, C., J.M. Ruijters, R.H. Lekanne Deprez, and A.F.M. Moorman. 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci. Lett. **339**:62-66.

Ren, Z., L.M. Steinberg, and J.M. Regan. 2008. Electricity production and microbial biofilm characterization in cellulose-fed microbial fuel cells. Water Sci. Technol. **58**:617-622.

Rinta-Kanto, J.M., A.J.A. Ouelette, G.L. Boyer, M.R. Twiss, T.B. Bridgeman, and S.W. Wilhelm. 2005. Quantification of toxic *Microcystis* spp. during the 2003 and 2004 blooms in western Lake Erie using quantitative real-time PCR. Environ. Sci. Technol. **39**:4198-4205.

Ritalahti, K.M., B.K. Amos, Y. Sung, Q. Wu, S.S. Koenigsburg, and F.E. Löffler. 2006. Quantitative PCR targeting 16S rRNA and reductive dehalogenase genes simultaneously monitors multiple *Dehalococcoides* strains. Appl. Environ. Microbiol. **72**: 2765-2774.

Rittmann, B.E. 2006. Microbial ecology to manage processes in environmental biotechnology. Trends in Biotechnol. **24**:261-266.

Rittmann, B.E. 2010. Environmental biotechnology in water and wastewater treatment. J. Environ. Eng. **136**:348-353.

Rose, T.M., E.R. Schultz, J.G. Henikoff, S. Pietrokovski, C.M. McCallum, and S. Henikoff. 1998. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. Nucleic Acids Res. **26**:1628-1635.

Saleh-Lakha, S., Shannon, K.E., Goyer, C. and J.T. Trevors. 2011. Challenges in quantifying microbial gene expression in soil using quantitative reverse transcription real-time PCR. J Microbiol Methods **85**:239-243.

Schaefer, C.E., X. Yang, O. Pelz, D. T. Tsao, S.H. Streger, and R.J. Steffan. 2010. Aerobic biodegradation of *iso*-butanol and ethanol and their relative effects on BTEX biodegradation in aquifer materials. Chemosphere **81**:1104-1110.

Schneegurt, M.A. and C.F. Kulpa. 1998. The application of molecular techniques in environmental biotechnology for monitoring microbial systems. Biotechnol. Appl. Biochem. **27**:73-79.

Sharma, S., V. Radl, B. Hai, K. Kloos, M. Mrkonjic Fuka, M. Engel, K. Schauss, and M. Schloter. 2007. Quantification of functional genes from procaryotes in soil by PCR. J. Microbiol. Methods **68**:445-452.

Sharp, J.O., C.M. Sales, J.C. LeBlanc, J. Liu, T.K. Wood, L.D. Eltis, W.W. Mohn, and L. Alvarez-Cohen. 2007. An inducible propane monooxygenase is responsiblefor *N*-nitrosodimethylamine degradation by *Rhodococcus* sp. strain RHA1. Appl. Environ. Microbiol. **73**:6930-6938.

Sipos, R., A.J. Székely, M. Palatinszky, S. Révész, K. Márialigeti, and M. Nikolausz. 2007. Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targeting bacterial community analysis. FEMS Microbiol. Ecol. **60**:341-350.

Smith, C.J., D.B. Nedwell, L.F. Dong, and A.M. Osborn. 2006. Evaluation of quantitative polymerase chain reaction-based approaches for determining gene copy and gene transcript numbers in environmental samples. Environ. Microbiol. **8**:804-815.

Smith, C.J. and A.M. Osborn. 2009. Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. FEMS Microbiol. Ecol. **67**:6-20.

Smith, S., L. Vigilant, and P.A. Morin. 2002. The effects of sequence length and oligonucleotide mismatches on 5' exonuclease assay efficiency. Nucleic Acids Res. **30**:111-121.

Song, M., S.G. Shin, and S. Hwang. 2010. Methanogenic population dynamics assessed by real-time quantitative PCR in sludge granule in upflow anaerobic sludge blanket treating swine wastewater. Bioresource Technol. **101**:S23-S28.

Steinberg, L.M. and J.M. Regan. 2009. *mcrA*-targeted real-time quantitative PCR method to examine methanogen communities. Appl. Environ. Microbiol. **75**:4435-4442.

Stults, J.R., O. Snoeyenbos-West, B. Methé, D.R. Lovley, and D.P. Chandler. 2001. Application of the 50 fluorogenic exonuclease assay (TaqMan) for quantitative ribosomal DNA and rRNA analysis in sediments. Appl. Environ. Microb. **67**:2781-2789.

Süss, B., G. Flekna, M. Wagner, and I. Hein. 2009. Studying the effect of single mismatches in primer and probe binding regions on amplification curves and quantification in real-time PCR. J. Microbiol. Methods **76**:316-319.

Tancsics, A., S. Szoboszlay, I. Szabo, M. Farkas, B. Kovacs, J. Kukolya, Z. Mayer, and B. Kriszt. 2012. Quantification of subfamily I.2.C catechol 2,3-dioxygenase mRNA transcripts in groundwater samples of an oxygen-limited BTEX-contaminated site. Environ. Sci. Technol. **46**:232-240.

Urbansky, E.T. 1998. Perchlorate chemistry: Implications for analysis and remediation. J. Bioremed. **2**:81-95.

Wahman, D.G., K.A. Wulfeck-Kleier, and J.G. Pressman. 2009. Monochloramine disinfection kinetics of *Nitrosomonas europaea* by propidium monoazide quantitative PCR and live/dead BacLight methods. Appl. Environ. Microbiol. **75**:5555-5562.

Whiley, D.M. and T.P. Sloots. 2005. Sequence variation in primer targets affects the accuracy of viral quantitative PCR. J. Clin. Virol. **34**:104-107.

White, H.K., C.E. Reimers, E.E. Cordes, G.F. Dilly, and P.R. Girguis. 2009. Quantitative population dynamics of microbial communities in plankton-fed microbial fuel cells. ISME J. **3**:635-646.

Wittwer, C.T., M.G. Herrmann, A.A. Moss, and R.P. Rasmussen. 1997. Continous fluorescence monitoring of rapid cycle DNA amplification. BioTechniques **22**:130-138.

Wolff, J. 1998. Perchlorate and the thyroid gland. Pharmacol. Rev. **50**:89-105.

Wu, L., D.K. Thompson, G. Li, R.A. Hurt, J.M. Tiedje, and J. Zhou. 2001. Development and evaluation of functional gene arrays for detection of selected genes in the environment. Appl. Environ. Microbiol. **67**:5780-5790.

Wu, Q., R.A. Sanford, and F.E. Löffler. 2006. Uranium(VI) reduction by *Anaeromyxobacter dehalogenans* strain 2CP-C. Appl. Environ. Microbiol. **72**:3608:3614.

Yu, Y., J. Kim, and S. Hwang. 2006. Use of real-time PCR for group-specific quantification of aceticlastic methanogens in anaerobic processes: Population dynamics and community structures. Biotechnol. Bioeng. **93**:424-433.

Zengler, K., G. Toledo, M. Rappé, J. Elkins, E.J. Mathur, J.M. Short, and M. Keller. 2002. Cultivating the uncultured. PNAS **99**:15681-15686.

Zhang, T. and H. H. P. Fang. 2006. Applications of real-time polymerase chain reaction for quantification of microorganisms in environmental samples. Appl. Microbiol. Biotechnol. **70**:281-289.