

# Deep Diversity: Novel Approach to Overcoming the PCR Bias Encountered During Environmental Analysis of Microbial Populations for $\alpha$ -Diversity

Gustavo A. Ramirez and Parag A. Vaishampayan

Jet Propulsion Laboratory, California Institute of Technology

**Alpha-diversity studies are of crucial importance to environmental microbiologists. The polymerase chain reaction (PCR) method has been paramount for studies interrogating microbial environmental samples for taxon richness. Phylogenetic studies using this technique are based on the amplification and comparison of the 16S rRNA coding regions. PCR, due to disproportionate distribution of microbial species in the environment, increasingly favors the amplification of the most predominant phylotypes with every subsequent reaction cycle. The genetic and chemical complexity of environmental samples are intrinsic factors that exacerbate an inherent bias in PCR-based quantitative and qualitative studies of microbial communities. We report that treatment of a genetically complex total genomic environmental DNA extract with Propidium Monoazide (PMA), a DNA intercalating molecule capable of forming a covalent cross-linkage to organic moieties upon light exposure, disproportionately inactivates predominant phylotypes and results in the exponential amplification of previously shadowed microbial  $\alpha$ -diversity quantified as a 19.5% increase in OUTs reported via phylogenetic screening using PhyloChip.**

## Introduction

Alpha-diversity is defined as the number of distinct species found within an ecosystem and is usually assessed by classifying heterologous 16S rRNA genes amplified via polymerase chain reaction (PCR) using primers of broad specificity. In nature, relatively small cohorts of bacterial species are the major contributors of cell numbers to their specific habitat. In phylogenetic studies, the organisms that account for the most abundant operational taxonomic units (OUTs) are extremely copious and thereby, cloak the theoretical, and hitherto cryptic, “rare microbial biosphere” which is putatively comprised of an extremely low, but insurmountably diverse, number of planktonic cells (12). 16S rRNA-based phylogenetic analysis methods have been employed to interrogate the species richness of numerous environments (2, 3, 7). Recently, high-throughput microarray technologies, such as the Berkeley Laboratory PhyloChip, have gained popularity for their ability to detect multiple bacterial and archaeal species from complex environmental samples sans time and resource consuming clone library construction and sequencing (11). The PhyloChip high-density microarray is composed of 297,851 oligonucleotide probes complementary to 842 prokaryotic sub-families used for massive parallel screening. Even this advanced technology is subject to the deficiencies of PCR-based amplification, leading to underrepresented  $\alpha$ -diversity, of complex heterologous templates.

The application of PCR-based techniques to genetically and chemically complex environmental samples has proven to be extremely cumbersome (2, 6, 7). The reaction

kinetics of PCR favor, with each subsequent reaction cycle, the amplification of predominant phylotypes in a mixture of heterologous templates. Competitive conditions further distort results by favoring the amplification of particular set 16S rDNAs based solely on undefined favorable reaction conditions (3). The presence of *Taq* polymerase inhibitors, highly diverse 16S rRNA gene sequence composition, formation of heteroduplexes and heterogeneity of the *rrn* operon are all factors that influence the PCR technique and lead to an inherit bias in results (1-6). Amplicons generated from heterologous templates with identical primer sites exhibit intrinsic competitive conditions. Amplification of the minor templates is hindered by the ratio inequality of the major templates (3). Inactivation and/or removal of the major templates is therefore, a necessary step to resolve the cornucopian diversity found in the extremely low abundance templates.

The absence of methods capable of selectively removing or inactivating predominant members of a complex mixture of phylotypes led us to examine the feasibility of using Propidium Monoazide (PMA), a dye capable of intercalating and covalently cross-linking DNA upon exposure to light (9), to inhibit PCR amplification of major taxonomic contributors in a Model DNA Community (MDC). The random binding pattern of PMA on a defined MDC sample was computationally modeled and results showed that predominant phylotypes were disproportionately inactivated. These theoretical observations led us to hypothesize that by using a sub-saturating PMA concentration (i.e. PMA concentration capable of inactivating a high proportion of total DNA but not all) on a defined MDC sample, predominant phylotypes would be inactivated to a higher extent, resulting in an increase of relative percent contributions of low abundance templates in the post-PMA treated sample. Subsequent PCR amplification, using universal 16S rDNA primers, should provide a more ingenuous estimate of total template diversity due to the amplification of extremely low quantity heterologous templates, readily detected as an increase in reported OTUs.

The photo-catalyzed covalent binding of PMA to DNA has been successfully employed in previous studies concerning the prevention of quantitative (q)-PCR detection of target sequences from non-viable cells (8, 9, 10). In this study, we initially assessed the efficiency of PMA mediated inactivation of Spacecraft Assembly Facility Isolate *Bacillus pumillus* SAFR-032 genomic DNA using species-specific qPCR. Subsequently, experiments using photo-catalyzed PMA to inactivate large fractions of total MDC templates, followed by 16S rDNA and species-specific qPCR, were performed in aims of corroborating the hypothesized disproportional, probabilistically favored, inactivation of major contributing templates of our MDC. Finally, an environmental soil DNA sample was treated with sub-saturating PMA amounts, PCR amplified using universal small ribosomal subunit primers and analyzed for biodiversity via PhyloChip, a high-throughput DNA Phylogenetic Microarray.

## Materials and Methods

**PMA treatment of MDC.** 1 $\mu$ l of 0.025 $\mu$ M PMA working stock was used to treat 4 $\mu$ l of MDC sample ( $V_f = 5\mu$ l). The sample, containing a final PMA concentration of 0.005 $\mu$ M was exposed on ice to a 500W halogen lamp at a distance of 25cm for 4 minutes. Subsequently, total 16S rDNA and species-specific qPCR analysis was performed in

triplicate. Note: All procedures involving the use of PMA prior to halogen light exposure were performed in the dark to avoid premature photo-activation.

**PMA treatment of *B. pumilus* SAFR-032 genomic DNA.** Working stock solutions of 5 $\mu$ M, 0.5 $\mu$ M, 0.05 $\mu$ M and 0.005 $\mu$ M PMA were made using an original 200mM stock. 4 $\mu$ l portions of genomic DNA sample were treated with 1 $\mu$ l of the appropriate working PMA stock concentration ( $V_f = 5\mu\text{l}/\text{sample}$ ) to yield final sample concentrations of 1, 0.1, 0.01, 0.001 $\mu$ M PMA. Subsequently samples were exposed on ice to a 500W halogen lamp at a distance of 25cm for 4 minutes. The exact same protocol, sans PMA, was employed to treat the non-PMA treated control dilutions. Each sample was qPCR analyzed in triplicate for total 16s rDNA copy numbers. Note: All procedures involving the use of PMA prior to halogen light exposure were performed in the dark to avoid premature photo-activation.

**Soil DNA extraction and PMA treatment.** 10.0 g of garden soil underwent total DNA extraction via Soil Max<sup>TM</sup> Kit from Mo Bio Laboratories, Inc. Quantitative-PCR 16S rDNA analysis, using 1492R and 1369F primers, revealed a total 16s rRNA gene count of approximately 1.55 X10<sup>7</sup> copies per  $\mu$ l. Three different concentrations of PMA (0.10, 2.50 and 3.75 $\mu$ M) were used to treat environmental DNA samples. Stock solutions of PMA were used to make five 20  $\mu$ L aliquots per treatment (Fig.1) all of which were then exposed on ice to a 500W halogen lamp at a distance of 25cm for 4 minutes. The exact same protocol was employed to treat the non-PMA treated control dilutions. Note: All procedures involving the use of PMA prior to halogen light exposure were performed in the dark to avoid premature photo-activation.

**DNA pooling for PhyloChip analysis.** Following light exposure, all five 20 $\mu$ l aliquots made for each PMA concentration (0.75, 0.5, 0.1  $\mu$ M 1:10 dilution, 1:2 dilution and untreated control) were pooled to make the final 100 $\mu$ l samples sent for GeneChip<sup>®</sup> Phylogenetic Microarray analysis at the Lawrence Berkeley Laboratory.

**Model DNA Community (MDC) control.** Previously isolated DNA from JPL's Biotechnology and Planetary Protection genetic inventory was used to design our model community DNA control. The model is comprised of 16S rDNA from 11 trans-domain microbial constituents (Table 1, Figure 2). The relative percentage contribution of each microbial DNA constituent is highly disproportional.

**Quantitative Real-Time PCR reaction parameters.** All reported gene numbers were acquired via quantitative real-time polymerase chain reaction on a BioRad CFX96 thermocycler. IQ<sup>TM</sup>SYBR<sup>®</sup> Green 2X Supermix containing dNTPs, 50U/ml *iTaq*<sup>TM</sup> DNA polymerase, 6mM MgCl<sub>2</sub>, SYBR<sup>®</sup> Green I, 20nM fluorescein real-time PCR reagent was used. Final RXN mix: 10 $\mu$ l of 2X reaction mix, 1 $\mu$ l each of 10-18nM F and R primers, 1 $\mu$ l of sample template DNA and dH<sub>2</sub>O up to 20 $\mu$ l. qPCR protocol: 1) 95.0°C for 3:00, 2) 95.0°C for 0:10, 3) 56.5°C for 0:30 Plate Read, 4) GO TO 2, 39 more times, 5) 95.0°C for 0:10, 6) Melt Curve 65°C to 95°C: Increment 0.5°C for 0:05.

## Results

**Total MDC inactivation.** Once the efficiency of genomic DNA inactivation using a photocatalyzed PMA treatment was confirmed, an identical standard curve protocol was performed for the MDC (*data not shown*). This experiment should be considered a crucial proof of concept since, to our knowledge, this is the first attempt at quantifying the effects of PMA inactivation on a defined mixture of phylotypes at the species-specific level. A 0.005 $\mu$ M PMA treatment, determined to be necessary for inactivation of 85-90% of total MDC, was performed. This was followed by three sets of qPCR analysis using universal 16S rDNA, *Methanobacterium formicicum*-specific and *Micrococcus luteus*-specific primer sets (Table 2). Total DNA inactivation in the MDC was quantified using universal 16S rDNA primers. The 0.005 $\mu$ M PMA treatment was theoretically established to reduce 85-90% of DNA comprising the MDC. Data shows an actual reduction of 70-80% of total DNA (Figure 3). Species-specific qPCR analysis for *M. formicicum*, the most predominant species in our MDC (Table 2), and *M. luteus*, a minor template contributor to the MDC, shows that 75-80% of the initial 16S rDNA copies for the predominant species were inactivated (Fig. 4A) while only ~58% of the minor template was affected (Fig. 4B).

**SAFR-032 genomic DNA inactivation.** In order to test the feasibility of using PMA to inactivate genomic DNA from an environmental sample, we first had to assess the efficiency of photo-catalyzed DNA inactivation of a defined genomic sample. Various concentrations of PMA were used to treat SAFR-032 genomic DNA in aims of determining, via species-specific qPCR, a sample-specific standard curve (Fig. 5). A 1 $\mu$ M PMA treatment was enough to inactivated >99% of the initial genomic SAFR-032 DNA while a ten fold dilution in PMA concentration inactivated only ~63% of initial genomic DNA thus establishing the feasibility of using predetermined PMA concentrations for inactivating a desired portion of total DNA in sample.

**Inactivation of environmental DNA sample.** The environmental sample used for this study was garden soil collected from the Sierra Madre Foothills, located adjacent to the Jet Propulsion Laboratory in La Canada Flintridge, Ca. The total 16S rRNA gene copies from our environmental DNA extract were quantified using universal 16S rDNA primers. Results show the soil chromosomal DNA extract contained a total of  $\sim 1.20 \times 10^7$  16S rRNA gene copies/  $\mu$ l. This results falls close to the widely accepted general range of  $10^8$ - $10^9$  bacterial cells present per gram of soil (13). 99%, 86% and 52% of total DNA inactivation for this sample was achieved by using 0.75 $\mu$ M, 0.5 $\mu$ M and 0.1 $\mu$ M PMA treatments, respectively (Fig. 6). PMA treated environmental DNA samples were sent out to the Lawrence Berkeley Laboratory for standard 16S rRNA gene PCR amplification and subsequent massive parallel phylogenetic analysis via PhyloChip DNA microarray.

**PhyloChip DNA Microarray.** Biodiversity was assessed via DNA microarray hybridization of 16s rRNA gene amplicons (Fig. 7). The control sample showed a total count of 5,011 distinct OTUs. A 1:2 dilution control showed a significant drop in OTUs that is indicative of the ineffectiveness of dilution alone to enhance resolution of taxonomic diversity. PMA-mediated inactivation of 52% of total environmental DNA resulted in a 7% drop in diversity, while 86% inactivation of total environmental DNA

caused a 2% increase in observed OTUs. Remarkably, the inactivation of 99% of total environmental DNA, achieved via a 0.75 $\mu$ M PMA treatment, showed a significant increase in total OTUs detected by the phylogenetic microarray, increasing the resolution of microbial diversity by 19.5%.

## Discussion

The qPCR-based analysis of the Model DNA Community (MDC) before and after treatment with PMA at the individual species level was crucial for evaluating the strength of our computational random binding simulation. The theoretical foundations for our experiments relied on the probabilistic odds of individual random binding events of PMA to DNA. Our computer model suggested that, due to the extreme disproportionality in phylotype distribution, the random binding of PMA to DNA molecules (and subsequent photo-catalyzed covalent binding resulting in PCR inactivation) comprising a complex phylotype cocktail, would favor the inactivation of the predominant phylotype templates. The theoretical model was corroborated by empirical data from our MDC experiment (Figures 3, 4A & 4B). The random inactivation of 70-80% of total MDC DNA, by PMA treatment and subsequent photo-catalyzed covalent bonding, resulted in the disproportional inactivation of heterologous templates where the predominant template (*M. formicicum*) was inactivated to a significantly higher extent than a minor contributing template (*M. luteus*). These results however, were only indicative of the efficiency of our method *in vitro* using a model exclusively comprised of 16r DNAs genes.

Genomic DNA inactivation of SAFR-032 was readily achieved using various concentrations of PMA (Figure 5). This showed that i) our PMA stock solution possessed intact chemical properties, ii) our experimental set up was appropriate for subsequent PMA experimentation and iii) that reasonably small quantities of PMA would suffice to inactivate relatively large amounts of genomic DNA.

The environmental chromosomal DNA sample results perfectly corroborate our hypothesis. We have shown that by treating a complex environmental genetic mixture with a PMA concentration sufficient to inactivate 99% of the total initial DNA (Figure 6), deeper diversity, in terms of OTUs when compared to a control, is resolved (Figure 7). It is of crucial importance to note that a 99% inactivated initial DNA sample has only 1/100<sup>th</sup> of the initial DNA concentration of the control sample. The fact that we have attained a 19.5% diversity increase, indicative of 978 additional OTUs, with an initial DNA quantity available for PCR amplification that is two logarithmic units less than the control attest the extreme potential of this technique to finally explore the hitherto cryptic “deep microbial biosphere”. Even though the DNA template concentration for the 99% inactivated sample was significantly less than that of the control at the first step of PCR, subsequent amplification and phylogenetic microarray analysis has confirmed that it was significantly more diverse. These observations are strongly indicative of the PMA-mediated inactivation of predominant environmental phylotypes resulting in a higher percent contribution to the total DNA of low abundance templates which ultimately results in the exponential amplification of, previously “shadowed”, minor contributing phylotypes.

A drawback from employing phylogenetic microarray technologies for biodiversity assessment is that the oligonucleotide probes on the chip represent only previously described microbial diversity. We have, nevertheless, shown that at least 20% of the true microbial diversity of a soil sample would have been overlooked if the standardized protocols were employed for interrogation. Future work employing PMA inactivation of predominant phylotypes in environmental samples must include the laborious task of constructing a clone library to assess the potential of the PMA technique for elucidating novel microbial diversity.

In conclusion, current estimates of microbial  $\alpha$ -diversity in environmental samples fall short of the true diversity of these complex communities. Metagenomic analysis of complex environmental genetic samples based on the PCR amplification of 16S rRNA gene sequences of heterologous templates is subject to the inherent PCR bias resulting from the disproportional distribution of phylotypes. We have shown, to our knowledge, the only method available for the PCR-based resolution of low abundance, previously unaccounted, phylotypes in the environment. The “rare microbial biosphere” is hypothesized to be comprised of an extremely low, but insurmountably diverse, number of planktonic cells harboring endless possibilities for discovering novel microbial diversity. Here, we have shown the efficiency of an easily performed and relatively inexpensive method, using PMA, to increase the  $\alpha$ -diversity resolution of existing PCR-based methodologies. The validation of this novel technique via massive parallel screening on a DNA microarray is a technological leap in the field of microbial ecology.

## Tables and Figures

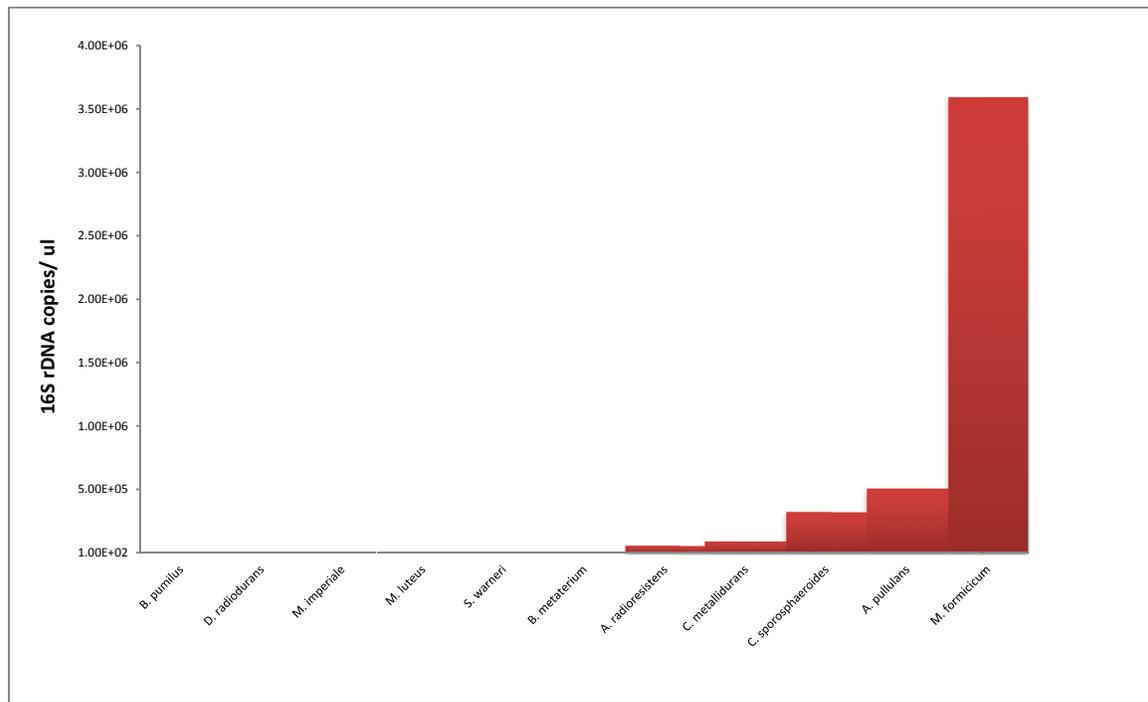
**Table 1.**

	<b>Microbes</b>	<b>Strains</b>	<b>Phylum</b>	<b>% contrib.</b>	<b>Total 16S RNA gene Copy Number/ <math>\mu</math>l</b>
1	<i>Methanobacterium formicicum</i>	DSM 1535	Euryarchaeota (Archaea)	78.04	3.59E+06
2	<i>Aureobasidium pullulans</i>	28v1	Ascomycota (Eukarya)	11.00	5.06E+05
3	<i>Clostridium sporosphaeroides</i>	DSM 1294	Firmicutes	7.00	3.22E+05
4	<i>Cupriavidus metallidurans</i>	CH34	( $\beta$ -)proteobacteria	2.00	9.20E+04
5	<i>Acinetobacter radioresistens</i>	50v1	( $\gamma$ -)proteobacteria	1.21	5.52E+04
6	<i>Bacillus megaterium</i>	KL-197	Firmicutes	0.50	2.30E+04
7	<i>Staphylococcus warneri</i>	82-4	Firmicutes	0.25	1.15E+04
8	<i>Micrococcus luteus</i>	ATCC 4698	Firmicutes	0.01	4.60E+02
9	<i>Microbacterium imperiale</i>	47v1	Actinobacteria	0.01	4.60E+02
10	<i>Deinococcus radiodurans</i>	ATCC 13939	Deinococcus-Thermus	0.01	4.60E+02
11	<i>Bacillus pumilus</i>	SAFR-032	Firmicutes	0.01	4.60E+02
			<b>Total copies/<math>\mu</math>l</b>	<b>100%</b>	<b>4.60E+06</b>

**Figure 1.**  
Preparation of PMA treated DNA

<b>Stock [PMA](<math>\mu</math>M)</b>	3.75	2.5	0.5
<b>Volume to Use (<math>\mu</math>L)</b>	4	4	4
<b>DNA (<math>\mu</math>L)</b>	16	16	16
<b>Final Volume (<math>\mu</math>L)</b>	20	20	20
<b>Final [PMA] (<math>\mu</math>M)</b>	0.75	0.5	0.1

**Figure 2.**



**Table 2.**

Primers	Target	Name	Sequence (5' to 3')
Universal	16S	27f	AGA GTT TGA TCM TGG CTC AG
		1392r	ACG GGC GGT GTG TRC
Species specific	<i>M. formicicum</i>	ssMFf	ATT GCT GGA GAT ACT ATT
		ssMFr	GGG ATT ATA GGA TTT CAC
Species specific	<i>M. luteus</i>	ssMLf	TAA CCT GCC CTT AAC TCT
		ssMLr	AAA CCG ATA AAT CTT TCC AA
Species specific	<i>B. pumilus</i>	600f	TGA AGC ACT TGA GAA ATT
		980r	TGC TGC AAA GAA AAT

Figure 3.

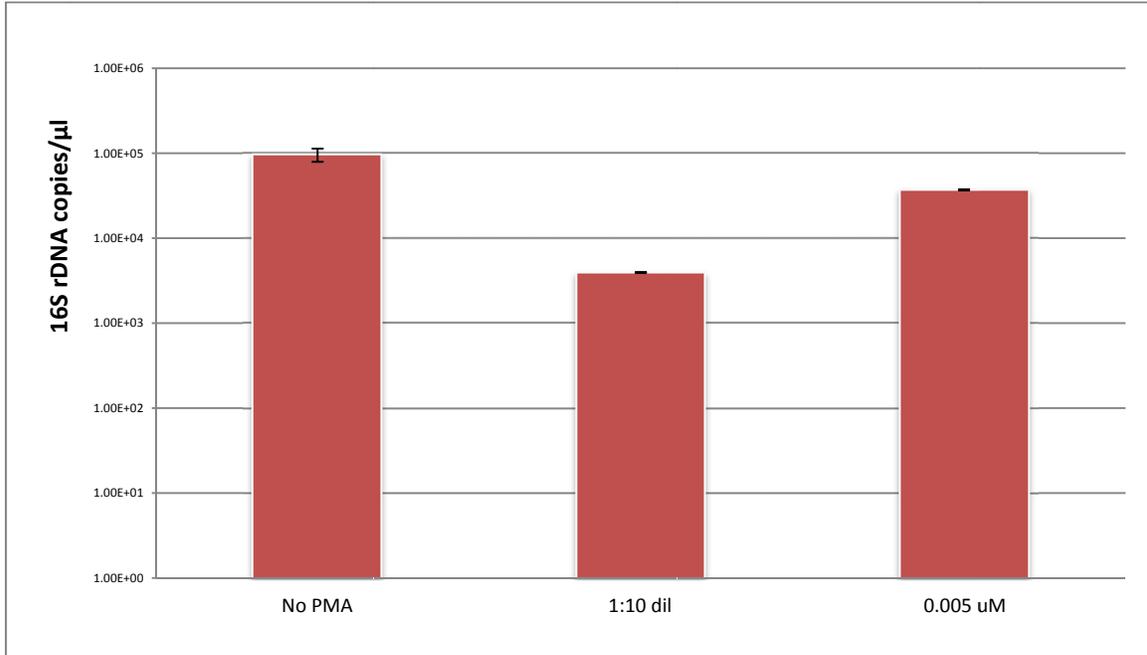
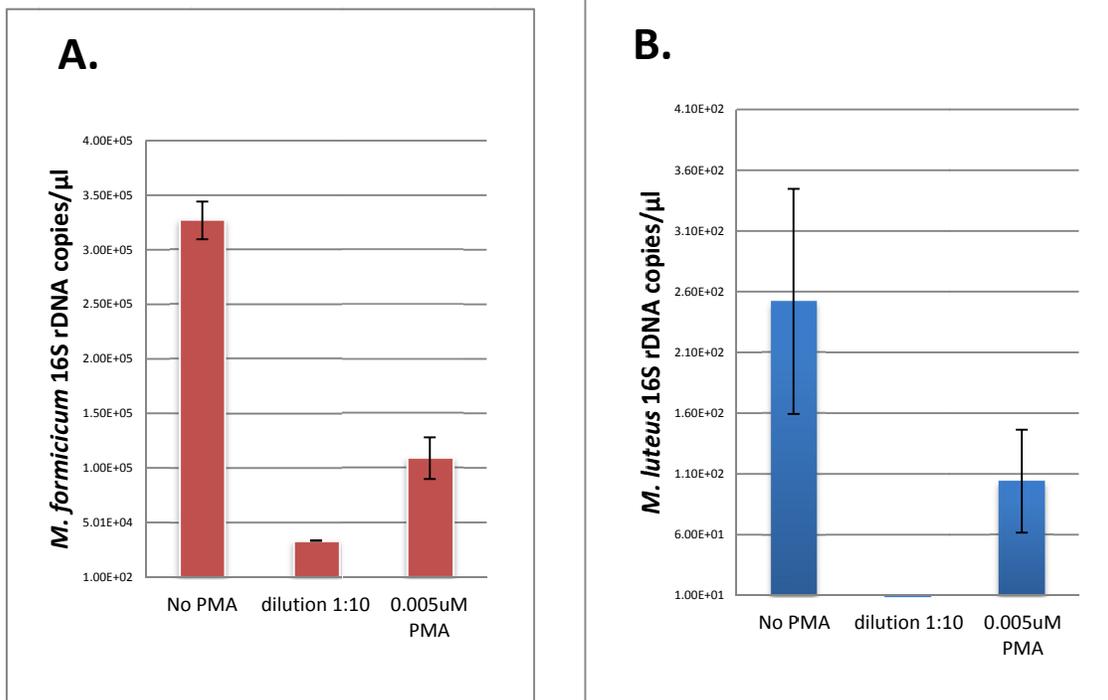
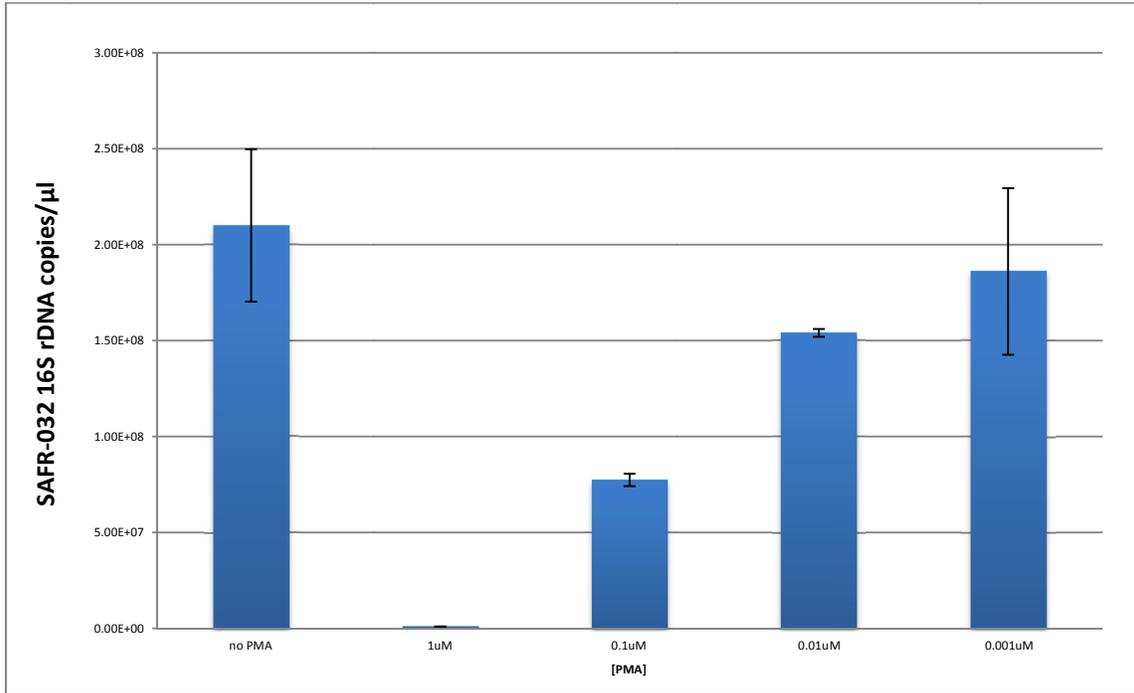


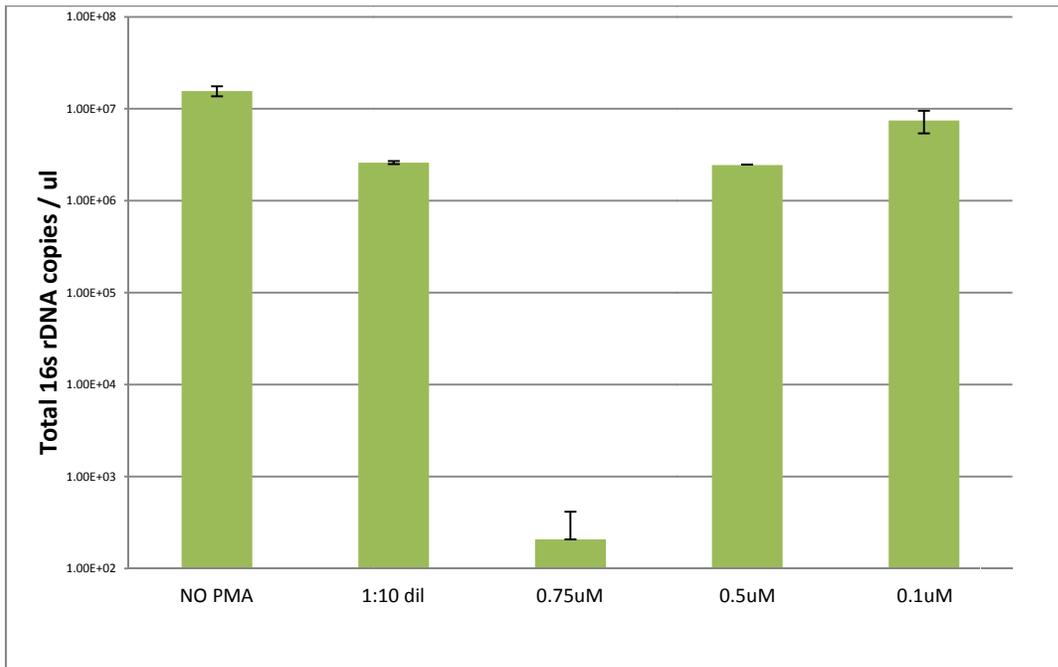
Figure 4.



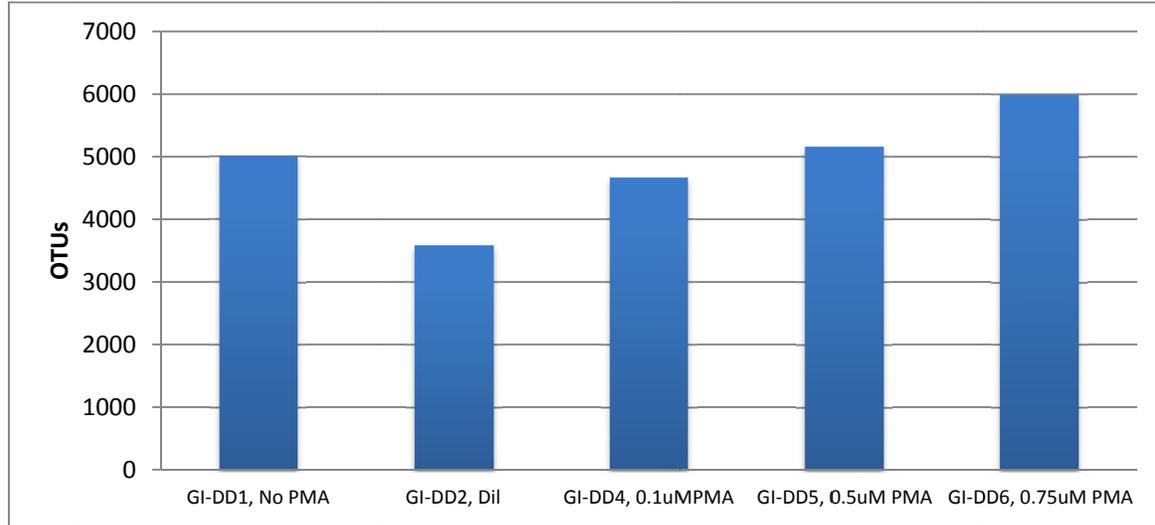
**Figure 5.**



**Figure 6.**



**Figure 7.**



## **Acknowledgments**

I am grateful to my mentor, Parag A. Vaishampayan, and my fellow intern, Arsh Chopra, for unflaggingly engaging in intellectual discussions with me which ultimately culminated in the production of this work. The authors would like to thank the intellectual and technical assistance of Dr. Kasthuri J. Venkateswaran, Myron T. La Duc, Emilee Bargoma and Dr. Moogega Cooper. This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by (NASA-USRP program) and the National Aeronautics and Space Administration.

## **References**

1. **Marcelino Suzuki and Stephen Giovannoni.** Bias Caused by Template Annealing in the Amplification of Mixtures of 16s rRNA Genes by PCR. *Appl. Environ. Microbiol.* **62**: 625-630.
2. **Winzingerode, F. V., U. B. Gobel, and E. Stackebrandt.** 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol. Rev.* **21**: 213-229.
3. **Sven Becker, Peter Böger, Ralfh Oehlmann, and Anneliese Ernst.** 2000. PCR Bias in Ecological Analysis: a Case Study for Quantitative Taq Nuclease Assays in Analyses of Microbial Communities. *Appl. Environ. Microbiol.* **66**: 4945-4953.
4. **Marcelino Suzuki, Michael S. Rappe, and Stephen Giovannoni.** Kinetic Bias in Estimates of Coastal Picoplankton Community Structure Obtained by Measurements of Small-Subunit rRNA Gene PCR Amplicon Length Heterogeneity. *Appl. Environ. Microbiol.* **64**: 4522-4529.

5. **Chandler, D. P., Fredrickson, J. K., and Brockman, F. J.** Effect of PCR template concentration on the composition and distribution of total 16s rDNA clone libraries. *Mol. Ecol.* **6**: 475-482.
6. **Xiaoyun Qiu, Liyou Wu, Heshu Huang, Patrick E. McDonel, Anthony V. Palumbo, James. Tiedje and Jizhong Zhou.** Evaluation of PCR-Generated Chimeras, Mutations and Heteroduplexes with 16s rRNA Gene-Based Cloning. *Appl. Environ. Microbiol.* **67**: 880-887.
7. **Jizhong Zhou et al.** 2002. Spatial and Resource Factors Influencing High Microbial Diversity in Soil. *Appl. Environ. Microbiol.* **60**: 326-334.
8. **Parshionikar et al.** Use of Propidium Monoazide in Reverse Transcriptase PCR To Distinguish between Infectious and noninfectious Enteric Viruses in Water Samples. *Appl. Environ. Microbiol.* **76**: 4318-4326.
9. **Nocker et al.** Use of Propidium Monoazide for Live/Dead Distinction in Microbial Ecology. *Appl. Environ. Microbiol.* **73**: 5111-5117.
10. **M. Varma et al.** Quantitative real-time PCR analysis of total and propidium monoazide-resistant fecal indicator bacteria in wastewater. *Water Research.* **43** (2009) 4790-4801
11. **Todd Z. DeSantis et al.** high Density universal 16s rRNA microarray Analysis reveals Broader Diversity than Typical clone Library When Sampling the Environment. *Microbial Ecology* **54** (371-383).
12. **Sogin et al.** Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences.* **103** (12115-12120).
13. **David M. Sylvia, Jeffrey J. Fuhrmann, Peter G. Hartel, David A. Zuberer.** Principles and Applications of Soil Microbiology. 2<sup>nd</sup> Edition. Ch.1. pg.9.