Evaluating PCR Bias Through Experimental Investigations of Complex Primer-Template Interactions

By

ANKUR NAQIB

B.Tech. Dr. D.Y. Patil University, 2008

Navi Mumbai, India

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Bioengineering
in the Graduate College of the
University of Illinois at Chicago, 2019
Chicago, Illinois

Defense Committee:

Yang Dai, Chair

Stefan J. Green, Advisor, Biological Sciences

Ao Ma

Rachel Poretksy, Biological Sciences

John Kelly, Loyola University

An ode to my grandparents, Naqib's and Misri's, for always instilling in me the value of education and a sense of morality.

To my parents, for always making me feel more like a friend than a son. Thank you for your selfless, unconditional and undying love.

To my wife Rubai, for standing by me every single time. Thank you for blindly trusting me and basically skydiving into this life with me without a parachute. I wish we never touch the ground!

ACKNOWLEDGEMENTS

I would like to thank my Advisor, Dr. Stefan J. Green, for seeing potential in me when I was a master's student at UIC. I will forever be indebted to you for providing me with all the space to learn from my own mistakes, for your constant guidance and support, and for always believing in doing things the right way. Your academic and life teachings will always stay with me. Thank you for being the inspiration I can always look up to. My journey in SQC has been one of the most joyous and defining moments of my life.

I would like to thank all the former and present SQC lab members. Thank you for all the help you provided during the fulfillment of this work.

My sincere and heartfelt gratitude to all my committee members: Dr. Dai, Dr. Ma, Dr. Poretksy and Dr. Kelly. Your valuable inputs shaped this work to where it is now.

A big thank you to Dr. Ali Keshavarzian and all his members at Rush University Medical Center. Our work collaboration has immensely contributed to my growth curve. I thoroughly enjoyed working with everyone one of you.

My sincere gratitude towards Dr. Irene H. Maumenee and Dr. Behrad Milani for letting me volunteer in your lab. You made my first lab experience at UIC amazing and worthwhile.

I would also like to thank all the members of Center for Research Informatics (CRI) for all the help and insights.

Last, but not the least, I would also like to thank RRC for providing all administrative help.

CONTRIBUTION OF AUTHORS

Chapter I of this dissertation highlights the extensive literature review that was conducted for my research. It also contains the issues that I am addressing through my research. Some parts of the introduction were previously published (Green, S. J., Venkatramanan, R., & Naqib, A. (2015). Deconstructing the polymerase chain reaction: understanding and correcting bias associated with primer degeneracies and primer-template mismatches. PloS one, 10(5), e0128122) and I am an author on the same. Thereafter, the dissertation contains four chapters that were designed to explain my research objectives. **Chapter II** of this dissertation is published work that I am an author on (Green, Stefan J., Raghavee Venkatramanan, and Ankur Naqib. "Deconstructing the polymerase chain reaction: understanding and correcting bias associated with primer degeneracies and primer-template mismatches." PloS one 10, no. 5 (2015): e0128122). The study was conducted under the guidance of Dr. Stefan Green (Sequencing Core, UIC) and with the support of Raghavee Venkatramanan (former UIC Department of Bioengineering graduate student). **Chapter III** of this dissertation is work that I am first author on and is under review for a publication (*Preprint* Naqib, Ankur, Silvana Poggi, and Stefan J. Green. "Deconstructing the Polymerase Chain Reaction II: An improved workflow and effects on artifact formation and primer degeneracy." PeerJ Preprints 7 (2019): e27525v1). The work was designed in collaboration with my advisor Dr. Stefan Green. The sequencing for this study was done in collaboration with Silvana Poggi (Northwestern University, formerly Sequencing Core at UIC). **Chapter IV** of this dissertation is published work and I am the first author on the study (Naqib A, Jeon T, Kunstman K, Wang W, Shen Y, Sweeney D, Hyde M, Green SJ. 2019. PCR effects of melting temperature adjustment of individual primers in degenerate primer pools. PeerJ 7:e6570 http://doi.org/10.7717/peerj.6570). This study was

designed in collaboration with my advisor Dr. Stefan Green and conducted in collaboration with Trisha Jeon, Kevin Kunstman, Weihua Wang, Yiding Shen, Dagmar Sweeney and Marieta Hyde (Sequencing Core UIC). **Chapter V** of this dissertation is my unpublished work that was designed in collaboration with Dr. Stefan Green and conducted in collaboration with Trisha Jeon. Dr. Stefan also provided me with experimental guidance and contributed to the writing. **Chapter VI** of this dissertation acts as a conclusion to the whole dissertation. It summarizes all my results, highlights implications of this research and discusses opportunities of future work in this area.

Table of Contents

Chapter I: Introduction.....01

Bias in the Polymerase Chain Reaction (PCR) 05

PCR bias caused by amplification cycle number 14

PCR bias caused by annealing temperature 18

PCR bias caused by genomic DNA concentration 19

PCR bias caused by degenerate primers 20

PCR bias caused by simultaneous linear and exponential amplification 22

Chapter II: Deconstructing the polymerase chain reaction: understanding and correcting bias associated with primer degeneracies and primer-template mismatches..... 25

Abstract 26

Introduction 27

Materials and Methods 29

DNA templates 29

Standard PCR (Targeted amplicon sequencing, TAS) 30

Specialty amplification reactions (PEX PCR method) 33

Next-generation amplicon sequencing 34

Analysis of mock community amplicon sequence data 35

Analysis of environmental community amplicon sequencing data 36

Analysis of primer utilization patterns 37

Results 38

Theory 38

Application of a new pipeline for PCR amplification of templates 42

Determination of optimum PEX PCR method annealing temperature 43

Analysis of a mock community using standard and PEX PCR method protocols 47

Analysis of environmental DNA using standard and PEX PCR method protocols 50

Analysis of primer utilization patterns 52

Discussion 54

Supporting Information 60

Chapter III: Deconstructing the Polymerase Chain Reaction II: An improved workflow and effects on artifact formation and primer degeneracy……..... 70

Abstract 71

Introduction 72

Materials and Methods 74

DNA templates 74

Primer Synthesis 74

Standard PCR Protocol 76

Deconstructed PCR (DePCR) Protocol 78

Sequence Data Analysis 79

Results 80

Theory 80

Validation of the DePCR method 83

Determination of linearity in DePCR amplification 90

Assessing the effect of individual primers in a degenerate primer pool 91

Discussion 97

Supporting Information 103

Chapter IV: PCR effects of melting temperature adjustment of individual primers in degenerate primer pools……...… 107

Abstract 108

Introduction 109

Materials and Methods 111

Primer Design and Synthesis 111

Genomic DNA Templates 112
Amplicon library preparation and sequencing 112
Bioinformatic Analysis of Sequence Data 114
Results 116
Primer design 116
Microbiome profiling using EMP and ShortEMP primers 119
Interrogation of a mock community with EMP, ShortEMP, Long EMP and NoLinker_ShortEMP primer sets 124
Assessing the need for phiX spike-in with ShortEMP primers 129
Discussion 130
Conclusions 132
Supporting Information 133
Chapter V: Quantitating primer-template interactions using a deconstructed PCR methodology……..... 146
Abstract 147
Introduction 148
Materials and Methods 151
Nucleic acids 151
Targeted-amplicon sequencing (TAS) protocol 152
Deconstructed PCR (DePCR) Protocol 153
Sequence Data Analysis 155
Results 157
Experimental design 157
Interrogation of single templates with primer pools of varying degeneracy 163
Interrogation of multi-template pools with a non-degenerate primer set 168
Interrogation of complex template pools with degenerate primer pools 171
Primer utilization profiles for each template within a complex template pool 175

Discussion 177

Conclusions 183

Supporting Information 185

Chapter VI: Conclusions……..... 223

Major advances 224

Cited Literature…… 231

VITA…….. 247

LIST OF TABLES

Table 1: Bias Table 07

Table 2: Primers used in this study 32

Table 3: Amplification and sequence analyses of mock community DNA 49

Table 4: Primers used in this study 75

Table 5: Rates of detectable chimeras in sequence data 84

Table 6: Alpha diversity indices of observed microbial communities 85

Table 7: Effects of amplification method and reverse primer variants on observed microbial community alpha diversity 93

Table 8: Alpha diversity indices of observed microbial communities 120

Table 9: Ideal score calculations for analysis of the 'Zymo' Standard 125

Table 10: Description of templates and primers used in experiments conducted as part of this study 159

LIST OF FIGURES

Figure 1: Types of DNA fragments found in PCR 23

Figure 2: Schematic of Targeted amplicon sequencing (TAS) and Polymerase Exonuclease (PEX) PCR methods 31

Figure 3: Types and abundance of DNA fragments found in PCR 39

Figure 4: Temperature gradient analysis of the PEX PCR and TAS methods using mock community DNA 44

Figure 5: Effect of PEX PCR Stage "A" annealing temperature on observed microbial community structure and primer utilization patterns 46

Figure 6: Relative abundance of mock DNA templates observed in sequencing of TAS and PEX PCR method reactions 48

Figure 7: Effect of PEX PCR and exonuclease treatment on observed microbial community structure and primer utilization patterns 51

Figure 8: Schematic of standard (TAS), polymerase-exonuclease (PEX) PCR, and Deconstructed PCR (DePCR) workflows 77

Figure 9: Polymerase-generated intermediates in the first stage of DePCR workflow 82

Figure 10: Effect of PCR methodology and annealing temperature on observed microbial communities 86

Figure 11: Effect of annealing temperature and amplification methodology on primer utilization profiles (PUPs) 88

Figure 12: Microbial community structure revealed using individual primer variants with TAS and DePCR amplification methodologies 95

Figure 13: Schematic of primer design and primer theoretical melting temperature distribution 117

Figure 14: Effect of annealing temperature and primer set on observed microbial community in Lake Michigan sediment 121

Figure 15: Effect of primer set on observed microbial communities in complex microbial samples 123

Figure 16: Comparison of representation of mock community standard using all primer sets and sequencing platforms 126

Figure 17: Effect of phiX spike-in on observed microbial communities with amplicons generated using ShortEMP primers 129

Figure 18: Schematic of Deconstructed PCR (DePCR) workflow 154

Figure 19: Primer, template and experimental design 161

Figure 20: Effect of PCR methodology and annealing temperature on PUPs of a single template 165

Figure 21: Effect of PCR methodology and annealing temperature on template profiles in amplification reactions utilizing a single primer 169

Figure 22: Effect of PCR methodology, annealing temperature, and primer pool on PUPs of ten templates 172

Figure 23: Effect of PCR methodology and annealing temperature on template profiles in amplification reactions utilizing varying primer pools 174

Figure 24: Template-specific primer utilization profiling 176

LIST OF ABBREVIATIONS

DNA	Deoxyribonucleic Acid
ANOSIM	Analysis of Similarity
BCD	Bray-Curtis Dissimilarity
BIOM	Biological Observational Matrix
DePCR	Deconstructed PCR
DGGE	Denaturing Gradient Gel Electrophoresis
EMP	Earth Microbiome Project
gDNA	genomic DNA
IDT	Integrated DNA Technologies
ITS	Internal Transcribed Spacer
KW	Kruskal-Wallis Tests
LMC	Lake Michigan Sediment
MID	Multiplex Identifier
mMDS	Metric Multi-dimensional Scaling
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
NMDS	Non-metric Multidimensional Scaling
OTU	Operational Taxonomic Units
PCR	Polymerase Chain Reaction
PEAR	Peared-End Read Merger
PEX PCR	Polymerase-exonuclease PCR
PGM	Personal Genome Machine
PUP	Primer Utilization Profile
QIIME	Quantitative Insights Into Microbial Ecology
RNA	Ribonucleic Acid
rpoC	RNA Polymerase Genes

RPV	Reverse Primer Variant
rRNA	Ribosomal RNA
SIMPROF	Similarity Profile Routine
SNP	Single Nucleotide Polymorphism
SPRI	Solid Phase Reversible Immobilization
SRA	Sequence Read Archive
SSU	Small Subunit
TAS	Targeted Amplicon Sequencing
Tm	Melting Temperature
TRFLP	Terminal Restriction Fragment Length Polymorphism
UICSQC	UIC Sequencing Core

SUMMARY

The polymerase chain reaction (PCR) is a well-established tool for amplification of regions of DNA and is used in a broad range of biological studies. PCR bias, in which some templates within a mixture of templates are preferentially amplified, is a well-known phenomenon. Despite substantial effort invested into correcting such bias, PCR-based studies continue to generate data that distort underlying template ratios. A major source of PCR bias is from primer-template interactions, leading to PCR selection favoring certain templates. Motives of this study were to understand better the causes of selection bias in PCRs with complex templates and complex degenerate primer pools, and to develop novel strategies to decrease bias. An experimental system was developed to reduce PCR bias by separating linear copying of templates from exponential amplification of amplicons (Deconstructed PCR or 'DePCR'), and this system also provides a mechanism to quantify primer-template interactions (Primer utilization profiles or 'PUPs'). DePCR was used to interrogate mock DNA communities and complex environmental samples, and all reactions were compared to standard PCR workflows. Experiments with annealing temperature gradients demonstrated a strong negative correlation between annealing temperature and the evenness of primer utilization in complex pools of degenerate primers. Critically, shifting primer utilization patterns mirrored shifts in observed microbial community structure. In experiments with mock DNA templates, DePCR demonstrates that although perfect match primer-template interactions are abundant, the dominant type of primer-template interactions are mismatch interactions, and mismatch amplification starts immediately during the first cycle of PCR. Furthermore, in DePCR reactions involving multiple mismatches, no strong effect on template profiles was observed. DePCR allows improved representation of templates, greater tolerance for mismatches between primers and templates, and greater success in

amplifying complex templates with low complexity primer pools. In addition, PUPs are empirical quantitative data derived from primer interactions with genomic DNA templates, and are a novel form of biological information that can be acquired only with DePCR. The DePCR method is simple to perform, is limited to PCR mixes and cleanup steps, has applicability to amplicon-based microbiome studies, and may also serve in other PCR-based protocols where primers and templates have mismatches.

Chapter I: Introduction

(Parts of this chapter have been previously published as Green, S. J., Venkatramanan, R., & Naqib, A. (2015). Deconstructing the polymerase chain reaction: understanding and correcting bias associated with primer degeneracies and primer-template mismatches. PloS one, 10(5), e0128122.)

PCR is one of the most essential advances in the field of biology, as it is robust and inexpensive method for generating millions of copies of DNA from limited templates. A common approach for studying microbial communities is to extract genomic DNA (gDNA) from multiple samples, PCR amplify this gDNA using gene-specific primers (usually targeting 16S ribosomal RNA genes) containing sequencing adapters and a sample-specific barcode, followed by equimolar pooling and sequencing [1]. As microbiome studies have become a critical component of the study of many human diseases [2, 3], it has become increasingly critical to decrease bias associated with PCR amplification and thereby improve the analysis of microbial communities. PCR bias, in which some templates within a mixture of templates are preferentially amplified, is a well-known phenomenon. Bias deriving from PCR amplification can be enhanced in the presence of a complex mixture of templates and in the presence of a degenerate primer pool (*i.e.,* a mixture of highly similar primers used simultaneously). Our motive with this study is to understand this bias better so as to be able to decrease its effects. In particular, this study is focused on characterizing and better understanding primer-template interactions in PCR systems with complex template and complex primer pools. As part of this work, we have developed a novel methodology (Deconstructed PCR, DePCR) to directly measure primer-template interactions, and have used this methodology to study fundamental properties of the polymerase chain reaction in complex primer-template systems. One obvious application of this work is amplicon-based microbiome studies, in which complex communities of organisms are interrogated with degenerate primers, but the methodology can also help in other systems where primer annealing and extension can be limited by mismatches between primer and template (*e.g.*, unknown single nucleotide polymorphisms, SNPS or in studies of methylation using PCR amplification of bisulfite treated DNA).

Thus, improving the way PCR experiments are run could lead to major improvements in the results of amplicon sequencing. Although the field of PCR-independent microbial community profiling (*e.g.,* shotgun metagenome sequencing) is rapidly developing, there are a number of features of 16S rRNA gene amplicon sequencing that still make it a valuable technique, for which reduced bias will be welcome. These include: substantially lower cost, high specificity to microorganisms even in the presence of substantial host nucleic acid, robust analytical techniques, and high annotation accuracy at genus or higher taxonomic levels. Besides PCR, 16s rRNA gene amplicon sequencing has been a long-standing strategy that allows bacterial identification that is more robust, reproducible, and accurate than that obtained by phenotypic testing [4]. The 16S rRNA amplicon sequencing approach has been the most commonly employed method to analyze microbiomes for the advantages that it is inexpensive and it has a vast archive of curated sequence database that can be used. However, known PCR biases are not rigorously addressed in most studies.

At least four major caveats are associated with amplicon-based 16S rRNA gene amplicon sequencing approaches: (i) Microorganisms contain a variable number of rRNA operons (e.g., [5], [6] and analyses of rRNA genes present a distorted representation of relative cellular abundance, (ii) PCR primer pools are often degenerate or the primers are anticipated to anneal to template sequences containing a variable number of mismatches with the primers, thereby producing bias in amplification efficiency among different templates, (iii) Samples are generally heavily amplified (30 cycles or more) leading to the possibility of extensive chimera formation, and (iv) Primer-template mismatches can lead to substantial underestimation of relevant taxa (**Chapter V**).

The objectives of this study were to understand the causes of bias in PCR amplifications in amplification reactions where both complex templates (*i.e.,* potential and variable mismatches with

primers) and complex (*i.e.,* degenerate) primer pools are present. In addition, this research sought to explore how different factors during PCR can be manipulated to reduce bias, but also other artifacts, such as chimeras. As a part of this research, I identified a novel form of bias (linear and exponential amplification bias, described in **Chapter II**), and developed a novel method to address it. This new method, initially called polymerase-exonuclease PCR (PEX PCR) and later updated with an improved worked more generally named 'Deconstructed PCR' (DePCR), addressed the newly identified form of bias, but also provided additional benefits (**Chapters III and V**). The deconstructed PCR approach (both PEX PCR and DePCR) provides the underpinnings for exploration of primer-template interactions by providing a novel source of information which I have named the 'Primer Utilization Profile' (PUP). The PUPs are empirical quantitative data derived from primer interactions with genomic DNA templates (**Chapters II, III, and V**). PUPs can be produced for standard PCR approaches; however, the information related to primer interaction with the source genomic DNA templates is lost due to a combination of primer-gDNA template and primer-amplicon interactions that occur during exponential amplification. Finally, a major determination of this study was that the DePCR methodology also significantly and substantially reduced the creation of artifact sequences known as chimeras [7]; **Chapter III**). The role of primer theoretical melting temperature (Tm) was also examined to determine if variability of Tm among primers in a degenerate primer pool contribute to PCR bias (**Chapter IV**). The novel methodology was also to explore the need for degenerate primer pools for amplification of complex DNA templates, and these studies included examining the activity of each primer from a primer pool independently (**Chapter III**), and in a systematic interrogation of a mock community with varying template and primer pools (**Chapter V**).

Bias in the Polymerase Chain Reaction (PCR)

The Polymerase Chain Reaction (PCR) has been an extraordinarily useful molecular biological technique, and has been used to create large numbers of DNA copies of specific regions for downstream analyses [8]. PCR is an *in vitro* technique for amplification of a region of DNA whose sequence is known or which lies between two regions of known sequence. Before PCR, DNA of interest could only be amplified by over-expression in cells and this limited yield. PCR soon became a standard method in the field of biology because of its selectivity, sensitivity and speed, as millions of DNA copies can be obtained from complex genomes in a matter of hours. Also favoring PCR is the low initial quantities of DNA required.

The field of microbiology was also a quick adopter of PCR for cultivation-independent analyses of complex microbial communities, and PCR amplicons generated from gDNA extracted from environmental samples were used for profiling initially using cloning and sequencing [9], then with gel separation techniques such as denaturing gradient gel electrophoresis [10], and finally with next-generation sequencing platforms [11]. The use of PCR coupled with sequencing was hugely beneficial for the field of microbiology, due to the difficulty in cultivating many environmental organisms, the obvious high diversity as observed microscopically, and the paucity of diagnostic morphological features of microbial cells [12].

Although PCR is an indispensable tool, limitations of the methodology were observed very early on. Detection of PCR artifacts, such as chimeras, were detected early [7]. In addition, clear distortions of underlying template ratios were also observed [13], [14], [15]. Such 'PCR bias' has been well-studied, particularly in the context of microbial ecology [14], [15], [16], and as such, PCR conditions have been shown to favor certain templates. Bias generated from selection bias has been attributed to a broad number of factors, including (but not limited to): annealing

temperature [17], [18], mismatches between template and primer [19], [20], location of mismatches between template and primer [21], interference from flanking regions due to during initial stages of PCR [22], too many PCR cycles [23], input DNA concentration [24], [25], [26], preferential amplification of low GC templates in a mixture [13], higher GC content in primer region/differences in primer binding energy [15], [27], template saturation at plateau phase of PCR [28], preferential formation of primer dimers from some primer variants when working with degenerate pools of primers [15], preferential amplification of unmethylated DNA [29], re-annealing of copies to templates and thereby inhibiting additional copying – particularly of dominant templates – thereby reducing amplification efficiency [30, 31], ramp rate for change in temperature during thermocycling allowing for formation of homoduplexes [32], and combinatorial effects of linear copying of gDNA and exponential amplification of PCR products occurring simultaneously and at different efficiencies [18]. Regardless of the exact cause of PCR bias, even small distortions in the evenness of amplification of templates due to PCR bias can become large over 30+ cycles of PCR can be substantial. Using a formula developed by Suzuki and Giovannoni [28], we can calculate that in a simple scenario with two templates with a starting template ratio of 1:1, if the efficiency of amplification of one template is 10% greater than the other, with 30 cycles of PCR the preferentially amplified sequence will be approximately 4.5X higher than the template with lower efficiency. Thus, small differences in efficiency can lead to large distortion effects. Below, I review in greater detail the prior literature on PCR bias caused by various factors, and also examine the methodology used to interrogate PCR bias. A summary of manuscript reviewed is shown in **Table 1**.

Table 1: Bias Table

S.No.	Manuscript	Year	Bias Discussed	Experimental Design	Implementation Effect	Reference	Similar Studies
1	Brakenhoff, R. H., J. G. Schoenmakers, and N. H. Lubsen	1991	Artifacts	The paper encounters a novel artifact of the PCR. Chimeric sequences could have resulted from somatic recombination or trans-splicing but are more likely an experimental artifact. The explanation is that partial cDNAs that are formed during cycle 1 of PCR serve as primer sites for reverse transcription by Taq polymerase. Retesting was also done to confirm the hypothesis with RNAase included. All five recombinant clones contained the correct transcript and no chimeric clones were seen.	Specificity of primers is of utmost importance. In usual PCR reaction, these kind of artifacts may contribute a lot towards the total output.	Brakenhoff, R.H., Schoenmakers, J.G. and Lubsen, N.H., 1991. Chimeric cDNA clones: a novel PCR artifact. Nucleic acids research, 19(8), p.1949.	
2	Wagner, A., N. Blackstone, P. Cartwright, M. Dick, B. Misof, P. Snow, G. P. Wagner, J. Bartels, M. Murtha, and J. Pendleton	1994	Primer Degeneracy	Analysis of the sometimes complex characteristics of multigene families has been greatly facilitated by use of the polymerase chain reaction (PCR). Using degenerate primers corresponding to highly conserved regions of homologous genes, PCR can be used to detect and identify members of these gene families in samples of genomic DNA or cDNA. A. In the context of sampling gene families, the paper studies those factors that produce skewness in the distribution of inserts, i.e., an apparent excess of inserts of some members of a gene family relative to others. It suggests two major classes of processes leading to such excess, PCR selection and PCR drift. PCR selection occurs when the reaction favors certain members of a gene. PCR drift is the result of random events occurring in the early cycles of the reaction. To study this, a single probabilistic model was created.	The reduce PCR drift type of bias, carrying out several independent reactions and pooling the products should reduce the skewness of the distribution. In cases of PCR selection type of bias (for e.g. wide variation in GC/AT ratio) set of degenerate primers used or of the target region of the gene family, or both, one may carry out the reaction only for the smallest necessary number of cycles or, alternatively, start the reaction with a small amount of DNA (small number of genomes) so as to override the effects of selection by the strong stochastic forces occurring in the first few cycles of the reaction.	Wagner, A., Blackstone, N., Cartwright, P., Dick, M., Misof, B., Snow, P., Wagner, G.P., Bartels, J., Murtha, M. and Pendleton, J., 1994. Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. Systematic Biology, 43(2), pp.250-261.	v. Wintzingerode, F., Göbel, U.B. and Stackebrandt, E., 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS microbiology reviews, 21(3), pp.213-229.
3	Suzuki and Giovannoni	1996	Primer Degeneracy PCR Cycle Number	Two different template mixtures were used. The templates were amplified against the 27F-1492R and 519F-1406R primer sets. Quantitative experiments were done using either the 27F-338R or the 519F-1406R primer set. The number of PCR cycles were also altered at 10, 15, 25, or 35 cycles.	Biases, that were strongly dependent on the choice of primer were observed. Results also indicate that the role of templates in introducing bias is on a lesser extent. For the 519F-1406R primer pair, little or no bias and a low product yield was observed. In contrast, for the 27F-338R primer pair, a strong bias and a much higher yield of product was observed. The authors also suggest that there might be two factors due to which this may have happened. Firstly, the 519F-1406R primer set is almost 3 times longer in comparison to the 27F-338R primer set, inhibiting Taq DNA polymerase's efficiency to amplify. Secondly, the 1406R reverse primer being a 15-mer should ideally anneal at 55°C with an efficiency lower than that of the 20-mer 27F or 18-mer 338R. Authors also suggest keeping the number of PCR cycles low to reduce bias.	Suzuki, M.T. & Giovannoni, S.J. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. 62, 625–630 (1996).	
4	Wang, G. C., and Y. Wang.	1996	Artifacts PCR Cycle Number	The paper discusses the robustness of PCR and then its drawbacks. The first drawback being that of preferential annealing. This results in an output where the frequency of a sequence occurring in a 16s library prepared from a sample does not reflect its actual relative abundance. The second drawback being that of chimeric artifacts. This occurs when the 16s rRNA genes of more than one species is PCR-amplified in a single reaction, chimeric or recombinant molecules may be generated which consist of mixtures of sequences from different 16s rRNA genes. The aim of this study is to study and evaluate the formation of PCR artifacts and also examine the effect of PCR cycles on the reaction. 16s rRNA genes of Streptosporangizlm nondiastaticzlm (IF0 13990), Streptosporangizlm pszeldowkare (IF01 399 1), Promicromonospora sditrmoe (IF0 14650) and Micromonospora chalcea were PCR amplified against the 5' TTA CCT GAT AGCGGCCGC AGA GTT TGATCC TGG CTC 3' and 5' TAC AGG ATC CGCGGCCGC TACGG(CT) TAC CTT GTT ACG ACT T 3' primer set.	Results of this study showed that Taq polymerase had little effect on formation of artifacts. The results of this study suggest it is advisable to use the least possible number of amplification cycles as the number of detectable artifacts were significantly lesser in samples that were ran for 10 cycles in comparison to the sample ran for 20 cycles. Also lesser elongation times of 2minutes resulted in more artifacts in comparison to the ones ran for 5 minutes. Also the paper advices to take greater caution in analysis of sequences PCR-amplified from complex genomes.	Wang, G.C. and Wang, Y., 1996. The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. Microbiology, 142(5), pp.1107-1114.	Wang, G.C. and Wang, Y., 1997. Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. Applied and environmental microbiology, 63(12), pp.4645-4650.

S.No.	Manuscript	Year	Bias Discussed	Experimental Design	Implementation Effect	Reference	Similar Studies
5	Mathieu-Daude´, F., J. Welsh, T. Vogt, and M. McClelland	1996	PCR Cycle Number	This study observes that the rate of amplification of abundant PCR products declines faster than that of the less abundant products. To tackle that, it hypothesizes that this bias in PCR products can be partly avoided by limiting the number of PCR cycles. This would also limit the abundance of the products and result in an abundance based normalizations. Samples were PCR against two primer sets 5′-AATGAAAGTTACGATAGCGG and 5′-AAAGACAACGGAGATGGCA for the ESAG transcript (GenBank accession no. U53929), and 5′-TGAAGCAGAAGACAATCAGG and 5′-AAAAATGCCAGTAGCAGGAC for the other transcript, called 'BET-2' (GenBank accession no. U49238). Varying number of cycles of 94°C for 30 s, 35°C for 30 s and 72°C for 50 s or five cycles of 94°C for 30 s, 40°C for 30 s and 72°C for 50 s, followed by various numbers of cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 50 s were performed.	The paper concludes that based on its results, re-hybridization appears to be responsible for the bias against PCR of abundant products in the late cycles of PCR. Also, the results seem more reliable when the PCR was done for lower number of cycles. However, when approaching saturation levels, the normalization phenomenon affects relative abundance.	Mathieu-Daudé, F., Welsh, J., Vogt, T. and McClelland, M., 1996. DNA rehybridization during PCR: the 'C o t effect'and its consequences. Nucleic acids research, 24(11), pp.2080-2086.	Suzuki, M., Rappé, M.S. and Giovannoni, S.J., 1998. Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity. Applied and environmental microbiology, 64(11), pp.4522-4529.
6	Polz and Cavanaugh et al.	1998	Primer Degeneracy	3 samples, out of which 2 were closely related and 1 distant, were used for this study. Genomic DNAs from these were amplified with degenerate primer set 27F and 1492R. Aim is to study two major classes of PCR processes, PCR selection (all mechanisms wherein certain templates are preferred due to either high copy numbers, higher GC-rich binding efficiency) and PCR drift (includes all the variations in the early PCR cycles when amplification is governed by genomic template).	Results indicate that PCR drift shows little contribution towards bias. Whereas, PCR selection emerges as the force behind unequal template amplification. To a large extent, PCR selection may also be caused by differences in the GC content at degenerate positions in the primer target sites in the 16S rDNAs. Results also show that effects of PCR selection can be reduced by low PCR cycles and higher template concentration.	Polz, M.F. & Cavanaugh, C.M. Bias in template-to-product ratios in multi-template PCR. Appl. Environ. Microbiol. 64, 3724–3730 (1998).	
7	Becker, S. et al.	2000	Primer Degeneracy	Pelagic Synechococcus sp. strains BO 8805, BO 8807, BO 8808, BO 8809, and BO 9404 and Synechocystis sp. strain BO 8402 isolated from the pelagic zone of Lake Constance were cultured. Microcystis sp. was also used as a sample. Anabaena variabilis strain ATCC 29413, and Anacystis nidulans (also known as Synechococcus leopoliensis strain SAG 1402-1 or Synechococcus sp. strain PCC 6301) were cultivated accordingly. DNA was extracted and RTPCR was ran using primers labeled and designed for Synechococcus spp..	The paper states that in case of analysis of microbial communities, one should have a large detection range and high specificity of primers. Also, loss of signal was observed and later attributed to a surplus of highly complex DNA which was later solved by increasing the primer concentration. Another issue was caused by amplicon accumulation that Taq was unable to detect.	Becker, S., Böger, P., Oehlmann, R. and Ernst, A., 2000. PCR bias in ecological analysis: a case study for quantitative Taq nuclease assays in analyses of microbial communities. Applied and Environmental Microbiology, 66(11), pp.4945-4953.	
8	Ishii and Fukui	2001	Annealing Temp.	Effect of the annealing temperature on the product ratio was investigated by denaturing gradient gel electrophoresis analysis of PCR products from a mixture of perfect-match and one-mismatch templates. These templates were generated by PCR from Pediococcus acidilactici for one mismatch and Micrococcus luteus for the perfect match. The primer set used was 341F and 907R and analyzed with the	The results of this study shows that the bias caused by the difference in primer binding energies was reduced by a lower annealing temperature. In contrast, Polz and Cavanaugh study showed that bias toward a primer with high binding energies was observed at a relatively low annealing temperature of 50°C. The explanation that the authors state for the disagreement between their results and the study conducted by Polz and Cavanaugh is the differences in the PCR conditions, such as the addition of 5% acetamide and the larger reaction volume in the latter's study. A reaction with acetamide might require a lower annealing temperature than one without acetamide.	Ishii, K. & Fukui, M. Optimization of annealing temperature to reduce bias caused by a primer mismatch in multi-template PCR. Appl. Environ. Microbiol. 67, 3753–3755 (2001).	
9	Qiu, X., L. Wu, H. Huang, P. E. McDonel, A. V. Palumbo, J. M. Tiedje, and J. Zhou.	2001	Artifacts PCR Cycle Number Input DNA conc Elongation Time	Model community of four species was constructed from alpha, beta and gamma subdivisions of Proteobacteria. This was done to evaluate PCR artifacts and bias caused because of that. Also, three different polymerases were used to asses each one's efficiency. 40 different gram-negative and positive bacteria were amplified against the eubacterium-specific primer set FD1 - R1540 and all treatments were carried out in triplicates. Different input template concentrations (0.1, 1.0 and 10 ng/ml) were used. Varying PCR cycle numbers (22, 25 and 28) were used. Elongation time was also varied (20s, 2mins and 4 mins).	PCR artifacts were found to be significantly different among the three Taq DNA polymerases: 20% for Z-Taq, with the highest processitivity; 15% for LA-Taq, with the highest fidelity and intermediate processitivity; and 7% for the conventionally used DNA polymerase, AmpliTaq. In contrast to the theoretical prediction, the frequency of chimeras for both Z-Taq (8.7%) and LA-Taq (6.2%) was higher than that for AmpliTaq (2.5%). The frequencies of chimeras and of heteroduplexes for Z-Taq were almost three times higher than those of AmpliTaq. Also, it was seen that chimeras increased with higher PCR cycles and higher starting template concentrations. PCR artifacts decreased with higher elongation time with the highest with observed at 20 seconds (25.5%). When elongation time was increased to 4 minutes, the artifacts percentage was reduced to 16%. Sample and species diversity also played a role in artifacts formation. Higher species diversity resulted in higher artifacts formation. The paper suggests that cycle number should be decided by the amount of template used, amplification efficiency, and existence and degree of inhibitory substances. All these factors should be experimentally determined for higher efficiency and throughput. PCR artifacts can also be minimized by using PCR products prior to or during the exponential period for cloning.	Qiu, X., Wu, L., Huang, H., McDonel, P.E., Palumbo, A.V., Tiedje, J.M. and Zhou, J., 2001. Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. Applied and environmental microbiology, 67(2), pp.880-887.	

S.No.	Manuscript	Year	Bias Discussed	Experimental Design	Implementation Effect	Reference	Similar Studies
10	Hongoh et al.	2003	Primer Degeneracy Annealing Temp.	Two 16s rDNA clone libraries were used. Annealing temperature for this experiment was kept at 55°C for 24 PCR amplification cycles. The primer sets 63F-1389R and T63F(5′-CAGGCCTAACACATGCAAGTT-3′)-1389R were used to amplify the region corresponding to 64–1388 in Escherichia coli (J01695). The final analysis was done using the Ribosomal Database Project (RDP). Separate libraries, K, M and N were prepared using primer 63F-1389R and amplified at an annealing temperature of 45°C.	The results indicate that 63F-1389R produced PCR products more efficiently than the primer 27F. Authors also concluded that a mismatch at the 3′-end between a PCR primer and its target seems more critical. These biased amplifications seem to be caused by primer mismatches of the forward primers, because no obvious difference was found between the libraries prepared using the reverse primers 1389R and 1492R. Furthermore, the biased amplification with 63F or 41F was improved when the degenerate primers 64F and 39F were used, which were designed to match spirochetal and Bacteroides sequences. Annealing temperature differences were observed when comparing libraries K and N, prepared using 63F-1389R at 45°C and 55°C, respectively. A significant difference in the expected diversity of detectable phylotypes was demonstrated by rarefaction analysis. This was in parallel to their previous model study that showed that a lower annealing temperature allowed annealing even at a mismatched site. These results suggest that a primer best matched with diverse bacterial groups, a lower annealing temperature and a decreased number of PCR cycles should be used to minimize amplification bias.	Hongoh, Y., Yuzawa, H., Ohkuma, M., & Kudo, T. (2003). Evaluation of primers and PCR conditions for the analysis of 16S rRNA genes from a natural environment. FEMS Microbiology Letters, 221(2), 299-304.	
11	Takahiro Kanagawa	2003	PCR Bias Artifacts PCR Cycle Number	This study focuses on the mechanism of PCR bias and artifact formation in multi-template PCR, and discusses suitable methods for the elimination of these problems to increase the reliability of the data from PCR-based analyses. The different types of PCR bias and occur because of:- Differences in Primer binding energy - The primer set used in multi-template PCR should have a sequence common to the targets. However, often there are no common sequences in the targets. If a primer has one mismatch with some targets, the amplification efficiency is usually very low, and therefore, a large bias in the amplification will occur. To avoid this bias, a degenerate primer which is a mixture of primers with a nucleotide sequence corresponding to the variation among homologs is often used. Chimeras: An incompletely extended primer can act as a primer in the subsequent PCR cycles. If the incomplete fragment anneals to a different template having a partially homologous sequence, one sees a chimera product. An experiment by Paabo et al. showed that by 40 cycles of PCR, chimeric products were generated which had the expected size and sequence. This shows that PCR products having a partially homologous sequence at the 3’ end can act as primers in the subsequent PCR cycles to produce chimeric molecules. Template: Mixture ratio: Suzuki and Giovannoni reported a strong bias towards 1: 1 mixtures of genes in final PCR products, regardless of the initial ratio of the templates. The bias was strongly dependent on the cycle number. The original differences in concentration decreased as the number of PCR cycles increased. This bias is explained by re-hybridization of the PCR products. When the amplification reaction proceeds, the concentration of PCR products becomes high enough to allow the re-hybridization of the products to some extent while the temperature is lower than the DNA melting point.	The study suggests the formation of chimeras occurs in the later cycles of PCR when the concentration of the incompletely extended primers is high enough to compete with the original primer for annealing. Therefore, chimera can be avoided by limiting the number of PCR cycles. Also, the study suggests lowering down the PCR cycle number to reduce the initial mixture: final product ration. The reasoning behind this as stated is that the amplification rate for abundant PCR products declines faster than that for less abundant products in the later PCR cycles, and the difference in starting template concentrations decreases.	Kanagawa, T. (2003). Bias and artifacts in multitemplate polymerase chain reactions (PCR). Journal of bioscience and bioengineering, 96(4), 317-323.	
12	Lueders, T. and Friedrich, M.W.	2003	Primer Degeneracy	Pure cultures of bacterium Methanobacterium bryantii DSM 863T, Methanosaeta concilii DSM 3671T, and Methanospirillum hungatei JF1 DSM 864T were obtained along with soil from a rice field. DNA was extracted and PCR was ran against Ar109f-Ar915r primer set. Another set of samples was amplified using FAM-labeled MCRf and MCRr primer pair.	The results show that varying annealing temperatures may alter primer-binding kinetics in a template mixture, especially if primers with degenerate positions are used. Final ratios of mcrA-targeted T-RFLP analysis were biased, most likely by PCR selection due to primer degeneracy. Biases involved in PCR should be evaluated for each primer set.	Lueders, T. and Friedrich, M.W., 2003. Evaluation of PCR amplification bias by terminal restriction fragment length polymorphism analysis of small-subunit rRNA and mcrA genes by using defined template mixtures of methanogenic pure cultures and soil DNA extracts. Applied and Environmental Microbiology, 69(1), pp.320-326.	
13	Kurata, S. et al.	2004	PCR Bias	Multi-template PCR strategies have regularly been used in ecology to assess environmental microbiomes. However, it has been shown that these analysis can show wrong abundance and diversity of genes due to bias and artifacts. This study looked at the 1:1 bias by employing a real time PCR. The paper also states that reannealing has not been shown to be a cause of bias. Whereas, the products formed by homoduplex formation during reannealing (decrease in temperature from denaturation to annealing) causes bias. Fluorescently labeled PCR products of bacterial 16S rRNA genes are used as templates. . The 16S rRNA genes were then amplified by PCR from the genomic DNA with the primer pair EC27f - E1389r primer set.	Results indicate that the ramp time from the denaturation to annealing step is one of the factors behind bias. The experiment also suggests that the best way to minimize bias towards a 1:1 ratio is to set the fastest ramp rate from denaturation to annealing. Care has to taken while doing this because heteroduplexes may form when PCR reaches plateau phase.	Kurata, S., Kanagawa, T., Magariyama, Y., Takatsu, K., Yamada, K., Yokomaku, T. and Kamagata, Y., 2004. Reevaluation and reduction of a PCR bias caused by reannealing of templates. Applied and environmental microbiology, 70(12), pp.7545-7549.	

S.No.	Manuscript	Year	Bias Discussed	Experimental Design	Implementation Effect	Reference	Similar Studies
14	Acinas, S.G. et al.	2005	PCR Cycle Number	2 large 16S rRNA gene libraries were generated from a single sample of bacterioplankton. First library was generated using a 35 cycle PCR and the second library through a 15 cycles of PCR amplification followed by 3 additional cycles. The 2-step process, as per the paper, allows to decrease PCR bias, *Taz* polymerase errors and reduce chimera formation.	Result table 1 shows that decrease in PCR cycle number (from 35 to 15+3) has a strong effect on chimera formation. PCR bias was previously linked to differences in amplification efficiencies as a consequence of differences in primer binding energy and to inhibition of amplification due to the reannealing of templates that occurs once they reach saturation concentration. The paper suggests to use several PCR replicates to reduce drift, lessen PCR amplification cycles to reduce chimeras and polymerase bias and high ramp rates between the denaturation and annealing steps and low annealing temperatures should be used, while long extension times (180 s) should be avoided.	Acinas, S.G., Sarma-Rupavtarm, R., Klepac-Ceraj, V. and Polz, M.F., 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Applied and environmental microbiology, 71(12), pp.8966-8969.	
15	Nikolausz et al.	2007	Primer Mismatch Annealing Temp. PCR Cycle Number	Aeromonas hydrophila (ATCC 7966); Bacillus cereus (ATCC 14579); Bacillus subtilis (ATCC 6633); and Pseudomonas fluorescens (ATCC 13525) bacterial strains were amplified against the 27F(or 63F)-1387R primer set. 27F forward primer has no mismatches while 63F has 3 mismatches. Reverse primer 1387R also has no mismatches. The annealing temperature ranged from 47 - 61°C. The cycle number also varied at 18, 24, 32 and 48 cycles.	A. hydrophila and P. fluorescens were preferentially amplified over both Bacillus strains when the 63F primer was used, while the 27F primer amplified all templates without bias. The extent of the preferential amplification showed an almost exponential relation with increasing annealing temperature from 47 to 61°C. PCR cycle number also had a little influence on the template-to-product ratios. The authors suggest using a low annealing temperature to significantly reduce preferential amplification.	Sipos, R., Székely, A. J., Palatinszky, M., Révész, S., Márialigeti, K., & Nikolausz, M. (2007). Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. FEMS Microbiology Ecology, 60(2), 341-350.	
16	Issa et al.	2007	Annealing Temp.	Aberrant DNA methylation plays an important role in both cancer initiation and progression, and this process is also implicated in other diseases, including imprinting disorders, diseases with trinucleotide expansions, and aging-related diseases. The central role of DNA methylation in maintaining cellular function, and the broad implications of DNA methylation in diseases have created a strong need for techniques to detect and measure DNA methylation reliably and quantitatively. The study states that all current methods rely on sodium-bisulfite treatment that helps create difference in sequences by converting un-methylated cytosines to uracils and leaving methylated cytosines unchanged. The main concern for PCR-based quantitative DNA methylation analysis is PCR bias, which is due to the fact that methylated and un-methylated DNA molecules sometimes amplify with greatly differing efficiencies.	The results of this study indicated that the reasons why increased annealing temperature can improve PCR efficiency for methylated DNA are unknown. One theory can be that after bisulfite treatment, methylated DNA has a higher GC content than un-methylated DNA, which could favor stable secondary structures that alter amplification efficiency. Raising the annealing temperature of PCR could melt these secondary structures and thus correct the lower amplification efficiency of methylated DNA. The authors suggest using a gradient annealing temperature for PCR. They found remarkable enhancements in the amplification efficiency for methylated DNA by increasing the annealing temperature for PCR, and thus overcome PCR bias in quantitative methylation analysis. Based on our results, we highly recommend mixing experiments using varying mixtures of methylated and un-methylated DNA and gradient annealing temperature for PCR to initially set up, evaluate, and calibrate each new assay.	Shen, L., Guo, Y., Chen, X., Ahmed, S., & Issa, J. (2007). Optimizing annealing temperature overcomes bias in bisulfite PCR methylation analysis. Biotechniques, 42(1), 48.	Warnecke, P. M., Stirzaker, C., Melki, J. R., Millar, D. S., Paul, C. L., & Clark, S. J. (1997). Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. Nucleic acids research, 25(21), 4422-4426.
17	Handelsman et al.	2008	Primer Degeneracy	As the 16S rRNA sequence database has grown, it has become evident that many sequences deviate within the most conserved regions targeted by “universal” 16S rRNA gene PCR primers. So to tackle that, primers have been modified with degenerate bases to allow primers to target a wide range of genes. Also because PCR polymerase requires primers ranging from 20-30nt length, 16s primers have been limited to target regions of only those lengths. So rather than using the usual primers of length 20-30nt, the authors use a smaller 16s RNA gene PCR primer to increase the scope further. These primer are known as "mini-primers". Two polymerase enzymes, S-Tbr and Taq were used for PCR. Three soil and nine microbial mat samples were purified, extracted and amplified using modified 27F-P and 1492R-P primers for initial testing. Several mini-primers were designed (9-11nt) and tested. These mini-primers were chosen for the most conserved 10nt regions within the 16s rRNA gene primers.	S-Tbr showed better response towards primers ranging from lengths 10-20nt. Taq on the other hand showed a band on the gel with only 20nt primer length. Mini-primer PCR amplified a greater proportion than did standard primers of sequences that were novel or that poorly matched a database of previously isolated 16S rRNA gene sequences. The expected promiscuous primer-template binding but this was not problematic in case of environmental samples. In case of mini-primer PCR characterization of the microbial community, the mini-primer method appears to amplify more novel sequences than the long-primer methods. Of the 1,129 predicted mini-primer amplicons in the NCBI environmental sequence database, 61 (5.4%) were found not to be 16S rRNA gene sequences; a similar calculation for long primers yields a similar false-positive rate (19/320 approx. 5.9%). In addition, the rates of identifying mitochondrial and chloroplast sequences were also the same (8/685 approx. 3/205 or approx. 1%) for both methods. Authors suggest that adding mini-primer PCR to the tools used for analyzing microbial communities may enable a more accurate measure of 16S rRNA gene sequences in environmental samples by expanding the sequences detectable by PCR. In addition to expanding the range of detectable targets, combining mini-primer PCR with standard techniques might increase the accuracy of environmental sampling by enabling estimates of microbial diversity to reflect sampled communities more closely. Lastly, the study contemplates that the combination of highly processive enzymes and shorter primers may decrease the cost of PCR. The increased processivity enables PCR to be performed with smaller amounts of enzyme, and shorter primers are less costly to synthesize.	Isenbarger, T. A., Finney, M., Rios-Velázquez, C., Handelsman, J., & Ruvkun, G. (2008). Mini-primer PCR, a new lens for viewing the microbial world. Applied and environmental microbiology, 74(3), 840-849.	

S.No.	Manuscript	Year	Bias Discussed	Experimental Design	Implementation Effect	Reference	Similar Studies
18	Dobrovic and Wojdacz et al.	2008	Annealing Temp.	This study concentrates on the important criteria that choice of primers is critical in bisulfite based methylation-screening protocols. The proportional amplification of all templates is critical but difficult to achieve due to PCR bias favoring the amplification of the un methylated template. The authors emphasize on the fact that amplification of methylated templates is necessary but choose a different strategy to improve it, optimization of annealing temperature. The degree of bias correction can be manipulated by varying the annealing temperature to control the stringency of binding of the primers to the template. At lower annealing temperatures, there is little favoring of the methylated template by the primers. At higher annealing temperatures, the primers will bind almost exclusively to the methylated template allowing the reversal of PCR bias and amplifying mostly methylated template. At an intermediate annealing temperature, the primer annealing bias that favors methylated templates will compensate for the amplification bias that favors un-methylated templates.	The study states that in methylation specific PCR (MSP) primers, it is important to include several CpG sites towards the 3' end of the primers as the methylation status of a specific locus is determined by the CpG sites within the primer sequence. The authors suggest that primers should be designed in a way that, include a limited number of CpG dinucleotides in the primer, these CpG's should be as far away as possible from 3' of the primer and melting temperature of the primer should be around 65°C.	Wojdacz, T. K., Hansen, L. L., & Dobrovic, A. (2008). A new approach to primer design for the control of PCR bias in methylation studies. BMC research notes, 1(1), 54.	Warnecke PM, Stirzaker C, Melki JR, Millar DS, Paul CL, Clark SJ: Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. Nucleic Acids Res. 1997, 25: 4422-4426. 10.1093/nar/25.21.4422.
19	Epstein et al.	2009	Primer Degeneracy	Marine sediments were collected and sequenced against two primer sets, 27F-1492R and 8F and 1542 R. Their strategy involves using a single primer set and a pooled primer set to predict the total microbial richness as it appears from individual libraries as well as pooled data. They hypothesize that pooled data are no more than a result of an increase in sequencing efforts.	The results show that just using one specific primer set to amplify 16s rRNA genes missed almost 50% of microbial diversity. The study recommends using primer pools, multiple DNA extraction techniques, and deep community sequencing to minimize the bias and help capture more species.	Hong, S., Bunge, J., Leslin, C., Jeon, S. & Epstein, S.S. Polymerase chain reaction primers miss half of rRNA microbial diversity. ISME J. 3, 1365–1373 (2009).	
20	Zhou et al.	2010	PCR Cycle number Starting Template Conc.	Sediment samples were used and sequenced using the 967F and 1046R primer set using the Solexa platform. Taxonomy classification was performed by assigning the reads of each sample to the 16s V6 region. Final analysis was done using Mothur software package.	Dilution of starting genomic DNA resulted in reduced richness 30 cycles of PCR resulted in increased artifacts and lower taxa richness in comparison to 25 cycles of PCR. Number of PCR cycles did not change the overall microbial community.	Wu, J. Y., Jiang, X. T., Jiang, Y. X., Lu, S. Y., Zou, F., & Zhou, H. W. (2010). Effects of polymerase, template dilution and cycle number on PCR based 16 S rRNA diversity analysis using the deep sequencing method. BMC microbiology, 10(1), 255.	
21	Bellemain et al.	2010	Primer Degeneracy	This study is regarding fungal sequencing of the ITS region to analyze fungal diversity in environmental samples. This study explore the potential amplification biases that various commonly utilized ITS primers might introduce during amplification of different parts of the ITS region in samples containing mixed templates. This study concentrated on mainly three things, firstly, to what degree the various primers mismatch with the template sequence and whether the mismatches are more in favor of some taxonomic groups. Secondly, to study the variation in length in the final PCR products, in relation to the taxonomy to assess amplification biases during real PCR amplification. This is because shorter DNA fragments are preferentially amplified from environmental samples containing DNA from a combination of various species. Finally, to what extent the various primers co-amplify plants, which often co-occur in environmental communities and samples? Three different datasets were amplified using different internal primer sets. Dataset 1 was amplified using ITS1-F-ITS2, ITS5-ITS2 and ITS1-ITS2, dataset t2 using ITS1-ITS4, ITS3-ITS4 and ITS5-ITS2 and dataset 3 using ITS3-ITS4 and ITS3-ITS4B. 0-3 mismatches were also allowed between each primer-template. The EcoPCR package was used to do the final analyses.	The results of the study showed that including Primer ITS1-F, some other primers hampered with a high proportion of mismatches relative to the target sequences, and most of them acted to introduce taxonomic biases during PCR. For example. Primers ITS1-F, ITS1 and ITS5, were biased towards amplification of basidiomycetes, whereas others, example ITS2, ITS3 and ITS4, were biased towards ascomycetes. The assumed basidiomycete-specific primer ITS4-B only amplified a minor proportion of basidiomycete ITS sequences, even under lenient PCR conditions. In case of primer mismatches, selected ITS primers showed large variability in their ability to amplify fungal sequences from the three subsets when tolerating varying number of mismatches. With the exception of ITS4-B, all primer pairs amplified at least 90% of the sequences when allowing two or three mismatches. The results also indicate that the percentages of sequences were quite similar for two and three mismatches, indicating that rather few sequences included three mismatches. The authors observe that because there is a systematic length differences between the ITS2 region and the entire ITS, ascomycetes will more easily amplify than basidiomycetes using these regions as targets. This bias can be avoided by using primers amplifying ITS1 only, but this would imply preferential amplification of 'non-dikarya' fungi. The authors suggest using different primer combinations or different parts of the ITS region to reduce bias.	Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P., & Kauserud, H. (2010). ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. BMC microbiology, 10(1), 189.	
22	Weon et al.	2012	PCR Cycle number	15 bacterial strains were cultured and pyrosequenced for the V1 region using 9F-541R primer set. Two types of polymerase (Phusion and Taq) were used and the whole reaction was amplified for 15 and 30 cycles each. Final analysis was done using Mothur software package.	16s rRNA recovered from all four datasets. The number of PCR cycles significantly affected the proportion of identified chimeric sequences, which were reduced from 32% (Taq) or 36%(Phusion) to below 1% by decreasing the PCR cycle number from 30 to 15. This study showed that bacterial richness was overestimated at increased PCR cycle number mostly due to the occurrence of chimeric sequences. Although the decrease in PCR cycle number will also reduce the amount of PCR product, it is recommended that PCR cycle number should be kept as low as possible for more accurate estimation of bacterial richness.	Ahn, J.-H., Kim, B.-Y., Song, J. & Weon, H.-Y. Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. J. Microbiol. 50, 1071–1074 (2012).	

S.No.	Manuscript	Year	Bias Discussed	Experimental Design	Implementation Effect	Reference	Similar Studies
23	Quan et al.	2012	Primer Degeneracy	Published study intends to study the issue of overestimation by assessing the coverage of 8 bacterial primers by using 7 metagenomic datasets as well as the Ribosomal Database Project (RDP). The 8 primers evaluated for this study are 27F(8F), 338F, 338R, 519F, 519R(536R), 907R(926R), 1390R(1406R) and 1492R. Influence of single primer mismatch occurring within the last 4 nucleotides was also studied.	The study revealed that the bias caused by primer-template mismatch and use of universal bacterial primers may misrepresent the real bacterial community of the samples. Single primer mismatch occurring within the last 4 nucleotides had a very weak influence at domain level. The most noticeable non-coverage rate was observed for 338F in the phylum *Lentisphaerae* and *OP3* . When allowed a single mismatched within the last 4 nucleotides, its non-coverage rate was only 3%; otherwise, it was as high as 100%. This indicated that primer 338F is not appropriate for either *Lentisphaerae* and *OP3*.	Mao, D.-P., Zhou, Q., Chen, C.-Y. & Quan, Z.-X. Coverage evaluation of universal bacterial primers using the metagenomic datasets. BMC Microbiol. 12, 66 (2012).	
24	Rasking and Pinto	2012	Artifacts	This study concentrates on understanding the effect of multi-template PCR bias on microbial community composition. The authors used three bacterial and three archaeal mock communities consisting of, respectively, 33 bacterial and 24 archaeal 16S rRNA gene sequences combined in different proportions to compare the influences of (1) sequencing depth, (2) sequencing artifacts (sequencing errors and chimeric PCR artifacts), and (3) biases in multi-template PCR. Primers Bact-338F/Bact-909R and Arch-340F/Arch-915R were designed for the V3-V5 region of the 16s rRNA gene. The final analyses was done using the Mothur package.	The study shows that greater sequencing depth does not always result in a more accurate representation of the sequenced community, since the errors in mean relative abundance due to multi-template PCR bias significantly alter the rank abundance distributions. In another observation made by the study, the number of sequences in each sample library may be affected by the GC content of the amplicon pool of each sample.	Pinto, A. J., & Raskin, L. (2012). PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. PloS one, 7(8), e43093.	
25	Klindworth, A. et al.	2013	Primer Degeneracy	Two separate PCR reactions were performed using 175 single primers and 512 primer pairs, and pyrosequenced. Overall alignment and taxonomic classification was done using SILVA-SINA.	Coverage of primers is different for different samples. For example, in case of Archaeal samples, primers S-D-Arch-0519-a-A-19 (A: 91.3%, B: 0.1%, E: 1.0%) and S-D-Arch-0787-a-A-20 (A: 87.4%, B: 7.8%, E: 0.0%) performed significantly well. The highest overall coverage and specificity for the domain Bacteria was detected for the primers S-D-Bact-1061-a-A-17 (A: 2.9%, B: 96.4%, E: 0.0%) and S-D-Bact-0564-a-S-15 (A: 16.3%, B: 96.0%, E: 0.0%). Furthermore, 39 primers show relatively high overall coverage for more than one domain. The results conclude that commonly used single primers exhibit significant differences in overall coverage and taxa at phylum level. Therefore, the authors suggest that primer should be carefully selected to avoid accumulative bias.	Klindworth, A. et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 41, e1 (2013).	
26	Tsuda and Kurokawa et al.	2013	Primer Degeneracy	Universal primers have revolutionized the deep sequencing of 16s rRNA genes but some universal primers also amplify eukaryotic rRNA genes, leading to a decrease in the efficiency of sequencing of prokaryotic 16S rRNA genes with possible mischaracterization of the diversity in the microbial community. This study used 50 candidates to calculate their coverage for prokaryotic and eukaryotic rRNA genes, including those from uncultured taxa and eukaryotic organelles, and a novel universal primer set, 342F-806R. This primer set was validated by the amplification of 16S rRNA genes from a soil metagenomic sample and subsequent pyrosequencing using the Roche 454 platform. The same sample was also amplified using 338F-533R and pyrosequenced. Shotgun sequencing was done using Illumina MiSeq platform. Final analysis was done using Mothur package.	Results indicate that non-degenerate 342F-806R doesn't produce substantial bias in the microbial community. The broad-range primers targeting bacterial and archaeal 16S rRNA genes tended to also amplify eukaryotic rRNA genes.	Mori, H., Maruyama, F., Kato, H., Toyoda, A., Dozono, A., Ohtsubo, Y., ... & Kurokawa, K. (2013). Design and experimental application of a novel non-degenerate universal primer set that amplifies prokaryotic 16S rRNA genes with a low possibility to amplify eukaryotic rRNA genes. Dna Research, 21(2), 217-227.	
27	Neufeld at al.	2014	Input DNA conc	This study concentrated on testing the effects of template concentration and pooling of PCR amplicons with paired-end Illumina sequencing of bacterial 16S rRNA genes in two soil samples. Using DNA extracts from soil and fecal samples as templates, pooled amplicons and individual reactions were sequenced for both high (5- to 10-ng) and low (0.1-ng) template concentrations.	The results indicates that the practice of pooling multiple PCR amplicons prior to sequencing contributes proportionally less to reducing bias. The results confirmed that high template concentrations (i.e., 5 to 10 ng per reaction) increased the accuracy of sample OTU profiles compared to low template concentrations (i.e., 0.1 ng per reaction).	Kennedy, K., Hall, M.W., Lynch, M.D.J., Moreno-Hagelsieb, G. & Neufeld, J.D. Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. Appl. Environ. Microbiol. 80, 5717–5722 (2014).	

S.No.	Manuscript	Year	Bias Discussed	Experimental Design	Implementation Effect	Reference	Similar Studies
28	Green et al.	2015	Primer Degeneracy Annealing Temp	GenomicDNA from mammalian feces and sediment samples from lake Huron were extracted, purified and amplified using the CS1_515F-CS2-806R primer set. Four mock communities Mock A, B, C and D were used, each almost identical barring a 10-bp region in the center of each fragments that was scrambled to create a unique identifier sequence. A novel PCR workflow, PEX-PCR was employed to test out the hypothesis. To validate the technique, a standard PCR assay (TAS) targeting the V4 variable region of microbial SSU rRNA genes was employed as a reference.	The samples that were run through PEX-PCR were able to anneal to primers even at a lower annealing temperature. Ideal Score was calculated for all samples and the ones processed with the PEX PCR method had significantly lower IS scores for all temperatures. Also the observed microbial community was significantly different when the samples was processed using different PEX PCR annealing temperatures due to annealing temperature-associated shifts in the relative abundance of individual taxa. This showed that TAS scrambled the signal and basically showed almost similar results even after changing annealing temperatures. This study showed ways of allowing exponential amplification of the target by primers with no mismatch. This helps in reducing bias associated with primer mismatches significantly by limiting the primer-template interactions to 2 cycles only. Also by employing a low annealing temperature, the PEX-PCR is more likely tolerant of low primer annealing efficiency of primer-template pairings with mismatches. Also, 3' mismatches can be overcome using the PEX PCR method.	Green, S. J., Venkatramanan, R., & Naqib, A. (2015). Deconstructing the polymerase chain reaction: understanding and correcting bias associated with primer degeneracies and primer-template mismatches. PloS one, 10(5), e0128122.	
29	D'Amore, R. et al.	2016	PCR Cycle number Starting Template Conc.	Two synthetic communities with an even and uneven distribution of archaeal and bacterial species were used as metagenomic control material. These were run using three different sequencing platforms, MiSeq, PacBio and IonTorrent along with varying PCR conditions such as PCR cycles and stating template conditions. Results were generated using AMPLImock pipeline.	The only consistent effect observed in this study is that the error rate increased marginally significant when associated with more PCR cycles. The analysis also shows that the amount of initial genomic DNA template concentration and PCR cycles acts as a factor in increasing artifacts such as chimeras. When starting template concentration was increased from 1ng to 10ng, percentage chimeric reads increased from approximately 8.75% to 21%. When PCR cycles were increased from 15 to 25 cycles, the percentage chimeric reads increased from approximately 10% to 90%.	D'Amore, R. et al. A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. BMC Genomics 17, 55 (2016).	
30	Gohl et al.	2016	PCR Cycle Number Annealing Temp. Primer Mismatches	A mock DNA community was obtained. Fecal samples were also collected from doucs. The samples were PCR amplified for the v4 region of the 16s rRNA gene using the 515F-806R primer set. Annealing temperature was set at 50°C. PCR products were sequenced on an Illumina MiSeq sequencer and final analysis was done using QIIME software package.	The results indicate a considerable reduction in chimeric reads when the number of PCR cycles are low. The authors have called primer mismatches as deleterious, one that can cause serious damages. They specifically talk about mismatches happening in the 3-4 bp at the 3' end of the primer. They also suggest not using sequencing primers that overlap with amplification primers.	Gohl, D. M., Vangay, P., Garbe, J., MacLean, A., Hauge, A., Becker, A., ... & Knights, D. (2016). Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. Nature biotechnology, 34(9), 942-949.	

PCR bias caused by amplification cycle number

One of the earliest studies of PCR bias was published in 1996 by Suzuki and Giovannoni [28]. In this study, bias caused by primer mismatch and PCR amplification cycle number was addressed. The study included samples that were either linearized plasmids containing cloned 16S ribosomal RNA gene inserts or fragments of 16S rRNA genes generated with primers 27F and 1492R by PCR. The sources of these amplicons were clone libraries of bacterial SSU rRNA genes. In the experiment, two different sets of templates were used to evaluate the introduction of biases by PCR. The first template consisted of a mixture of purified plasmids containing SSU rRNA genes from SAR202 and SAR464 (two bacterial 16S rRNA genes) and linearized by digestion with a restriction endonuclease. The second set of templates consisted of mixtures of the PCR amplified fragments of SAR202, SAR432, and SAR464 templates, each of which had been amplified separately by PCR from linearized plasmids. All template mixtures were added to final concentrations of 0.1 ng/ml. PCR conditions for each analysis were similar, with variables including: (a) primers used, (b) templates used, and (c) number of PCR cycles. The number of PCR cycles used were 10, 15, 25, or 35 cycles. In order to understand and interpret PCR bias better, kinetic numerical models were developed using the software Stella (High Performance Systems, Inc., Hanover, N.H.). When using two different primers to amplify 16S rRNA gene template mixtures from three different bacteria, a strong divergence in results between primer sets was observed. Under the experimental condition used, if we assume PCR to be non-biased, the abundance of all three bacteria should be fairly equal at the end of the amplification. The results of this study indicate otherwise; a strong bias was observed, which was dependent more on the choice of primers than to the choice of templates. For the 519F-1406R primer pair, little or no bias and a low product yield was observed. In contrast, for the 27F-338R primer pair, a

strong bias and a much higher yield of product was observed. This observed results were then matched with an amplification model. The authors developed a hypothesis that indicates that an increase in the product of a dominant template will start to inhibit further copying of this template through competition between primers and amplification products for priming sites. Thus, the authors indicate, that in reactions with mixed templates and high amplification efficiency, the template with higher initial concentration will reach inhibitory concentrations sooner, leading to even final product abundance, regardless of input concentration [28]. This condition is worsened by increasing numbers of PCR cycles, and thus it is recommended to limit the number of PCR cycles to remain at concentrations of products below inhibitory levels. The longer amplicon, the 519F-1406R primer set, has an overall lower amplification efficiency. Therefore, products levels do not rise to the same levels as for the shorter primer set during the PCR cycle, thereby remaining at sub-inhibitory levels and resulting in more even amplification of templates.

PCR bias caused by high cycle number was also discussed by [30]. Samples in this study were PCR amplified using two primer set targeting gene transcripts. This study observed that the rate of amplification of abundant PCR products declined faster than that of the less abundant products, similar in manner to Suzuki and Giovannoni [28]. In particular, this phenomenon appears to occur in the late cycles of PCR amplification when PCR products reach a high concentration. This reannealing can cause interference with primer binding and elongation. This phenomena is known as the 'C_ot effect', first suggested by Innis and Gelfand [33]. The paper concludes that based on its results, re-hybridization appears to be responsible as a bias against PCR of abundant products in the late cycles of PCR.

Wang and Wang [34, 35] further explored the effects of PCR amplification cycles in creating chimeras and causing PCR bias. In addition to issues with preferential annealing leading to distorted relative abundance, Wang and Wang also identified cycle number as a major contributor to the creation of chimeric artifacts. Chimera formation occurs in microbiome amplifications when the 16S rRNA genes of more than one species are amplified in a single reaction, and artifactual sequences not present in the original sample are created when incompletely amplified fragments from one gene anneal to a different gene. Such artifacts can lead to substantial effects on diversity analyses performed as part of microbial ecology studies [36]. The aim of this study was to evaluate the formation of PCR artifacts and also examine the effect of PCR cycles on the reaction. 16S rRNA genes of four Actinobacteria were used to test for chimera formation. Results indicated that the type of *Taq* polymerase had little effect on formation of artifacts, but that detectable artifacts were significantly lower in samples that were amplified for 10 cycles relative to those amplified for 20 cycles. In addition, shorter elongation times (2 min vs 5 min) resulted in more artifacts, presumably due to a higher proportion of incompletely copied fragments in reactions with 2 min elongation times.

Zhou et al. [37] systematically tested bias caused by starting concentrations of genomic DNA (0.1, 1.0 and 10 ng/ml), PCR elongation time (20s, 2mins and 4 mins), polymerase type (Z-Taq, LA-Taq, and AmpliTaq), and cycle number (22, 25 and 28 cycles) using model communities constructed from Alpha-, Beta- and Gamma classes of Proteobacteria. 40 different Gram-negative and positive bacteria were amplified using the primer set FD1 - R1540, and all reaction conditions were performed in triplicate. PCR artifacts were significantly different in reactions with three different polymerases, and were opposite of the theoretical prediction. Briefly, the frequency of chimeras for both Z-Taq (8.7%) and LA-Taq (6.2%) polymerases with higher

processivity and accuracy was higher than that for AmpliTaq (2.5%). The frequencies of chimeras and of heteroduplexes for Z-Taq were almost three times higher than those of AmpliTaq. Chimeras also increased with higher PCR cycles and higher starting template concentrations. PCR artifacts decreased with higher elongation time with the highest with observed at 20 seconds (25.5%). Sample and species diversity also played a role in artifact formation, with higher species diversity resulting in higher artifact formation. The paper suggests that cycle number should be decided by the amount of template used, amplification efficiency, and existence and degree of inhibitory substances. All these factors should be experimentally determined for higher efficiency and throughput. Such recommendations, however, are rarely used!

In PCR amplification reactions intended for next-generation sequencing (NGS), a two-stage PCR protocol is often used [38]. Briefly, genomic DNA is amplified using locus-specific primers containing 5' linkers. The amplicons generated from these reactions are then transferred to a second stage amplification employing primers containing Illumina sequencing adapters, a sample-specific barcode, and the same linkers, but located at the 3' ends of the oligonucleotides. In a systematic test, I examined the relationship between numbers of PCR cycles performed in the first stage and second stage on rate of chimera formation [39]. In this test it was observed that the only relevant variable was total number of PCR cycles between both stages, with rates of chimera formation near 50% with high total cycle number. These data are relevant as the two-stage PCR protocol is used throughout the study.

PCR bias caused by annealing temperature

The issue of bias caused by annealing temperature was discussed by [27]. Annealing temperature has been thought to play a role in biased amplification, as sequences with high GC contents may difficult to amplify due to low efficiency of template dissociation. Thus, additives such as acetamide – which serve as DNA denaturants – may improve evenness of amplification. More broadly, however, Ishii and Fukui [27] considered the kinetics of binding between a complementary single oligonucleotide, such as a primer or a probe, and targeted DNA vary depending on temperature. At lower hybridization temperatures, a primer with a perfect match to the target and a primer with one mismatch to the target can hybridize at similar rates. Ishii and Fukui suggest that PCR bias caused by primer binding energies can be reduced by running PCR at low annealing temperatures, and the authors addressed this using an experimental system. Genomic DNA from two bacteria was used as template for amplification with primers 341F-907R. To reduce the complexity of the study, the 907R primer was free of degeneracies (note: we also used this approach to reduce experimental complexity in our experimental study described in **Chapter V**). Aside from annealing temperature (45, 50, 55, and 60°C), PCR conditions were identical for all reactions. A differential yield of amplification product was suggested to be due to a difference in GC content of templates. For example, in a reaction with two templates both perfectly matching the primers and present at roughly equimolar starting concentration, PCR bias led to final ratios of 1.61, 1.70, and 2.12 at the end of reactions run at 45°C, 50°C and 55°C. In a test with one template with a single mismatch with the primer set, the final ratios were 1.47, 0.87 and 0.17 at the same three temperatures. The dominant template had a higher GC content than the more poorly copied template, suggesting that primer-template mismatches exert a stronger effect than GC content. Furthermore, distortion was substantially

lower at lower annealing temperatures, indicating that lowered annealing temperature can reduce bias associated with primer-template mismatches. Hongoh et al. [40] demonstrated an increase in number of detectable microbial phylotypes from environmental samples when annealing temperature was lowered, and PCR cycles were increased (note: increased PCR cycles can increase chimera formation, and this can lead to a spurious increase in microbial phylotypes. We observed consistently lower phylotypes using the DePCR method relative to standard PCR, and this was concordant with significantly lower chimera formation rates; **Chapter III**). These results indicate that annealing temperature is a very important variable, and lower PCR annealing temperatures can be used for detecting a broader range of organisms. Not all reaction systems favor low annealing temperatures, however. In amplification of bisulfite-treated DNA, increased annealing temperatures favor improved evenness of recovery of methylated and un-methylated DNA (*e.g.*, Shen et al. [41]). In such systems, differing GC content in methylated and un-methylated DNA may lead to secondary structures inhibitory to amplification. Thus, high annealing temperatures may help remove such inhibition.

PCR bias caused by genomic DNA concentration

The end product of a PCR protocol can also be governed by how much starting template concentration was used. Zhou et al. [37] examined the effect of higher input gDNA concentration at the start of PCR reactions, and observed that higher concentrations coupled with higher amplification cycles resulted in greater frequency of detectable chimeras. Similarly, elevated chimera formation was observed with increased DNA input in a recent study [26]. When starting DNA template concentration were increased from 1 ng to 10 ng, the percentage chimeric reads increased from approximately 8.75% to 21%, and when PCR cycles were increased from 15 to

25 cycles, the percentage chimeric reads increased from approximately 10% to 90%. In a separate study by Neufeld et al. [25], the effects of template concentration and pooling of PCR replicates were examined in complex environmental samples. Using gDNA extracts from soil and fecal samples as templates, pooled amplicons and individual reactions were sequenced for at high (5- to 10-ng) and low (0.1-ng) input template concentrations. The results of the study indicated that the practice of pooling multiple PCR amplicons prior to sequencing contributes proportionally less to reducing bias, while high template concentrations (*i.e.,* 5 to 10 ng per reaction) increased the accuracy of OTU profiles compared to low template concentrations (*i.e.,* 0.1 ng per reaction). This likely represents reduced stochastic effects with increasing DNA concentration. Conceivably, with increased DNA input levels, the number of cycles of PCR amplification could be reduced, leading to lower overall artifact formation.

PCR bias caused by degenerate primers

One of the first studies to discuss primer degeneracy and its impact on the overall PCR output was by Wagner et al. [14]. In the context of sampling gene families, the paper studied those factors that produced skewed distribution of templates. This study suggested two major classes of processes leading to such distortion - PCR selection and PCR drift. PCR selection occurs when the reaction favors certain members of a gene, and PCR drift is the result of random events occurring in the early cycles of the reaction. Wagner et al. [14] suggested that both selection and drift likely occur concurrently. Following on this, Polz and Cavanaugh made several recommendations regarding bias reduction. First, the authors recommend avoiding degeneracies when designing universal primers, starting with high template concentration to avoid stochastic

effects, pooling of replicate amplifications to reduce PCR drift, and finally, reducing amplification cycles to reduce skewed results caused by PCR selection.

Although reduced primer degeneracy may be desirable, it is not always feasible. In microbial DNA samples, when primers anneal to gDNA templates, the potential positions and numbers of mismatches is very large due to high sequence diversity of ribosomal RNA genes, even in highly conserved primer regions [42, 43]. This fundamental problem was addressed by Isenbarger et al. [44]. The authors indicated that as the 16S rRNA gene sequence database has grown, it has become evident that many sequences deviate within the most conserved regions targeted by "universal" 16S rRNA gene PCR primers. One solution to this was to utilize short primers ('miniprimers') targeting only the most highly conserved nucleotide stretches of 16S rRNA genes. However, the short primers required custom polymerases to allow for polymerase extension.

Bellemain et al. [45] explored potential amplification biases derived from the use of commonly utilized primers targeting fungal internal transcribed spacer [46] regions. This study concentrated on three areas: (i) to what degree the various primers mismatch with the template sequence and whether the mismatches favor of some taxonomic groups; (ii) to what degree variation in length in the final PCR products led to bias in amplification; and (iii) to what extent the various primers co-amplify plants, which often co-occur in environmental communities and samples? This study therefore addresses the need for degenerate primer pools to target a broad range of microbial taxa, as well as the potential for mismatch amplification with off-target sequences. The results of the study showed that including many primers had a high proportion of mismatches relative to the target sequences, and most of them acted to introduce taxonomic biases during PCR. In case of primer mismatches, selected ITS primers showed a large variability in their ability to amplify

fungal sequences while tolerating a varying number of mismatches. With one exception, however, all primer pairs amplified at least 90% of the sequences when allowing (primarily) two mismatches.

PCR bias caused by simultaneous linear and exponential amplification

As part of this work, a previously undescribed form of PCR bias that leads to differential amplification of templates has been identified. We note that amplification efficiency has been shown to change during PCR, and starting efficiency of template amplification may not match end-point amplification efficiency [28]. However, in prior studies, the focus has been competition between primers and amplicons for annealing sites on templates. What has not previously been considered is the difference in primer-template interactions between primer-input DNA template and primer-DNA copy interactions. In a recent publication [18], two distinct types of primer-template interactions were identified, and it was suggested that these types of interactions likely operate at different efficiencies. The first primer-template interaction is that of the PCR primers with the source genomic DNA; such interactions include both perfect matches and primer annealing to regions containing mismatches of unpredictable number and location. Subsequently, however, PCR primers interact with primer sites that have been created during the PCR cycles, and are the inverse complement of the synthesized oligonucleotides used as PCR primers. Here, the primer-template interactions can also include perfect matches and mismatches, but the scale of mismatch annealing is proportional to the number of degeneracies in the primer pool. Primer annealing to genomic DNA were labeled "natural" interactions and primer annealing to PCR amplicons were labeled as "artificial" interactions (**Figure 1**). Regardless of the degeneracy of the primer pool, natural template-primer interactions are likely to be more

complex due to the great potential for the presence of primer-template mismatches as a result of the underlying genetic diversity at primer sites. Artificial template-primer interactions are less complex, and the number of potential mismatches between primer and template is proportional to the diversity of the primer pool employed.

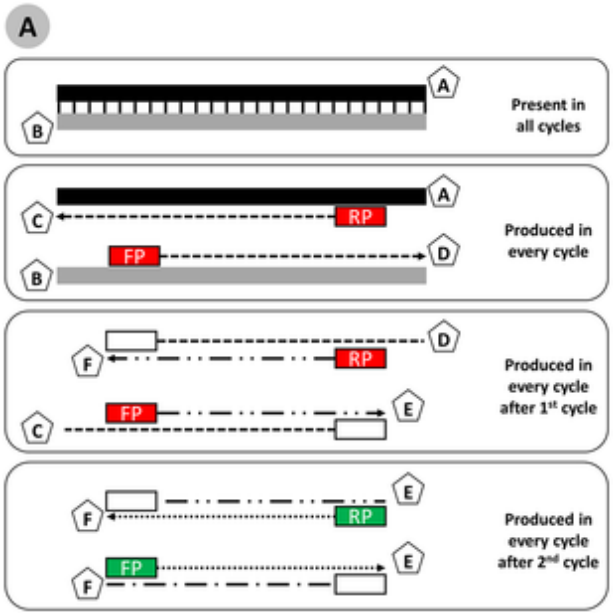


Figure 1: Types of DNA fragments found in PCR

(A). Template DNA fragments (containing strands "A" and "B") are added to PCR reactions and are conserved throughout the reaction. Fragments "A" and "B" serve as templates for copying in each cycle, with hybrid molecules "C" and "D" produced in a linear fashion each cycle. In cycles two and above, "C" and "D" are copied, creating hybrid molecules "E" and "F" in a linear fashion. In cycles three and above, the "E" and "F" fragments generated in prior cycles are copied into inverse complement fragments "F" and "E", respectively, in an exponential fashion. Red boxes indicate 'natural' primer annealing to genomic DNA template or copy of gDNA template. Green boxes indicate 'artificial' primer annealing to primer sites that are copies of

oligonucleotide primers added to the PCR mixture, and incorporated during previous cycles Green et al. [18].

The solution to this form of bias is the deconstructed PCR methodology that serves as the foundation for this thesis. The essential function of the DePCR approach is to allow only the minimum number of cycles (*i.e.*, 2) for 'natural' primer-template interactions. The method does not solve any fundamental differences in efficiency during these two cycles, but limits these lower efficiency interactions to only two cycles. After the first two cycles are complete, the locus-specific primers with 5' linker overhangs are removed using exonuclease or amplicon cleanup, and subsequent exponential amplification is performed with non-degenerate primers targeting linker sequences.

Chapter II: Deconstructing the polymerase chain reaction: understanding and correcting bias associated with primer degeneracies and primer-template mismatches

(Previously published as Green, S. J., Venkatramanan, R., & Naqib, A. (2015). Deconstructing the polymerase chain reaction: understanding and correcting bias associated with primer degeneracies and primer-template mismatches. PloS one, 10(5), e0128122.)

Abstract

The polymerase chain reaction (PCR) is sensitive to mismatches between primer and template, and mismatches can lead to inefficient amplification of targeted regions of DNA template. In PCRs in which a degenerate primer pool is employed, each primer can behave differently. Therefore, inefficiencies due to different primer melting temperatures within a degenerate primer pool, in addition to mismatches between primer binding sites and primers, can lead to a distortion of the true relative abundance of targets in the original DNA pool. A theoretical analysis indicated that a combination of primer-template and primer-amplicon interactions during PCR cycles 3-12 is potentially responsible for this distortion. To test this hypothesis, we developed a novel amplification strategy, entitled "Polymerase-exonuclease (PEX) PCR", in which primer-template interactions and primer-amplicon interactions are separated. The PEX PCR method substantially and significantly improved the evenness of recovery of sequences from a mock community of known composition, and allowed for amplification of templates with introduced mismatches near the 3' end of the primer annealing sites. When the PEX PCR method was applied to genomic DNA extracted from complex environmental samples, a significant shift in the observed microbial community was detected. Furthermore, the PEX PCR method provides a mechanism to identify which primers in a primer pool are annealing to target gDNA. Primer utilization patterns revealed that at high annealing temperatures in the PEX PCR method, perfect match annealing predominates, while at lower annealing temperatures, primers with up to four mismatches with templates can contribute substantially to amplification. The PEX PCR method is simple to perform, is limited to PCR mixes and a single exonuclease step which can be performed without reaction cleanup, and is recommended for reactions in which degenerate primer pools are used or when mismatches between primers and template are possible.

Introduction

To target single gene fragments from multiple organisms within a complex community of known and unknown organisms using PCR has required careful bioinformatics analyses and empirical testing of many primers. Ideal criteria for primers include: (i) primers should match genes of all known organisms within the group of interest, and should be able to target genes from unknown organisms in sub-taxonomic levels (*i.e.* domain-level primers targeting Bacteria should be conserved among all known bacteria, with the assumption that unknown bacteria will also contain these conserved regions), (ii) primer pairs should be balanced for melting temperature and produce robust amplification, and (iii) primers need to span one or more hyper-variable regions of the gene. Even in highly conserved regions of the ribosomal RNA (rRNA) genes, no primer set matches these criteria perfectly, although many excellent primer sets have been developed [10, 44, 47-49]. To increase target range, pools of primers (degenerate primers) are used. Ribosomal RNA genes are the preferred targets for broad-spectrum analyses as the level of genetic diversity in conserved regions of rRNA genes is lower than that present in functional genes where amino acid sequences can be highly conserved even in the presence of substantial DNA-level changes due to the degeneracy of the genetic code. Therefore, degenerate primers used for PCR amplification of rRNA genes generally have lower levels of degeneracy than primers used for amplification of functional genes.

Amplicon sequencing approaches for microbial surveys, including those targeting rRNA genes, are limited in several ways: (i) amplicon sequencing studies of mixed microbial communities fundamentally distort the "true" structure of the community through systematic bias associated with input genomic DNA (gDNA) concentration, differential amplification efficiency due to mismatches between primer and template, differential amplification efficiency within

mixtures of degenerate primer pools and varying melting temperature, through differential consumption of specific primers in degenerate primer pools during later cycles of PCR, through PCR saturation during late PCR stages, and through preferential amplification of specific targets [15, 17, 24, 27, 28, 50-53]; (ii) even degenerate primer pools are not degenerate enough to target all the intended targets [52], and this is particularly true for many functional genes; (iii) ribosomal RNA genes come with a built-in bias: namely, the copy number variation from organism to organism distorts the observed structure of the community, and favors organisms with high copy number (this has been partially addressed in [54]); and (iv) quantitative analyses are even more sensitive to varying amplification efficiencies of degenerate primers as high efficiency of PCR amplification is necessary, and distortions may occur if the quantitative calibration standards have no mismatches between primer and target, while environmental samples contain mixtures of perfectly matching and single- and multiple-mismatch targets. Taken together, direct PCR-sequencing and quantitative PCR analysis of microbial rRNA (and other) genes is certainly providing a distorted composition and structure, and some taxa are not detected at all [55].

This study describes a new methodological approach to PCR amplification to address distortions due to variable primer melting temperature in degenerate primer pools and distortions arising from mismatches between the primer and template sequence. The method also provides a mechanism to determine which primers from a primer pool are involved in annealing to gDNA templates, and what number of mismatches can be tolerated by a given primer pair.

To validate the technique, a standard PCR assay targeting the V4 variable region of microbial SSU rRNA genes was employed. This assay has been previously described [56], [49], and is employed by the Earth Microbiome Project (EMP). The primer set (515F and 806R),

targeting Bacteria and Archaea, has found wide acceptance in part due to an appropriate amplicon length for sequencing on the Roche 454, Illumina MiSeq and Ion Torrent Personal Genome Machine (PGM) next-generation sequencing (NGS) platforms. The primer set is highly degenerate (2-fold degeneracy in the 515F primer pool, 18-fold degeneracy in the 806R primer pool). The new methodology was optimized using DNA from a constructed mock community. After optimization, complex environmental gDNA samples were analyzed using the new method and the results compared with those generated with standard PCR methods. The method, termed polymerase-exonuclease (PEX) PCR, significantly improved the evenness of recovery of templates in a mock simplified community, and had a significant effect on the observed structure of complex microbial communities.

Materials and Methods

DNA templates

Genomic DNA was extracted from fecal pellets of domesticated chinchillas ("Chin") using the Tissue DNA Purification Kit and Maxwell 16 System (Promega Corporation, Madison, WI). No specific permissions were required for the collection of the chinchilla feces. Genomic DNA was extracted from multiple sediment samples from Lake Huron (Lat/Long coordinates: 44°05.9933 N, 082°30.1474; 44°19.9650 N, 082°49.9548 W) using the PowerSoil DNA extraction kit (Mo Bio Laboratories, Carlsbad, CA), and pooled to make a complex gDNA pool ("Sed"). All lake sediment samples were collected on the RV Lake Guardian operated by the EPA, and no sampling permits were required. The field studies did not involve endangered or protected species. DNA quantity was measured using a Qubit 2.0 fluorometer with the dsDNA

BR Assay (Life Technologies, Grand Island, NY). Artificial double-stranded DNAs (gBlocks Gene Fragments) were synthesized by Integrated DNA Technologies, Inc. (IDT; Coralville, Iowa). The synthesized 492 bp gene fragments were derived from a portion of the SSU rRNA gene of *Rhodanobacter denitrificans* 2APBS1 [57, 58], covering the 515F and 806R primer positions with approximately 75 bp on either side of the targeted region. The sequence of the primer annealing sites was altered to introduce mismatches with primers or to match only a single primer from the degenerate primer pool. To allow for identification of each of the four gene fragments (*i.e.*, Mock A, B, C and D) after sequencing, a 10-bp region in the center of each of the fragments was scrambled to create a unique identifier sequence (**Table S1**). Other than alterations in the primer sites and the 10-bp region in the middle of the gene fragment, the sequences were identical. The exact primer site sequences and number of mismatches with each of the primers in the degenerate pool are shown in **Table S2**. The gene fragments were delivered as 200 ng stocks, and were dissolved in 20 microliters of TE buffer, yielding solutions with 10 ng/ul (~1.85 x 10^{10} copies/ul). A single equimolar pool of the four synthesized double-stranded gene fragments was made ("Mock"), and a 1/100th dilution of this pool was used for all subsequent analyses.

Standard PCR (Targeted amplicon sequencing, TAS)

The benchmark or standard PCR was a targeted-amplicon sequencing (TAS) approach, similar to that described by Bybee et al. [59, 60]. Briefly, genomic DNA is amplified using primers targeting the gene of interest, but containing 5' linker sequences which do not anneal to the genomic DNA template. Subsequently, the amplicons generated during the first stage of PCR are use as template for amplification with primers containing sequencing adapters, sample-

specific barcode sequences, and the same linker sequences, but located at the 3' end of the primer (TAS PCR method, **Figure 2**; **Figure S1**.). In this study, TAS sequencing was performed in two stages ("A" and "B") of 28 cycles and 8 cycles, respectively, generating amplicons ready for sequencing on an Ion Torrent Personal Genome Machine (PGM) sequencer or Illumina MiSeq. Unless specified, gDNA was PCR amplified with primers CS1_515F and CS2_806R (Primer set 2; **Table 2**). The primers contained 5' sequence tags (known as common sequence 1 and 2, CS1 and CS2) as described previously [61]. PCR amplifications were performed in 10 microliter reactions. A master mix for each reaction was made using the 2X AccuPrime SuperMix II (Life Technologies). The final concentration of CS1_515F and CS2_806R primers was 500 nM. Approximately 25 ng of environmental gDNA or 1 µl of the "mock" DNA was added to each reaction. Cycling conditions were as follows: 95°C for 5 minutes, followed by 28 cycles of 95°C for 30", variable annealing temperature for 45" and 68°C for 30". A final, 7 minute elongation step was performed at 68°C.

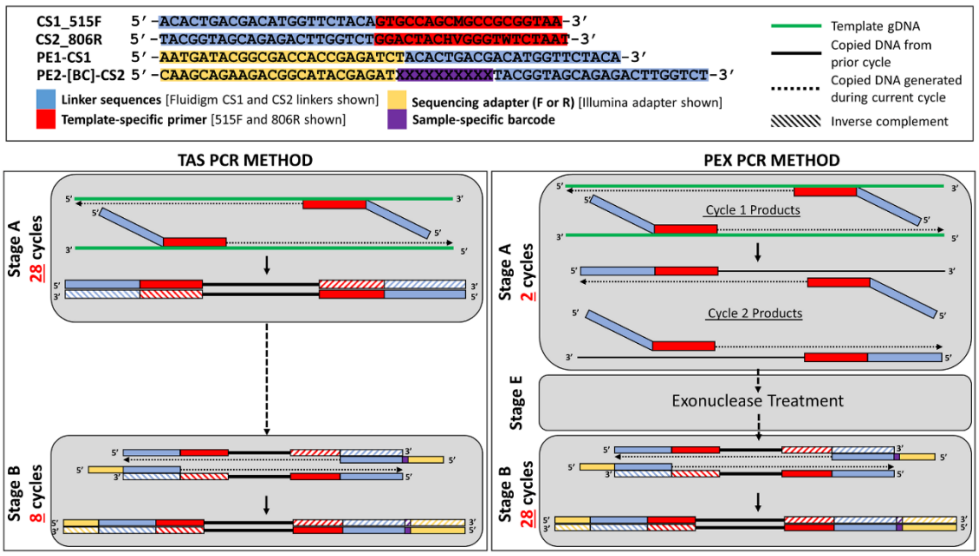


Figure 2: Schematic of Targeted amplicon sequencing (TAS) and Polymerase\ Exonuclease (PEX) PCR methods

Table 2: Primers used in this study

Primer Set	Primer Name	Sequences (5'-3')*	Reference
1	515F	GTGCCAGC**M**GCCGCGGTAA	Caporaso et al. [1]
	806R	GGACTAC**HV**GGGT**W**TCTAAT	Caporaso et al. [1]
2	CS1_515F	ACACTGACGACATGGTTCTACA-GTGCCAGC**M**GCCGCGGTAA	Caporaso et al. [1]; This study
	CS2_806R	TACGGTAGCAGAGACTTGGTCT-GGACTAC**HV**GGGT**W**TCTAAT	Caporaso et al. [1]; This study
3	CS1_515F	ACACTGACGACATGGTTCTACA-GTGCCAGC**M**GCCGCGGTAA	Caporaso et al. [1]; This study
	CS2_806R-NI	TACGGTAGCAGAGACTTGGTCT-GGACTAC**55**GGGT**W**TCTAAT	Caporaso et al. [1]; This study
4	Ion Torrent A Adapter - Barcode - CS2	CCATCTCATCCCTGCGTGTCTCCGACTCAG-**[BC]**-TACGGTAGCAGAGACTTGGTCT	Fluidigm
	Ion Torrent P1 Adapter - CS1	CCTCTCTATGGGCAGTCGGTGAT-ACACTGACGACATGGTTCTACA	Fluidigm
5	Illumina PE1 - CS1	AATGATACGGCGACCACCGAGATCT-ACACTGACGACATGGTTCTACA	Fluidigm
	Illumina PE2 - Barcode - CS2	CAAGCAGAAGACGGCATACGAGAT-**[BC]**-TACGGTAGCAGAGACTTGGTCT	Fluidigm

* Underlined sequences represent the common sequence linkers. M = A or C; H = A or C or T; V= A or C or G; W = A or T; 5 = 5-nitroindole substitution. **[BC]** = 10 base barcode that is unique to each sample.

A second PCR amplification (Stage B; **Figure 2**; **Table S1**), used to incorporate barcodes and sequencing adapters into the final PCR product, was performed in 10 microliter reactions, using the same master mix conditions as described above. Each well received a separate primer pair

containing a unique 10-base barcode, obtained from the Fluidigm Access Array Barcode Library for Ion Torrent or Illumina Sequencers (Primer set 4 or 5; **Table 2**). The final concentration of each primer was 400 nM. Cycling conditions were as follows: 95°C for 5 minutes, followed by 8 cycles of 95°C for 30", 60°C for 30" and 68°C for 30". A final, 7 minute elongation step was performed at 68°C.

Specialty amplification reactions (PEX PCR method)

An amplification strategy was developed to reduce amplification bias. This novel strategy, entitled PEX PCR (Polymerase-EXonuclease PCR), also included the two polymerase-mediated stages with the same primer sets as described above (PEX PCR method, **Figure 2**; **Table S1)**. Primer concentration during the first stage was reduced to 125 nM. The first stage (stage "A") consisted of only 2 cycles, and the second amplification reaction (Stage "B"; **Figure 2**) consisted of 28 cycles. Otherwise, reactions were set up identically as described above. Tubes containing master mix, gDNA or "mock" DNA, and CS1_515F and CS2_806R primers were heated to 95°C for five minutes and then to the specified annealing temperature (30°C, 35°C, 40°C, 45°C, 50°C, or 55°C) for 20 minutes. This cycle was repeated once more (two cycles total). In control reactions used to determine if the stage "A" primers were active during the stage "B" reactions (**Figure 2)**, only a single stage "A" cycle was performed.

After stage "A", samples were either directly treated with exonuclease I or diluted $1/10^{th}$ in sterile water (Mo Bio Laboratories), and treated with exonuclease I (Stage "E"). Some trial reactions did not receive exonuclease treatment. Exonuclease digestion was performed with five microliters of diluted or undiluted sample from the first two-cycle stage of the PEX method. Five

µl of sample were mixed with two µl of ExoSAP (ExoSAP-IT for PCR Product, Affymetrix, Santa Clara, CA) and incubated according to the manufacturer's instructions (37°C for 15', 80°C for 15'). Three microliters of the ExoSAP-treated sample was transferred to the second PCR reaction containing primer sets 4 or 5 (**Table 2**), as described above. Each sample and replicate received a primer set with a unique barcode. PCR was performed as described above, but with 28 cycles instead of 8. In tests without exonuclease treatment, 1 µl of sample from the first two-cycle stage of PEX PCR method was transferred to the second 28-cycle stage of the reaction.

In addition, some reactions were performed with a modified 806R primer ("806R-NI"; Primer set 3, **Table 2**). The most highly degenerate positions of the 806R primer (two adjacent positions are 3-fold degenerate, "H" and "V", **Table 2**) were replaced with so-called "universal base" 5-nitroindole substitutions [62]. All primers were synthesized by IDT. PCR conditions were not altered when 806R-NI primers were utilized.

Next-generation amplicon sequencing

Final PCR amplicons from TAS and PEX PCR methods were pooled in equal volume and purified using solid phase reversible immobilization (SPRI) cleanup, using AMPure XP beads (Beckman Coulter, Brea CA) at a ratio of 0.6X (v:v) SPRI solution to sample. Final quality control was performed using the D1000 ScreenTape assay implemented on a TapeStation2200 instrument (Agilent Technologies, Santa Clara, CA) and Qubit analysis. The pooled libraries were quantitated by qPCR using a library quantification kit (KAPA Biosystems, Wilmington, MA). For Illumina sequencing, the library pool was spiked with 15% non-indexed PhiX control library provided by Illumina and then loaded onto a MiSeq v2 flow cell at a

concentration of 8 pM for cluster formation and sequencing (paired-end reads, 2x250 bases). Custom sequencing and index read primers (according to Fluidigm Access Array Illumina sequencing guide) were added to the appropriate wells of the reagent cartridge. Illumina MiSeq sequencing was performed at the W.M. Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign, and data were analyzed using the Casava1.8 pipeline. For Ion Torrent sequencing, pooled libraries were diluted to 8.5 pM for emulsion PCR. Libraries were prepared for sequencing using automated emulsion PCR employing the Ion PGM template OT2 400 kit. Sequencing was performed on an Ion Torrent Personal Genome Machine (PGM) with a 318 chip, using the Ion PGM sequencing 400 kit (Life Technologies). Barcode sequences from Fluidigm were provided to the PGM server, and sequences were automatically binned according to 10-base multiplex identifier (MID) sequences. Raw reads were recovered from the PGM server as FASTQ files. Ion Torrent PGM sequencing was performed at the DNA Services Facility at the University of Illinois at Chicago.

Analysis of mock community amplicon sequence data

Under ideal conditions, PCR amplification of the mock community would result in equal amplification of the four templates, resulting in 25% of all sequence reads attributed to each of the four target templates. To assess the effectiveness of each modification to the amplification reactions, FASTQ files (Ion Torrent and Illumina) were imported into the software package CLC genomics workbench 7.0 (CLC Bio, Qiagen, Boston, MA). Raw sequence data were trimmed using quality trimming algorithms (quality threshold, 0.05 for Ion Torrent and 0.01 for Illumina), and common sequences (Ion Torrent only). Sequence data from the “mock” communities were

mapped against the four variants of *R. denitrificans* 2APBS1 gene (**Table S1**) within the software package CLC Genomics workbench. Counts for each variant were generated.

Divergence from equal relative abundance was calculated by summing the difference between expected abundance (*e.g.*, 25% for each variant) and measured abundance for each variant ("Ideal score (IS)"; $\sum_{i=1}^{n} abs(1/n - Pi) * 100$ where n= number of targets in equimolar pool and P*i*= percentage of NGS sequencing reads mapping to target i). Ideal scores were calculated for all four possible templates (IS4), for three target templates (excluding the two-mismatch template; IS3), and for the two no-mismatch templates (IS2). Lower scores represent a closer representation of the ideal (for evenly distributed templates, potential scores range from 0 to 2*(100/n)*(n-1)).

Analysis of environmental community amplicon sequence data

For analysis of amplicons generated from environmental gDNA, sequences were initially processed in CLC genomics workbench. Illumina reads were merged, and subsequently quality-trimmed to remove low-quality reads. Ion Torrent data were quality trimmed only, and reads of less than 200 bases were removed from the analysis. For analysis of microbial community structure, primer sequences (515F and 806R) were removed. The remaining sequences were exported as FASTA files and processed with the software package QIIME [1]. Briefly, sequences were screened for chimeras using the usearch61 algorithm (Edgar 2010), and putative chimeric sequences were removed from the dataset. Subsequently, each sample sequence set was sub-sampled to the smallest sample size to avoid analytical issues associated with variable

library size [63]. Sub-sampled data were pooled and renamed, and clustered into operational taxonomic units (OTU) at 97% similarity. Representative sequences from each OTU were extracted, and these sequences were classified using the "assign_taxonomy" algorithm implementing the RDP classifier, with the Greengenes reference OTU build [64, 65]. A biological observation matrix (BIOM; [77]) was generated at taxonomic levels from phylum to genus using the **"**make_OTU_table**"** algorithm. The BIOMs were imported into the software package Primer6 [66] for analysis and visualization. Figures were generated using the software package OriginPro8.5 (OriginLab Corporation, Northampton, MA) and in the software packages Excel and Powerpoint (Microsoft, Redmond, Washington).

Analysis of primer utilization patterns

In standard bioinformatics analyses, sequence data at primer sites are removed prior to bioinformatics analyses. To examine which primers annealed to template strands in TAS and PEX PCR reactions, quality-trimmed sequences containing primer sequences were exported from CLC genomics as FASTA files, or mapped to mock community reference sequences, and subsequently exported. Sequences were imported into Excel, and searching algorithms were implemented to detect each primer variant from the degenerate primer pools. For reactions in which the 806R primer contained 5-nitroindole substitutions (*i.e.* primer set 3, **Table 2**), an additional 14 potential variants for the 806R primer were examined, since any base can potentially be incorporated opposite to a 5-nitroindole substitution.

Data Access

The gene amplicon sequence data generated as part of this study have been submitted to the NCBI BioProject database under accession number PRJNA262579. Sample details and FASTQ file names are provided in **Table S3**.

Results

Theory

During the polymerase chain reaction, two distinct types of primer-template interactions can occur, and these types of interactions likely operate at different efficiencies. The first primer-template interaction is that of the PCR primers with the source genomic DNA; such interactions include both perfect matches and primer annealing to regions containing mismatches of unpredictable number and location. Subsequently, however, PCR primers interact with primer sites that have been created during the PCR cycle, and are the inverse complement of the synthesized oligonucleotides used as PCR primers. Here, the primer-template interactions can also include perfect matches and mismatches, but the scale of mismatch annealing is proportional to the number of degeneracies in the primer pool. We have termed primer annealing to genomic DNA as "natural" interactions and primer annealing to PCR amplicons as "artificial" interactions (**Figure 3A**).

A critical observation of these two type of primer-template interactions is that there are two mechanisms for generating the final fragments generated by PCR (*i.e.*, "E" and "F" fragments). This includes natural template-primer annealing processes (*i.e.* "E" and "F" fragments from "C" and "D" fragments) and artificial template-primer annealing processes (*i.e.*

"E" and "F" fragments from other "F" and "E" fragments) (**Figure 3A**). Artificial template-primer interactions occur only in cycles three and above, since no "E" and "F" molecules exist in the reaction until the end of cycle two. Regardless of the degeneracy of the primer pool, natural template-primer interactions are likely to be more complex due to the great potential for the presence of primer-template mismatches as a result of the underlying genetic diversity at primer sites. Artificial template-primer interactions are less complex, and the number of potential mismatches between primer and template is proportional to the diversity of the primer pool employed.

Artificial template-primer interactions ultimately dominate PCR as artificial template-primer interactions yield exponential amplification, while natural template-primer annealing interactions yield linear amplification. During the early cycles of PCR (*i.e.*, up to the 12th cycle), however, natural template-primer annealing interactions can contribute substantially (>1%) to the overall pool of fragments that are the final yield of PCR (**Figure 3B**), assuming both processes operate at 100% efficiency.

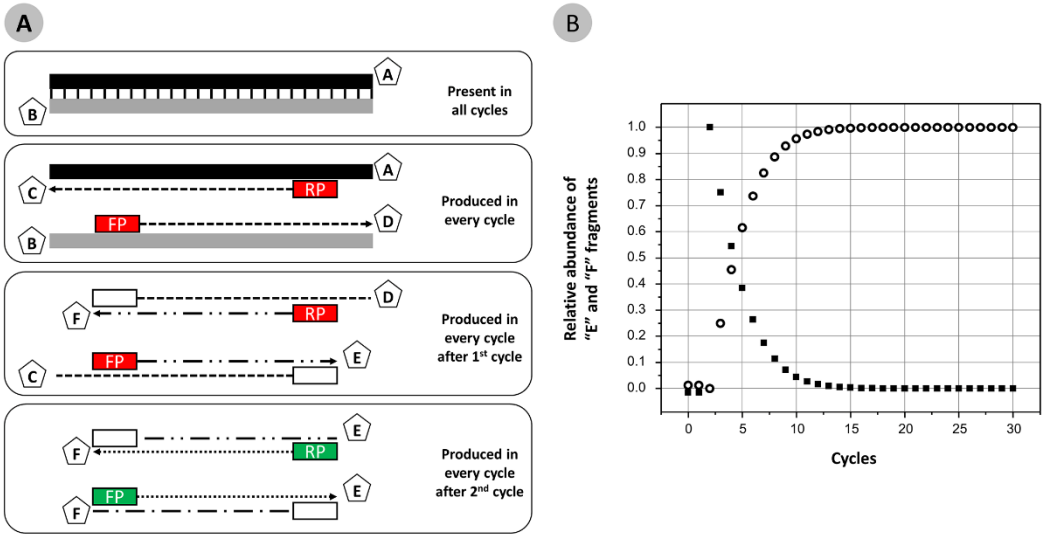


Figure 3: Types and abundance of DNA fragments found in PCR

(A) Template DNA fragments (containing strands "A" and "B") are added to PCR reactions and are conserved throughout the reaction. Fragments "A" and "B" serve as templates for copying in each cycle, with hybrid molecules "C" and "D" produced in a linear fashion each cycle. In cycles two and above, "C" and "D" are copied, creating hybrid molecules "E" and "F" in a linear fashion. In cycles three and above, the "E" and "F" fragments generated in prior cycles are copied into inverse complement fragments "F" and "E", respectively, in an exponential fashion. Red boxes indicate 'natural' primer annealing to genomic DNA template or copy of gDNA template. Green boxes indicate 'artificial' primer annealing to primer sites that are copies of oligonucleotide primers added to the PCR mixture, and incorporated during previous cycles. **(B)** The relative abundance of "E" and "F" fragments generated by 'natural' template-primer interactions ("C","D" → "E","F"; shown as solid squares) and by artificial template-primer interactions ("E","F" → "F","E"; shown as open circles) varies by cycle. At the end of cycle two, all "E" and "F" fragments have been generated only by 'natural' template-primer interactions.

At the end of cycle 2, however, no exponential amplification can have occurred. In fact, there has (providing that the reaction operates at 100% efficiency) been a non-destructive conversion of gene targets present in large gDNA fragments to short gene fragments bounded at either end by the forward and reverse primer sites. These fragments are not true double-strands, but hybrids of one long strand with only one primer site, and a short strand with both primer sites present (Third box, **Figure 3A**). In theory, for every gDNA copy of double-stranded DNA there

exist two single stranded fragments of the same region. Thus, the end of the second cycle of PCR marks a critical transition from linear copying only (cycles 1 and 2) to linear and exponential amplification (cycles 3 and above).

In this study, 'natural' template-primer interactions were separated from 'artificial' template-primer interactions. To achieve this separation, PCR primers with 5' linker sequences which do not interact with the native genomic template DNA were synthesized (*e.g.*, Primer Set 2, **Table 2**; **Figure 2**). These linkers are intended to have no similarity to any known biological sequences. Thus, at the end of cycle two, E and F fragments are bounded by the two separate linkers (a "forward" linker and a "reverse" linker). After the second cycle is complete, stage "A"of the overall reaction is terminated, and the original primers are removed using DNA exonuclease I digestion. The gDNA (A,B fragments), first and second cycle copies (C,D fragments) and second-cycle only copies (E,F fragments) are then transferred to a second PCR in which new primers, targeting the linker sequences, are used instead of template-specific sequences (**Figure 2**, **Figure S1**). Only the "E" and "F" fragments are suitable templates for such amplification. The purpose of this is to perform the minimum number of cycles with degenerate primers operating under 'natural' template-primer interaction conditions and rapidly shift the exponential amplification of amplicons to 'artificial' template-primer interactions employing non-degenerate primers. Thus, the template-specific primers are not involved in amplification. In this way, bias associated with degenerate primers and with template-primer mismatches may be reduced.

Application of a new pipeline for PCR amplification of templates

To separate 'natural' and 'artificial' template-primer interactions, a simple workflow, customizable to any primer set, was developed **(Figure 2**, **Figure S1**). After two cycles of denaturation, annealing and extension, the reaction was terminated (Stage "A"). Subsequently, the sample incubated with an Exonuclease I to digest single stranded DNA (*i.e.*, unincorporated primers). The lowered initial primer concentrations (125 nM) were used since the exonuclease activity was insufficient to remove primers from reactions with higher concentration. In some cases, stage 1 cycling products were diluted $1/10^{th}$ prior to exonuclease digestion. Exonuclease activity in the Exo-SAP product is optimized for post-PCR conditions in which much of the primer pool is consumed in the amplification of the template. Since no amplification occurs in the first two cycles of the PEX PCR method, only a small amount of primer is consumed in the reaction. We initially observed this by testing for second stage amplification after only one cycle instead of two. We anticipated that after one cycle only, the second stage amplification could not occur as no template molecules would contain both the forward and reverse common sequences (*i.e.*, no E and F fragments would be produced before cycle 2). When amplification was observed after only 1 cycle, this was taken to indicate that left-over primer from the first stage of the reaction was still active during the second stage of the reaction, generating additional copies during second stage cycling. This was subsequently verified when examining primer utilization patterns (see below). Thus, conditions were optimized until stage "B" PCR produced no amplicons when only one cycle of stage "A" reaction was performed. For the primer set CS1_515F and CS2_806R, these conditions could be met by lowering the initial primer concentration to 125 nM (since most of the primer is not used in the first two cycles), and subsequently diluting the reaction $1/10^{th}$ before Exo-SAP treatment. Although manufacturer's

details (*i.e.* Exo-SAP can degrade 5 μM of primer) suggest that this dilution is not necessary, we often observed stage "B" (**Figure 2**) amplification without dilution. Finally, exonuclease I is a 3'-5' processive enzyme, and at the 3' end of the hybrid molecules generated after the second cycle, the DNA is double-stranded and therefore not a target for the exonuclease. After treatment of the diluted first stage sample with Exo-SAP, 3 microliters of treated sample was transferred to a second PCR reaction, as was performed for the TAS (standard) PCR method. Here, the sequencing adapter and sample-specific unique barcode were incorporated using PCR. In the TAS approach, only 8 cycles were used to incorporate sequencing adapters, but for the PEX PCR method, 28 amplification cycles were performed since this reaction is intended to amplify targets and incorporate sequencing adapters.

Reactions in which the standard 806R primer was replaced with the 806R primer containing 5-nitroindole substitutions were not found to amplify well under standard conditions (TAS PCR), and weak or no yield was generated. However, the nitroindole substituted primer did work effectively in reactions employing the PEX PCR method protocol, and this may be related to relatively weak template-primer interactions associated with the "universal base" substitutions.

Determination of optimum PEX PCR method annealing temperature

To determine the optimum temperature for the PEX PCR protocol with primer set 2 (**Table 2**), a temperature gradient was performed with stage "A" reactions containing the mock community DNA. At each temperature (30°C-55°C, 5°C increments), two cycles of 20 minute annealing and elongation were performed. After exonuclease treatment (stage "E"), each pool of

fragments was amplified with the stage "B" primers, containing a unique barcode. Each condition was performed in quadruplicate, and a minimum of 8,493 sequences per sample were generated (Ion Torrent PGM). The distribution of reads among the four templates at each temperature is shown in **Figure 4A**. Ideal scores were calculated for all four templates (IS4), for all templates excepting the two mismatch template (IS3) and for only perfectly matching templates (IS2), and compared to results from TAS PCR (below) (**S4 File**). Samples processed with the PEX PCR method had significantly lower IS scores for all temperatures when considering 3 or 4 templates (IS3 and IS4; **S4 File**). When considering only the two perfect match templates (Mock A and Mock B), IS2 scores were significantly lower for all PEX PCR reaction temperatures except 30°C and 40°C (IS2; **S4 File**). Overall, the two mismatch template (Mock D) was not well amplified at any temperature, although slightly higher amplification (still below 1% of total reads) was observed in PEX PCR reactions with 40°C and 45°C annealing temperatures. Reactions using 45°C and 50°C reaction temperatures during the first stage of the PEX PCR method were capable of nearly even amplification all templates except Mock D (**Figure 4A**; **S4 File**).

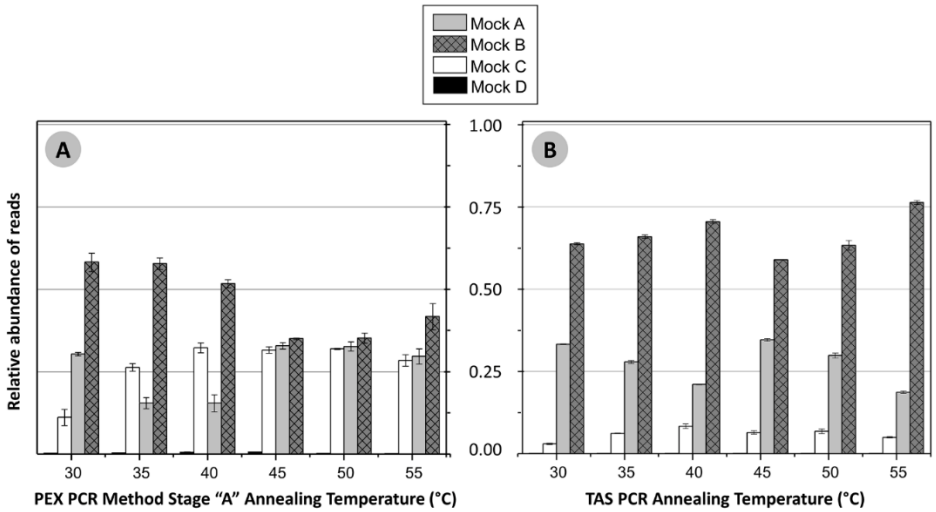


Figure 4: Temperature gradient analysis of the PEX PCR and TAS methods using mock community DNA

The relative abundance of reads mapping to the each of the four target templates (Mock A, Mock B, Mock C and Mock D) is shown for each temperature. The error bars represent standard deviation associated with two to four replicates per sample. (A) Results from PEX PCR and (B) Results from TAS PCR.

For comparison with TAS (standard) PCR, a modified stage 1 PCR amplification was performed on mock community DNA to best approximate cycling conditions in the PEX PCR method. Briefly, a two-step PCR cycle was used, and a single annealing temperature was held for two minutes (*i.e.*, 95°C – 30", AT – 2'; repeated 28 times). The distribution of reads (minimum of 5,902 sequences per replicate) among the four templates at each temperature is shown in **Figure 4B** and **S4 File**. The TAS protocol preferentially amplified the template perfectly matching the reverse primer with the highest Tm (Mock B), and this varied little with annealing temperature. TAS PCR poorly amplified templates with introduced mismatches (*i.e.* Mock C and D).

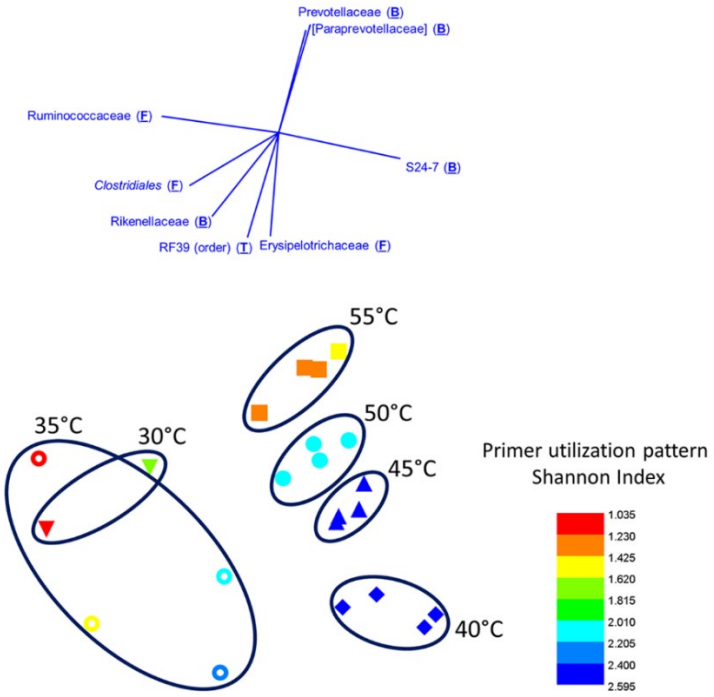


Figure 5: Effect of PEX PCR Stage "A" annealing temperature on observed microbial community structure and primer utilization patterns

Non-metric multidimensional scaling (NMDS) plot of fecal microbiome, performed at the taxonomic level of family and based on Bray-Curtis similarity (2D stress = 0.05). Samples were rarefied to 1,250 sequences per sample and no transformation was applied. The analysis is based on a single genomic DNA sample, with PEX PCR stage "A" annealing performed at 30°C (down-facing triangles), 35°C (open circles), 40°C (diamonds), 45°C (up-facing triangles), 50°C (closed circles) and 55°C (squares). Symbols are color-coded by the diversity (Shannon Index) of reverse primers (i.e. 806R) utilized in annealing and elongation during stage "A" of PEX PCR. Maximum possible Shannon index for 18 primers in the primer pool is 2.89. Vectors indicate taxa with Pearson correlation of >0.8 with MDS1 and MDS2 axes.

A temperature gradient was also performed using the PEX PCR method, with a single gDNA sample extracted from mammalian feces ("Chin") used as the template (**Figure 5**). Data were processed as described above, and analyses were performed at the taxonomic level of family. Poor amplification of environmental gDNA was found with PEX PCR method stage A annealing temperatures below 40°C, and replicates were more variable than at higher temperatures. The observed microbial community was similar at all temperatures above 40°C (>80% Bray-Curtis similarity), with temperatures 45°C-55°C most similar (>90% Bray-Curtis similarity). Nonetheless, the observed microbial community was significantly different when the "Chin" sample was processed using different PEX PCR method stage "A" annealing temperatures (ANOSIM, $R=0.746$, p-value<0.0002) due to annealing temperature-associated shifts in the relative abundance of individual taxa. For example, the relative abundance of sequences from bacteria of the family Prevotellaceae was correlated with stage "A" annealing temperature (**Figure 5**). Based on results from the mock community, and analysis of "Chin" gDNA, a reaction temperature of 45°C was chosen for the first stage of the PEX PCR method in subsequent analyses using the CS1_515F and CS2_806R primer set.

Analysis of a mock community using standard and PEX PCR method protocols

A series of tests with mock community DNA were performed to determine if the PEX PCR method protocol improved the evenness of recovery of target templates with varying sequences at primer sites. Mock community templates were amplified at 45°C and 55°C annealing temperatures using TAS PCR, at 45°C using the PEX PCR method with and without exonuclease treatment, and at 45°C using the PEX PCR method without exonuclease, but with the 806R-NI primers. Sequencing for these amplicons was performed using an Illumina MiSeq.

The results of these analyses, similar to that of prior sequencing on the Ion Torrent PGM, revealed that TAS PCR, at 55°C annealing temperature, grossly distorted the underlying ratio of templates in the mock community (**Table 3**, Test 1; **Figure 6**). By lowering the annealing temperature to 45°C, the reaction could more evenly amplify perfectly matching templates (*i.e.* Mock A and Mock B), but still poorly amplified both DNA fragments containing mismatches introduced at the 3' end of the priming sites.

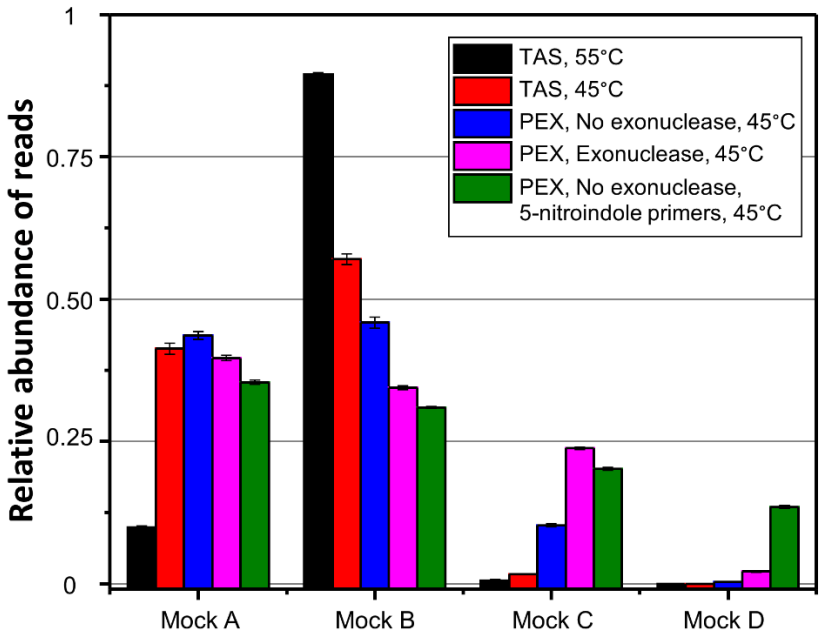


Figure 6: Relative abundance of mock DNA templates observed in sequencing of TAS and PEX PCR method reactions

The error bars represent standard deviation associated with two replicates per sample.

Table 3: Amplification and sequence analyses of mock community DNA

	Conditions Employed				Average percentage of reads mapping to references				Ideal Scores (IS)		
Test	**Method#**	**AT#**	**Exo#**	**Reverse Primer**	**Mock A**	**Mock B**	**Mock C**	**Mock D**	**4 targets**	**3 targets**	**2 targets**
1	TAS	45°C	No	806R	41.27	57.03	1.70	0.00	97*	63*	16*
	TAS	55°C	No	806R	9.91	89.47	0.61	0.01	129	112	80
2	TAS	45°C	No	806R	41.27	57.03	1.70	0.00	97	63	16
	PEX	45°C	No	806R	43.60	45.86	10.27	0.27	79*	46*	3*
3	PEX	45°C	No	806R	43.60	45.86	10.27	0.27	79	46	3^
	PEX	45°C	Yes	806R	39.66	34.40	23.79	2.16	48*	18*	7
4	PEX	45°C	Yes	806R	39.66	34.40	23.79	2.16	48	18*	7
	PEX	45°C	No	806R-NI	35.39	30.96	20.19	13.46	33*	20	7^{NS}

TAS PCR = Targeted amplicon sequencing (standard PCR approach); PEX = Polymerase exonuclease PCR method; Exo = Exonuclease treatment

* Significant decrease relative to alternate method; $p<0.01$, two-tailed TTEST (unequal variance)

^ Significant decrease relative to alternate method; $p<0.05$, two-tailed TTEST (unequal variance)

NS = Not significant, two-tailed TTEST (unequal variance)

DNA fragments containing a single introduced mismatch at the forward primer site (Mock C) were amplified more effectively using the PEX PCR method, even without exonuclease (**Table 3**, Test 2). However, this fragment was most effectively targeted when the exonuclease step was included in the PEX PCR reaction (**Table 3**, Test 3). The two mismatch fragment (Mock D) was

not amplified effectively even in the PEX PCR reaction with exonuclease, and less than 3% of the reads were derived from the two mismatch fragment (Mock D). The use of 5-nitroindole substitutions in the 806R primer pool, decreasing the level of degeneracy of the reverse primer pool from 18X to 2X, yielded improved amplification of the two mismatch template (Mock D) and slightly higher than 13% of the reads were derived from this template (**Table 3**, Test 4). Overall, the "ideal score", an estimate of how close to the true ratio of the input template the observed distribution is, demonstrated substantial and significant improvement with (i) a decrease in annealing temperature to 45°C, (ii) the use of the PEX PCR method, and (iii) substitution of 5-nitroindole for 3-fold degenerate positions in the 806R primer. Removal of exonuclease treatment decreased the evenness of the amplification reaction.

Analysis of environmental DNA using standard and PEX PCR method protocols

Genomic DNA extracted from sediment from Lake Huron was recovered and analyzed using the TAS PCR approach, PEX PCR method and PEX PCR method without exonuclease (**Figure 7**). In addition, the effect of primers containing 5-nitroindole substitutions was also examined (**Figure S2**). All reactions were performed with an annealing temperature of 45°C and sequencing was performed on the Illumina MiSeq instrument. Bacterial SSU rRNA amplicon sequences were clustered and analyzed at the taxonomic level of family. A significant effect of method (TAS vs PEX PCR method) was observed (ANOSIM, $R=0.778$, $p<0.02$), as well as a significant effect of reverse primer pool (806R vs 806R-NI; $R=0.763$, $p<0.0003$). The observed sediment sample community differed when processed using the PEX PCR method with or without exonuclease (across both regular and nitroindole primers), though analysis of similarity was not significant ($R=0.178$; $p<0.07$). Calculation of Shannon indices for the "Sed" sample

indicated that the family diversity observed in PEX PCR amplification was slightly, but significantly, higher than that observed for TAS amplification (**Figure 7**). Replicates of lake sediment microbial communities in analyses with the 806R-NI primer showed poorer amplification and much greater variation than replicates from the same sample with the standard 806R primer pool (**Figure S2**).

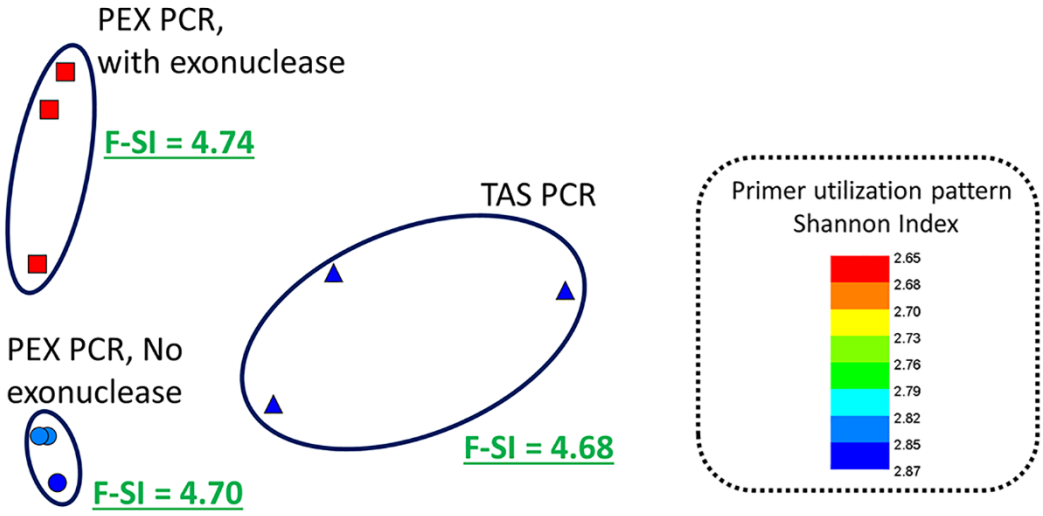


Figure 7: Effect of PEX PCR and exonuclease treatment on observed microbial community structure and primer utilization patterns

Non-metric multidimensional scaling (NMDS) plot of lake sediment microbiome, performed at the taxonomic level of family and based on Bray-Curtis similarity (2D stress = 0.02). Samples were rarefied to 35,500 sequences per sample and no transformation was applied. All reactions were performed with an annealing temperature of 45°C, using PEX PCR with exonuclease (squares), PEX PCR without exonuclease (circles), and TAS PCR (triangles). Symbols are color-coded by the diversity (Shannon Index) of reverse primers (i.e. 806R) detected in the sequences. Maximum possible Shannon index for 18 primers in the primer pool is 2.89. Small, but

significant differences in the observed family-level Shannon index (F-SI) were observed between PEX PCRs (with and without exonuclease treatment) and TAS PCRs using a two-tailed ttest ($p<0.05$).

Analysis of primer utilization patterns

The PEX PCR method allows for direct examination of which primers from a primer pool are involved in annealing and polymerase extension during the first cycles of amplification reactions, when 'natural' primer-template interactions occur. In standard PCRs, the primer sites are not informative, because during many cycles of amplification, different primers from a primer pool amplify target molecules during 'artificial' primer-template amplification cycles (*e.g.* cycles 3 and above; **Figure 3B**). In the PEX PCR method, the CS1 and CS2 portions of the primer are used for exponential amplification, and the sequence of the initial primer from the primer pool annealing to the genomic DNA template is preserved during subsequent amplification cycles in Stage "B" (**Figure 2**). The primer utilization pattern was initially examined in the analysis of mock community DNA across a temperature gradient (**Figure 5**; **Figure S3**). All samples amplified with the TAS method had highly similar primer utilization profiles (>90% Bray-Curtis similarity), and these profiles showed that all primers within the reverse primer pool were involved in amplification (**Figure S3**). This is an indication of the signal "scrambling" that occurs in standard PCR; when 'artificial' primer-template interactions dominate, primers with mismatches to amplicons can still readily anneal to these targets. A strong effect of annealing temperature on primer utilization pattern in PEX PCR-amplified samples was observed. At the highest annealing temperature (55°C), a strong skewing towards the primer with the lowest Tm was observed (**Figure S3**). In addition, at this 55°C annealing

temperature, the primer with the highest Tm was also a large contributor to annealing and extension. Since the mock community is composed of two templates perfectly matching the lowest Tm primer, one template with a single mismatch with the lowest Tm primer, and one template with no mismatches with the highest Tm primer (**Table S2**), this indicates that at the highest annealing temperatures, perfect match or few mismatch interactions are heavily favored. When the annealing temperature in the PEX PCR method was dropped to 45°C, a more even distribution of primer utilization was observed, and the primer with the lowest Tm was less strongly favored. To rapidly analyze primer utilization patterns, a Shannon Index (SI) was calculated for the distribution of reverse primers (18 possible combinations; maximum SI=2.89). Between annealing temperatures of 55° to 40°C, a dramatic increase in the diversity of primers used for amplification was observed (**Figure 5**); below 40°C, an inconsistent signal was obtained (**Figure 5**).

To examine this phenomenon more closely, the primer utilization patterns were examined for each of the four separate templates within the mock community, using both the TAS and PEX PCR methods (**Figure S4**). For TAS amplification, annealing temperature did not substantially alter primer utilization patterns, though each target had different patterns (too few sequences were recovered for template Mock D for analysis). For PEX PCR amplification, both temperature and template type altered primer utilization patterns. At high annealing temperatures, the lowest Tm primer was predominantly associated with amplification of all targets except Mock B, although the highest Tm primer contributed substantially. For the Mock B templates, the highest Tm primers were the most strongly associated with annealing and elongation. When using the PEX PCR method at 45°C, a broader array of primers were involved in annealing and elongation of Mock A, Mock C, and Mock D targets. A wide array of primers

were also involved in annealing and elongation of the Mock B template, but here the primer with the highest Tm (perfectly matching the reverse primer site) was abundantly utilized.

When examining the primer utilization patterns for the "Sed" sample (lake sediment), a very diverse utilization pattern was observed for the PEX PCR method (**Figure 7**). This finding is consistent with the high microbial diversity in the "Sed" sample, even when compared to the fecal microbiome sample. The TAS method showed the most diverse pattern, consistent with the hypothesis that standard PCR "scrambles" the signal of primer utilization. When the same sample was processed using the PEX PCR method, but without exonuclease treatment, the resulting primer utilization pattern was most similar to that of the sample processed with the PEX PCR method, but a more even distribution of primer utilization (*i.e.* higher primer Shannon index) was observed, similar to the TAS PCR sample (**Figure 7**). This observation is an indication that carryover of primers from stage "A" to the stage "B" of the PEX PCR method is responsible for distortion during PCR amplification. Samples processed with the primer pool containing 5-nitroindole substitutions had a strong selection for a single primer from the pool (**Figure S2**). This represents a preference for specific bases when the polymerase uses a 5-nitroindole base as a template for copying. In reactions with primers containing 5-nitroindole substitutions, recovery of an additional 14 primer variants was observed (**Figure S2**), though at a very low level (<0.1%).

Discussion

We describe here a novel technique for amplifying DNA to reduce the negative effects of mismatches between primer and template on the efficiency of amplification of target templates.

This method is an extremely general and simple modification to the PCR reaction, and involves three enzymatic steps, including a polymerase-mediated linear copying of genomic DNA templates, an exonuclease digestion of primers, and a final PCR amplification using primers targeting target-independent linker sequences. As shown in this manuscript, this approach allows the exponential amplification of the target of interest to be performed with primers that have no mismatches with any templates in the reaction. This significantly reduces bias associated with mismatches and degenerate primers that can accumulate during PCR by limiting primer-template interactions to two cycles only.

The objectives of this study were to: (i) demonstrate that the first and second stages of PCR could be separated, and still generate reliable amplification; (ii) determine if degeneracies in the primer pools led to obvious distortion of the observed microbial community, and if the PEX PCR technique could be used to circumvent or reduce such distortion, (iii) develop a robust workflow for this approach to be implemented for any primer set or sets, and (iv) identify applications to which this approach is best suited. The results herein demonstrate that indeed the two defined types of interactions within PCR (*i.e.* natural and artificial interactions) can and should be separated when degenerate primers are used or when mismatches with the template are anticipated. The stages should be separated because genomic DNA template-primer interactions have the greatest potential for bias due to mismatches derived from true mismatches in the gDNA and from degeneracies in the primer pool. Our analysis of an artificially synthesized mock community demonstrates the strong potential for a degenerate primer pool of oligonucleotides of varying melting temperatures to preferentially select templates based on sequence variations in the primer site. Our strategy limits the gDNA template-primer interaction to two cycles, with all subsequent amplification cycles employing non-degenerate, non-template

interactions. Furthermore, because only two cycles are utilized during the first stage of gDNA-primer interactions, unusual annealing and elongation conditions can be utilized. In this study, we employed long annealing times (20 minutes) at low annealing temperatures to allow for adequate time for the polymerase to bind target sites and elongate without raising the reaction temperature. Such reaction conditions are likely more tolerant of low primer annealing efficiency of primer-template pairings with mismatches. In a systematic study of primer-template interactions, Wu et al. [67] reported that single mismatches occurring in the last 3-4 positions from the 3' end of the primer yielded minimal or no primer extension. We observe here that 3' mismatches can be overcome using the PEX PCR method, and that primers with 3 or 4 mismatches can still anneal with genomic DNA targets and yield polymerase extension. This was revealed through an analysis of primer utilization patterns in mock community analyses. We note that such an analysis cannot be performed using standard PCR approaches, and requires the PEX PCR method.

Furthermore, we demonstrate through the comparison of PEX PCR method with and without exonuclease that the primers targeting the source gDNA (or mock DNA) do contribute to the distortion during later cycles of PCR even in the presence of high concentrations of the second stage PCR primers. The complete removal of unincorporated primers from the first stage reaction, although desirable, was found to be difficult. Even after dilution, exonuclease digestion, and lowering of the primer concentration during the first two cycles, some limited amount of first stage primer may be propagated to the second stage PCR. Despite this, treatment with exonuclease significantly reduces the impact of primer carry-over, and is an essential part of the PEX PCR methodology. In analyses of both mock community and environmental gDNA, exonuclease treatment significantly alters the observed microbial community. It is possible that

in place of exonuclease treatment, blocking oligonucleotides of the inverse complement of the forward and reverse template-specific primers (*e.g.* inverse complement of 515F and 806R primers without CS1 and CS2 linkers) could be added to the second PCR stage of the PEX PCR reaction to prevent gDNA template-primer interactions.

The PEX PCR method also provides a novel and robust mechanism to explore primer-template interactions in analyses of complex gDNA samples. The PEX PCR method preserves the sequence of the primer annealing to the gDNA template during the first two cycles of the reaction, and these can be bioinformatically interrogated to determine which primers within a degenerate pool are truly involved in annealing and extension. We observed that at high annealing temperatures, perfect match annealing is favored, and a lower diversity of the primers in the pool were utilized. This appears to be detrimental for the amplification reaction, as perfectly matching primers are present at a low overall abundance in a heavily degenerate primer pool. At lower annealing temperatures, a broader spectrum of primers, containing mismatches with the template, are involved in annealing and elongation. This appears to be beneficial, as analyses of the mock community under lower annealing conditions, generated better representations of the true underlying distribution of mock DNAs. Primers with 0-4 mismatches with various templates were observed to anneal and allow for polymerase extension. When 3' mismatches were introduced into mock DNA templates, the primer utilization distribution shifted towards primers with 1 or 2 total mismatches to the template. Therefore, the heavy degeneracy of primer sets may not be beneficial when using the PEX PCR method. Instead, it may be appropriate to select "intermediate" primers that have at most 1 or 2 mismatches to all potential priming sites, with the assumption that every variant does not require a unique primer to be targeted. Further research is needed to determine the best combination of primer degeneracy and

annealing temperatures for other, more degenerate targets such as microbial functional genes. Such a strategy may enable a direct PCR-based method for analysis of a single copy gene present in all microorganisms for the purpose of community structure analyses. We further note that additional strategies may be required to allow the PEX PCR method to work effectively at lower annealing temperatures (*e.g.* <45°C), such as the introduction of single-stranded DNA binding protein into the amplification master mixes.

The PEX PCR method is recommended for: (i) any PCR reaction in which a degenerate primer pool is used; (ii) any PCR reaction in which a non-degenerate primer is used but where DNA template variability at the priming site is possible; (iii) reactions in which high-level degeneracy may be utilized to target all known variants of a gene; and (iv) when multiple primer pairs are to be utilized simultaneously. We show here that the method can be used to amplify and sequence templates with mismatches at the 3' end of the primer site, which have been shown to be highly destabilizing in PCR [67, 68]. Crosby and Criddle [69] previously employed a strategy using hybridization capture followed by random-primed amplification and sequencing to target DNA-directed RNA polymerase genes (*rpoC*). The PEX PCR method may be adaptable to a direct PCR amplification of *rpoC* genes from all microorganisms within a single amplification reaction. This would preserve the original relative abundance found in the template DNA, and provide a direct proxy for relative abundance of organisms in the sample. This is unlike amplification and sequencing of rRNA genes, as performed in this study, since a wide range of gene copies of rRNA operons are found across the domains Bacteria and Archaea [6]. We do not yet know if the level of degeneracy at conserved regions of the *rpoC* (or other similar) gene is likely to be a major impediment during the first two cycles of stage "A" of the PEX PCR method, but this is a clear next step in the development of this technique. If primer dimerization

in reactions with extremely degenerate primer pools is observed, purification of stage "A" components using a size-selection protocol (*e.g.* AMPure beads) instead of exonuclease will reduce the transfer of primer dimers which are insensitive to single-strand exonuclease activity.

The PEX PCR method may also find wide-spread application for quantitative PCRs in which degenerate primers are employed. Quantitative PCRs could be performed by initially processing DNA samples prepared using stage "A" of the PEX PCR method. Subsequently, qPCR would be performed using primers targeting linker sequences instead of template-specific primers. This could potentially greatly increase qPCR efficiency and target range for broad-target degenerate primers common to environmental microbiology, and avoid problems deriving from differential amplification efficiencies for different targets. A similar approach has in fact been developed, with the aim to reduce the impact of bacterial DNA contamination in qPCR reactions [70].

Finally, we note that this method has conceptual similarities to a study previously performed by Crosby and Criddle [69], in which linker sequences connected to random primers were used to amplify functional genes that were captured using hybridization probes. In that study, two cycles of annealing and elongation were used for labeling, with subsequent amplification. In addition, Illumina has developed a target-capture approach in which two primers with linkers either at the 5' or 3' ends are allowed to anneal to a single strand of template DNA (*i.e.* TruSeq Amplicon). After polymerase extension, ligation is used to link the elongated fragment to the 3' terminal primer with a 3' flanking linker. Subsequently, PCR amplification using the linker sequences is used to prepare fragments for sequencing.

Supporting Information

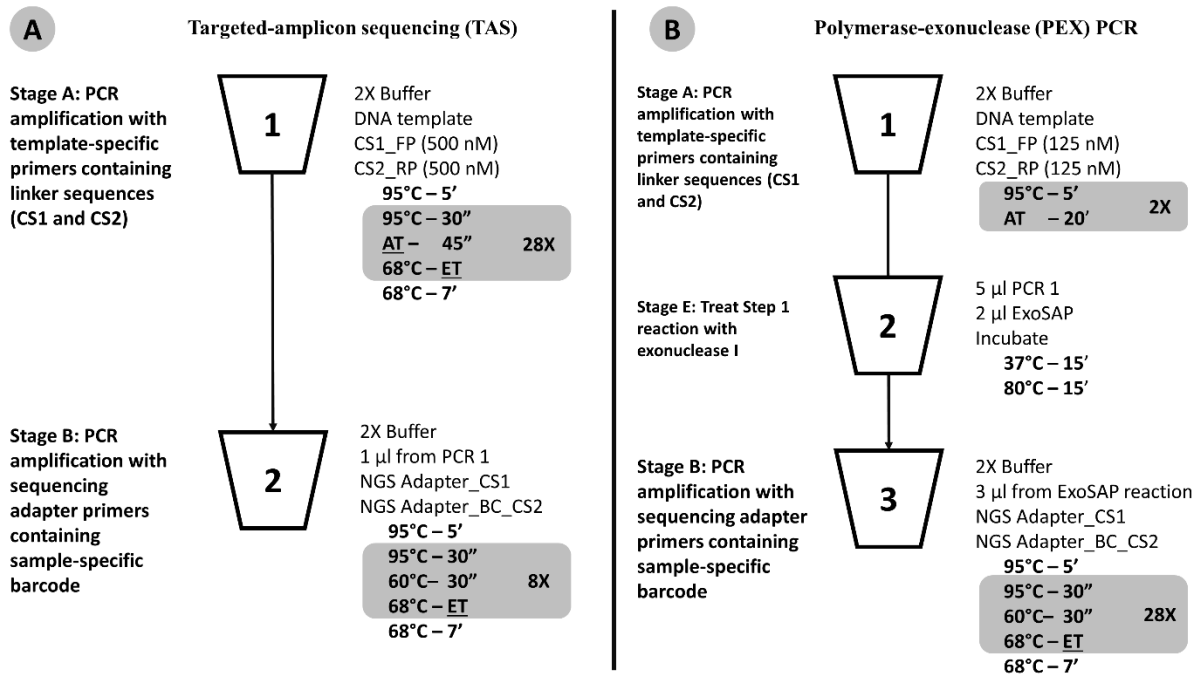


Figure S1. Sequencing workflow using targeted-amplicon sequencing (TAS) (A) and the polymerase-exonuclease-PCR (PEX PCR) method (B).

The TAS workflow consists of two PCR stages in which template-specific primers containing 5' linker sequences are used to amplify from template DNA. Subsequently, an aliquot of the first PCR is transferred to a second reaction for amplification with primers containing NGS sequencing adapters and a sample-specific barcode. In the PEX PCR method, a modified workflow is used; the first stage reaction is truncated after 2 cycles, primers are removed using exonuclease digestion, and the exonuclease-treated reaction mixture is subsequently PCR-amplified with primers containing sequencing adapters and barcodes. AT = annealing temperature; ET = Elongation time.

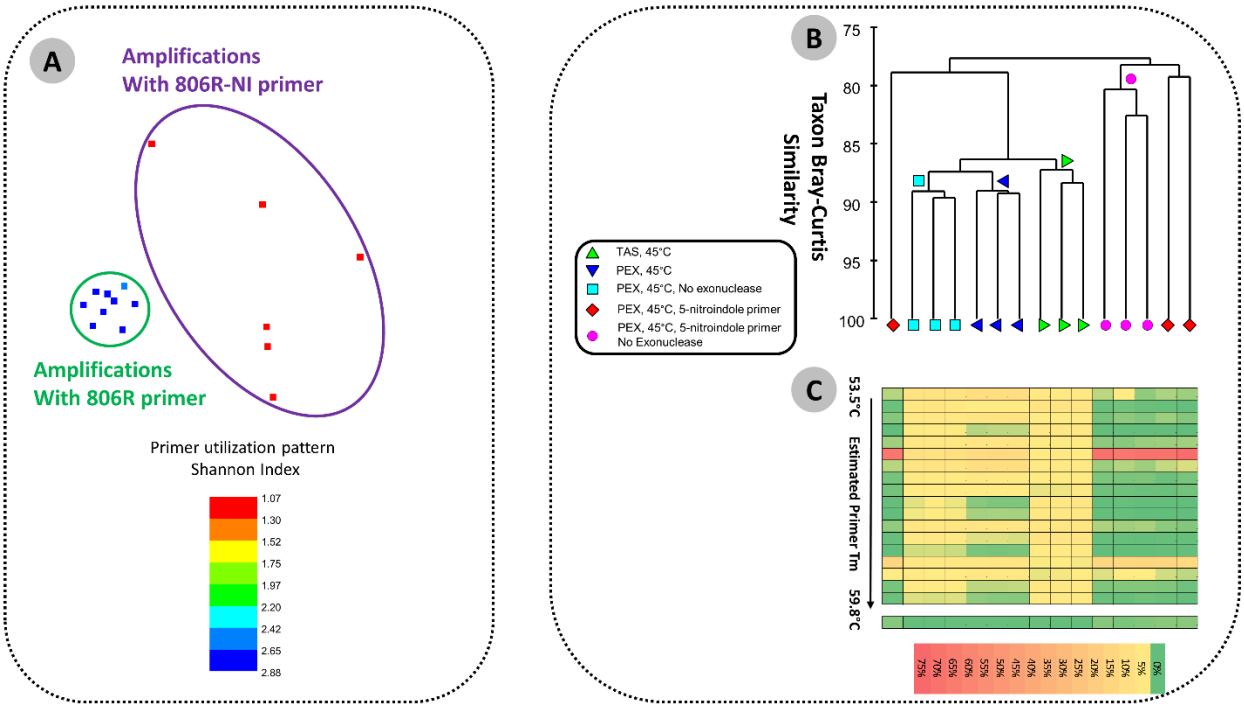


Figure S2. Effect of 5-nitroindole substitution and amplification strategy on observed microbial community structure and primer utilization patterns.

(A) Non-metric multidimensional scaling (NMDS) plot of lake sediment microbiome, performed at the taxonomic level of family and based on Bray-Curtis similarity (2D stress =0.02). Samples were rarefied to 4,750 sequences per sample and no transformation was applied. Symbols are color-coded by the diversity (Shannon Index) of reverse primers (*i.e.* 806R) detected in the sequences. Maximum possible Shannon index for 18 primers in the primer pool is 2.89. Reactions in which the 806R primer with 5-nitroindole (806R-NI) substitutions was used were less reproducible. **(B)** Group-average dendrogram of observed lake sediment microbial community structure from a single sample as amplification method is altered. gDNA was PCR amplified using the standard TAS reaction and with PEX PCR reactions with and without exonuclease and with and without primers containing 5-nitroindole substitutions. Bray-Curtis similarity scores were generated based on family-level taxonomic classification, generated as

described in the text. Data were standardized but not transformed. Clusters containing all three replicates from a single treatment are indicated by a symbol at the node. **(C)** Dendrogram and heatmap of reverse (806R) primer utilization patterns for the same samples. Bray-Curtis similarity was generated based on standardized abundance of each of 18 primer variants present in the reverse primer pool. The heatmap indicates relative abundance of each primer variant for each sample, with primers ordered by increasing theoretical Tm. A separate column (at the very bottom) indicates the relative abundance of variants potentially present when 5-nitroindole primers are used.

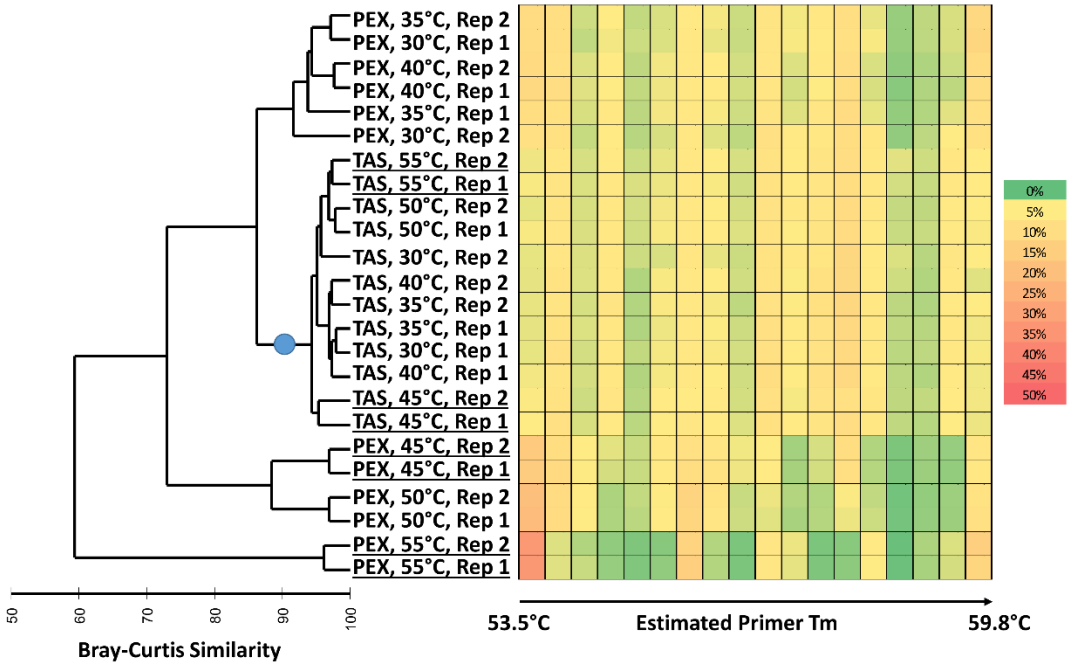


Figure S3. Effect of method and annealing temperature on primer utilization patterns in mock DNA.

Dendrogram and heatmap of reverse primer utilization patterns for the mock community analyzed using the TAS and PEX PCR methods, at temperatures from 30°-55°C. Bray-Curtis similarity was generated based on standardized abundance of each of 18 primer variants present

in the reverse primer pool. The heatmap indicates relative abundance of each primer variant for each sample, with primers ordered by increasing theoretical Tm. All reactions using the TAS method clustered together (node indicated by a blue circle). Underlined samples are analyzed at the individual template level in **Figure S4**.

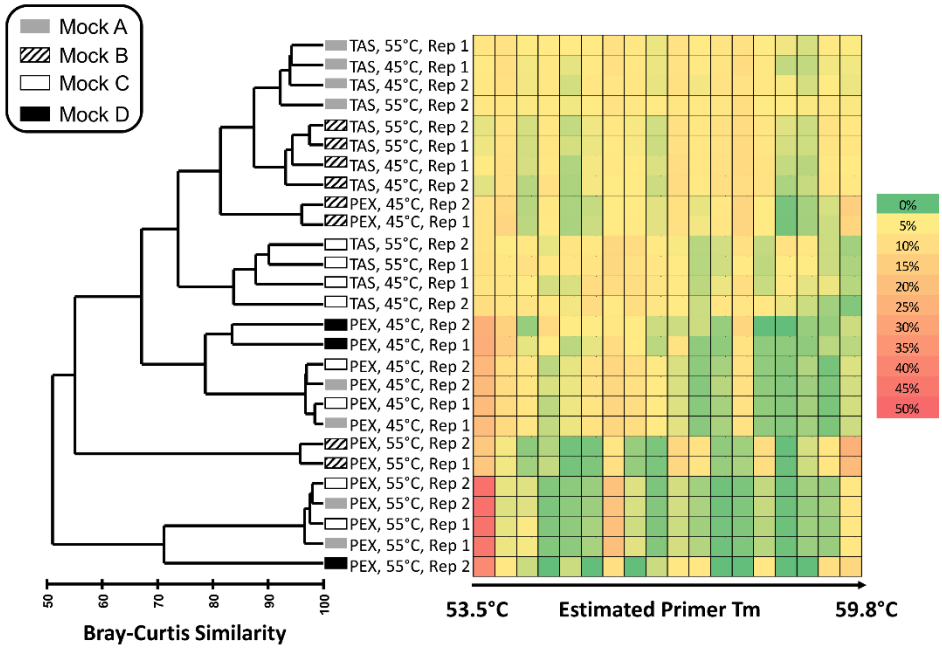


Figure S4. Effect of method and annealing temperature on primer utilization patterns for each mock template.

Dendrogram and heatmap of reverse primer utilization patterns for the mock community analyzed using the TAS and PEX PCR methods, at temperatures of 45° and 55°C. Bray-Curtis similarity was generated based on standardized abundance of each of 18 primer variants present in the reverse primer pool. The heatmap indicates relative abundance of each primer variant for each template within the mock community DNA pool, with primers ordered by increasing theoretical Tm.

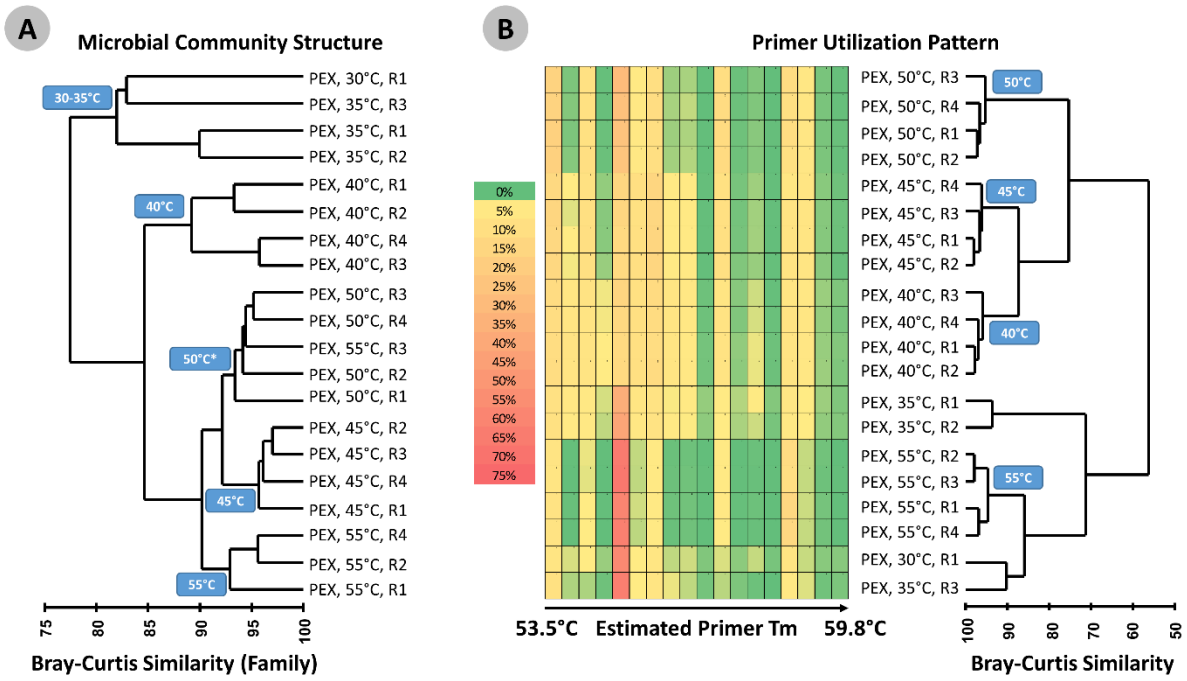


Figure S5. Effect of annealing temperature on observed microbial community structure and primer utilization patterns.

(A) Group-average dendrogram of observed mammalian fecal microbial community structure from a single sample ("Chin") as PEX PCR stage 1 annealing temperature is altered. Labeled nodes indicate grouping of replicates from a single annealing temperature (* indicates a single replicate from 55°C is included). Bray-Curtis similarity scores were generated based on family-level biological data, generated as described in the text. Data were standardized but not transformed. **(B)** Dendrogram and heatmap of reverse primer utilization patterns for the same samples. Bray-Curtis similarity was generated based on standardized abundance of each of 18 primer variants present in the reverse primer pool. The heatmap (0-75%) indicates the relative abundance of each primer variant for each sample, with primers ordered by increasing theoretical Tm.

Table S1: Sequences of artificial DNA fragments

>Mock A (FP Tm = Low / RP Tm = Low)

TAAAGCACTTTTATCAGGAGCGAAATACCACGGGTTAATACCCTATGGGGCTGACG
GTACCTGAGGAATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAG
GGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGGGTGCGTAGGCGGTTACTTAA
GTCTGTCGTGAAATCCCCGGGCTCAACCTGGGAATGGCGATGGATACTGGGTGGCTA
GAGTGTGTCAGAGGATGGTGGAATTCCCGGTGTAGCGGTGAAATGCGTAGAGATCG
GGAGGAACATCAGTGGCGAAGGCGGCCATCTGGGACAACACTGACGCTGAAGCAC
GAAAGCGTGGGGAGCAAACAGGATTAGATACCCTAGTAGTCCACGCCCTAAACGAT
GCGAACTGGATGTTGGTCTCAACTCGGAGATCAGTGTCGAAGCTAACGCGTTAAGTT
CGCCGCCTGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAA

>Mock B (FP Tm = Low / RP Tm = High)

TAAAGCACTTTTATCAGGAGCGAAATACCACGGGTTAATACCCTATGGGGCTGACG
GTACCTGAGGAATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAG
GGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGGGTGCGTAGGCGGTTACTTAA
GTCTGTCGTGAAATCCCCGGGCTCAACCTGGGAATGATAGGTAGCGCTGGGTGGCTA
GAGTGTGTCAGAGGATGGTGGAATTCCCGGTGTAGCGGTGAAATGCGTAGAGATCG
GGAGGAACATCAGTGGCGAAGGCGGCCATCTGGGACAACACTGACGCTGAAGCAC
GAAAGCGTGGGGAGCAAACAGGATTAGAAACCCGGGTAGTCCACGCCCTAAACGAT
GCGAACTGGATGTTGGTCTCAACTCGGAGATCAGTGTCGAAGCTAACGCGTTAAGTT
CGCCGCCTGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAA

>Mock C (FP Tm = Low + mismatch / RP Tm = Low)

TAAAGCACTTTTATCAGGAGCGAAATACCACGGGTTAATACCCTATGGGGCTGACG
GTACCTGAGGAATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTCATACGAAG
GGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGGGTGCGTAGGCGGTTACTTAA
GTCTGTCGTGAAATCCCCGGGCTCAACCTGGGAATGGACTGAAGTGCTGGGTGGCTA
GAGTGTGTCAGAGGATGGTGGAATTCCCGGTGTAGCGGTGAAATGCGTAGAGATCG
GGAGGAACATCAGTGGCGAAGGCGGCCATCTGGGACAACACTGACGCTGAAGCAC
GAAAGCGTGGGGAGCAAACAGGATTAGATACCCTAGTAGTCCACGCCCTAAACGAT
GCGAACTGGATGTTGGTCTCAACTCGGAGATCAGTGTCGAAGCTAACGCGTTAAGTT
CGCCGCCTGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAA

>Mock D (FP Tm = Low + mismatch / RP Tm = Low + mismatch)

TAAAGCACTTTTATCAGGAGCGAAATACCACGGGTTAATACCCTATGGGGCTGACG
GTACCTGAGGAATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTCATACGAAG
GGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGGGTGCGTAGGCGGTTACTTAA
GTCTGTCGTGAAATCCCCGGGCTCAACCTGGGAATGGGATAGCGATCTGGGTGGCTA
GAGTGTGTCAGAGGATGGTGGAATTCCCGGTGTAGCGGTGAAATGCGTAGAGATCG
GGAGGAACATCAGTGGCGAAGGCGGCCATCTGGGACAACACTGACGCTGAAGCAC
GAAAGCGTGGGGAGCAAACAGGAGTAGATACCCTAGTAGTCCACGCCCTAAACGAT
GCGAACTGGATGTTGGTCTCAACTCGGAGATCAGTGTCGAAGCTAACGCGTTAAGTT
CGCCGCCTGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAA

Artificial fragments are 492 bp gene fragments synthesized by Integrated DNA Technologies (Coralville, Iowa). The fragments are derived from the sequence of the small subunit (SSU or 16S) ribosomal RNA (rRNA) gene of *Rhodanobacter denitrificans* 2APBS1, a denitrifying Gammaproteobacteria (GenBank accession number: NR_102497.1). The sequences shown represent one strand (5'-3') of the double-stranded molecules. The sequences highlighted in green represent the target position of the 515F primer, while the light blue sequences represent the target position of 806R primer [56]. Positions highlighted in red indicate introduced mismatches relative to the primer pool, introduced into the synthesized DNA fragments to allow for testing of effects of mismatches between template and primer. The region highlighted in purple is a 10-bp region which is diagnostic for each of the four sequences. The 10 bp are re-arranged in each region to allow for identification of the template, while maintaining overall similarity of the gene fragment. The last two gene fragments contain mismatches that are not targeted by any of the primer variants found in the standard degenerate 515F and 806R primers. The forward primer (515F) primer site is highlighted in bright green, and the reverse primer site is highlighted in cyan when matching the lowest Tm reverse primer or olive green when matching the highest Tm reverse primer.

Table S2: Primer sites and mismatches with mock double-stranded DNA templates

			Mock A	Mock B	Mock C	Mock D
			(L/L)	(L/H)	(L+1mm/L)	(L+1mm/L+1mm)
Forward primers	**Primer sequence**	**Tm***	GTGCCAGCAGCC GCGGTAA	GTGCCAGCAGCC GCGGTAA	GTGCCAGCAGCC GCGGTCA	GTGCCAGCAGCC GCGGTCA
Degenerate Pool	**GTGCCAGCMGCCGCGGTAA**		0	0	1	1
FP-Variant 1	GTGCCAGCAGCCGCGGTAA	68.5	0	0	1	1
FP-Variant 2	GTGCCAGCCGCCGCGGTAA	70.7	1	1	2	2
Reverse Primers	**Primer sequence**	**Tm***	GGACTACTAGGG TATCTAAT	GGACTACCCGGG TTTCTAAT	GGACTACTAGGG TATCTAAT	GGACTACTAGGG TATCTACT
Degenerate Pool	**GGACTACHVGGGTWTCTAAT**		0	0	0	1
RP-Variant 1	GGACTACTAGGGTATCTAAT	53.5	0	3	0	1
RP-Variant 2	GGACTACTAGGGTTTCTAAT	54.6	1	2	1	2
RP-Variant 3	GGACTACAAGGGTATCTAAT	54.7	1	3	1	2
RP-Variant 4	GGACTACAAGGGTTTCTAAT	55.7	2	2	2	3
RP-Variant 5	GGACTACCAGGGTATCTAAT	56.4	1	2	1	2
RP-Variant 6	GGACTACAGGGGTATCTAAT	56.4	2	3	2	3
RP-Variant 7	GGACTACTGGGGTATCTAAT	56.4	1	3	1	2
RP-Variant 8	GGACTACTCGGGTATCTAAT	56.4	1	2	1	2
RP-Variant 9	GGACTACACGGGTATCTAAT	57	2	3	2	3
RP-Variant 10	GGACTACTCGGGTTTCTAAT	57.4	2	1	2	3
RP-Variant 11	GGACTACCAGGGTTTCTAAT	57.5	2	1	2	3
RP-Variant 12	GGACTACAGGGGTTTCTAAT	57.5	3	2	3	4
RP-Variant 13	GGACTACTGGGGTTTCTAAT	57.5	2	2	2	3
RP-Variant 14	GGACTACACGGGTTTCTAAT	58	3	1	3	4
RP-Variant 15	GGACTACCGGGGTATCTAAT	58.7	2	2	2	3
RP-Variant 16	GGACTACCCGGGTATCTAAT	58.7	2	1	2	3
RP-Variant 17	GGACTACCGGGGTTTCTAAT	59.8	3	1	3	4
RP-Variant 18	GGACTACCCGGGTTTCTAAT	59.8	3	0	3	4
* Tm (°C) calculated using OligoAnalyzer3.1, assuming 125 nM primer concentration, 2 mM Mg^{2+}, and 0.2 mM dNTPs.						

Degenerate positions in the primer pools are indicated with red, underlined letters. For each of the mock DNA templates (Mock A, B, C and D) number of variants with each individual primer

in the pool is indicated. Additional mismatches were introduced near the 3' end of the annealing site for forward and reverse primers, as indicated by red, underlined letters for Mock C and D templates.

Table S3: List of FASTQ filenames and associated sample preparation details (BioProject PRJNA262579).

Sample Name	Sequencer	Barcode / MID	DNA Template	PEP/TAS	Temp(°C)	Exo-nuclease?	Primer (806R/806_NI)
M1_1_2cy_28cy	MiSeq	TGCTACATCA	Mock	PEP	45	No	806R
M1_2_2cy_28cy	MiSeq	TGATAGAGAG	Mock	PEP	45	No	806R
M1_3_2cy_28cy	MiSeq	CTCAGCAGTG	Mock	PEP	45	No	806R
M1e_1_2cy_28cy	MiSeq	CAGCTATAGC	Mock	PEP	45	Yes	806R
M1e_2_2cy_28cy	MiSeq	GACTCATGCT	Mock	PEP	45	Yes	806R
M1e_3_2cy_28cy	MiSeq	CACATACAGT	Mock	PEP	45	Yes	806R
m1_1_45c_28_8	MiSeq	AATATGCTGC	Mock	TAS	45	No	806R
m1_2_45c_28_8	MiSeq	AGAGGTCGGA	Mock	TAS	45	No	806R
m1_3_45c_28_8	MiSeq	ATCTGTCCAT	Mock	TAS	45	No	806R
M2_1_2cy_28cy	MiSeq	ACTCGATAGT	Mock	PEP	45	No	806R_NI
M2_2_2cy_28cy	MiSeq	CACGAGATGA	Mock	PEP	45	No	806R_NI
M2_3_2cy_28cy	MiSeq	TATAGAGATC	Mock	PEP	45	No	806R_NI
M1_1_55c_28_8	MiSeq	TCTAGCGTGG	Mock	TAS	55	No	806R
M1_2_55c_28_8	MiSeq	TCTCGGATAG	Mock	TAS	55	No	806R
M1_3_55c_28_8	MiSeq	GTATAACGCT	Mock	TAS	55	No	806R
S1.1.2cy.28cy	MiSeq	TCATATCGCG	Sed	PEP	45	No	806R
S1.2.2cy.28cy	MiSeq	TGCGAGACGT	Sed	PEP	45	No	806R
S1.3.2cy.28cy	MiSeq	TACTGCAGCG	Sed	PEP	45	No	806R
s1.1.45cy.28.8cy	MiSeq	GACAGCAAGC	Sed	TAS	45	No	806R
s1.2.45cy.28.8cy	MiSeq	AAGTACACTC	Sed	TAS	45	No	806R
s1.3.45cy.28.8cy	MiSeq	AGTGGCAGGT	Sed	TAS	45	No	806R
S1e.1.2cy.28cy	MiSeq	GCACGCGTAT	Sed	PEP	45	Yes	806R
S1e.2.2cy.28cy	MiSeq	ACTAGCTGTC	Sed	PEP	45	Yes	806R
S1e.3.2cy.28cy	MiSeq	CGAGCTAGCA	Sed	PEP	45	Yes	806R
S2.1.2cy.28cy	MiSeq	TCATCATGCG	Sed	PEP	45	No	806_NI
S2.2.2cy.28cy	MiSeq	ACGTGCTCTG	Sed	PEP	45	No	806_NI
S2.3.2cy.28cy	MiSeq	TACATGATAG	Sed	PEP	45	No	806_NI
S2e.1.2cy.28cy	MiSeq	AGAGTCGCGT	Sed	PEP	45	Yes	806_NI
S2e.2.2cy.28cy	MiSeq	GATATATGTC	Sed	PEP	45	Yes	806_NI
S2e.3.2cy.28cy	MiSeq	ATCATATCTC	Sed	PEP	45	Yes	806_NI
Ma1	PGM	MID-19R	Mock	PEP	30	Yes	806R
Ma2	PGM	MID-20R	Mock	PEP	35	Yes	806R
Ma3	PGM	MID-21R	Mock	PEP	40	Yes	806R
Ma4	PGM	MID-22R	Mock	PEP	45	Yes	806R
Ma5	PGM	MID-23R	Mock	PEP	50	Yes	806R
Ma6	PGM	MID-24R	Mock	PEP	55	Yes	806R

Ma1D	PGM	MID-27R	Mock	PEP	30	Yes	806R
Ma2D	PGM	MID-28R	Mock	PEP	35	Yes	806R
Ma3D	PGM	MID-29R	Mock	PEP	40	Yes	806R
Ma4D	PGM	MID-30R	Mock	PEP	45	Yes	806R
Ma5D	PGM	MID-31R	Mock	PEP	50	Yes	806R
Ma6D	PGM	MID-32R	Mock	PEP	55	Yes	806R
Mb1	PGM	MID-51R	Mock	PEP	30	Yes	806R
Mb2	PGM	MID-52R	Mock	PEP	35	Yes	806R
Mb3	PGM	MID-53R	Mock	PEP	40	Yes	806R
Mb4	PGM	MID-54R	Mock	PEP	45	Yes	806R
Mb5	PGM	MID-55R	Mock	PEP	50	Yes	806R
Mb6	PGM	MID-56R	Mock	PEP	55	Yes	806R
Mb1D	PGM	MID-59R	Mock	PEP	30	Yes	806R
Mb2D	PGM	MID-60R	Mock	PEP	35	Yes	806R
Mb3D	PGM	MID-61R	Mock	PEP	40	Yes	806R
Mb4D	PGM	MID-62R	Mock	PEP	45	Yes	806R
Mb5D	PGM	MID-63R	Mock	PEP	50	Yes	806R
Mb6D	PGM	MID-64R	Mock	PEP	55	Yes	806R
Mc1	PGM	MID-83R	Mock	TAS	30	No	806R
Mc2	PGM	MID-84R	Mock	TAS	35	No	806R
Mc3	PGM	MID-85R	Mock	TAS	40	No	806R
Mc4	PGM	MID-86R	Mock	TAS	45	No	806R
Mc5	PGM	MID-87R	Mock	TAS	50	No	806R
Mc6	PGM	MID-88R	Mock	TAS	55	No	806R
Mc1D	PGM	MID-91R	Mock	TAS	30	No	806R
Mc2D	PGM	MID-92R	Mock	TAS	35	No	806R
Mc3D	PGM	MID-93R	Mock	TAS	40	No	806R
Mc4D	PGM	MID-94R	Mock	TAS	45	No	806R
Mc5D	PGM	MID-95R	Mock	TAS	50	No	806R
Mc6D	PGM	MID-96R	Mock	TAS	55	No	806R
A1	PGM	MID-1R	Chin	PEP	30	Yes	806R
A2	PGM	MID-2R	Chin	PEP	35	Yes	806R
A3	PGM	MID-3R	Chin	PEP	40	Yes	806R
A4	PGM	MID-4R	Chin	PEP	45	Yes	806R
A5	PGM	MID-5R	Chin	PEP	50	Yes	806R
A6	PGM	MID-6R	Chin	PEP	55	Yes	806R
A1D	PGM	MID-10R	Chin	PEP	30	Yes	806R
A2D	PGM	MID-11R	Chin	PEP	35	Yes	806R
A3D	PGM	MID-13R	Chin	PEP	40	Yes	806R
A4D	PGM	MID-14R	Chin	PEP	45	Yes	806R
A5D	PGM	MID-15R	Chin	PEP	50	Yes	806R
A6D	PGM	MID-16R	Chin	PEP	55	Yes	806R
B2	PGM	MID-36R	Chin	PEP	35	Yes	806R
B3	PGM	MID-37R	Chin	PEP	40	Yes	806R
B4	PGM	MID-38R	Chin	PEP	45	Yes	806R
B5	PGM	MID-39R	Chin	PEP	50	Yes	806R
B6	PGM	MID-40R	Chin	PEP	55	Yes	806R
B1D	PGM	MID-43R	Chin	PEP	30	Yes	806R

B2D	PGM	MID-44R	Chin	PEP	35	Yes	806R
B3D	PGM	MID-45R	Chin	PEP	40	Yes	806R
B4D	PGM	MID-46R	Chin	PEP	45	Yes	806R
B5D	PGM	MID-47R	Chin	PEP	50	Yes	806R
B6D	PGM	MID-48R	Chin	PEP	55	Yes	806R

S4 File: Temperature gradient tests of PEX and TAS methods using mock community DNA ("Chin") at varying annealing temperatures

Method	Annealing Temperature	Exonuclease	Reverse Primer	Average percent abundance of reads mapping to references				Ideal Score	Ideal Score	Ideal Score
Stage "A" Annealing Temperature, PEX PCR Method				**Mock A**	**Mock B**	**Mock C**	**Mock D**	**4 targets**	**3 targets**	**2 targets**
PEX PCR	30	Yes	806R	30.40	58.19	11.12	0.29	77*	50*	31[NS]
PEX PCR	35	Yes	806R	15.49	57.76	26.36	0.39	68*	49*	58
PEX PCR	40	Yes	806R	15.46	51.70	32.29	0.56	68*	38*	54[NS]
PEX PCR	45	Yes	806R	32.85	35.00	31.56	0.59	49*	4*	3*
PEX PCR	50	Yes	806R	32.62	35.23	31.93	0.23	50*	4*	4*
PEX PCR	55	Yes	806R	29.68	41.73	28.44	0.15	50*	17*	17*
Stage "A" Annealing Temperature, TAS PCR Method										
TAS PCR	30	No	806R	33.26	63.77	2.94	0.04	94	61	31
TAS PCR	35	No	806R	27.89	65.93	6.11	0.08	88	65	41[#]
TAS PCR	40	No	806R	21.06	70.54	8.34	0.06	91	74	54
TAS PCR	45	No	806R	34.58	58.95	6.42	0.05	87	54	26
TAS PCR	50	No	806R	29.81	63.35	6.78	0.06	86	60	36
TAS PCR	55	No	806R	18.66	76.33	4.93	0.08	103	86	61

* Significant decrease in PEX PCR method relative to TAS PCR method; p<0.03, two-tailed TTEST (unequal variance)
# Significant decrease in TAS PCR relative to PEX PCR method; p<0.03, two-tailed TTEST (unequal variance)
NS = Not significant, two-tailed TTEST (unequal variance)

Chapter III: Deconstructing the Polymerase Chain Reaction II: An improved workflow and effects on artifact formation and primer degeneracy.

Abstract

Polymerase chain reaction (PCR) amplification of complex microbial genomic DNA templates with degenerate primers can lead to distortion of the underlying community structure due to inefficient primer-template interactions leading to bias. We previously described a method of deconstructed PCR ("PEX PCR") to separate linear copying and exponential amplification stages of PCR to reduce PCR bias [18]. In this manuscript, we describe an improved deconstructed PCR ("DePCR") protocol separating linear and exponential stages of PCR and allowing higher throughput of sample processing. We demonstrate that the new protocol shares the same benefits of the original and show that the protocol dramatically and significantly decreases the formation of chimeric sequences during PCR. By employing PCR with annealing temperature gradients, we further show that there is a strong negative correlation between annealing temperature and the evenness of primer utilization in a complex pool of degenerate primers. Shifting primer utilization patterns mirrored shifts in observed microbial community structure in a complex microbial DNA template. We further employed the DePCR method to amplify the same microbial DNA template independently with each primer variant from a degenerate primer pool. The non-degenerate primers generated a broad range of observed microbial communities, but some were highly similar to communities observed with degenerate primer pools. The same experiment conducted with standard PCR led to consistently divergent observed microbial community structure. The DePCR method is simple to perform, is limited to PCR mixes and cleanup steps, and is recommended for reactions in which degenerate primer pools are used or when mismatches between primers and template are possible.

Introduction

The small subunit (SSU) ribosomal RNA (rRNA) gene is the most frequently targeted gene in studies of complex microbial systems. A common approach for microbial community studies is to extract genomic DNA (gDNA) from multiple samples, PCR amplify gDNA using locus-specific SSU rRNA gene primers containing sequencing adapters and a sample-specific barcode, and equimolar pooling and sequencing [49]. A number of major caveats are associated with such an approach: (i) Microorganisms contain a variable number of rRNA operons [71, 72] and analyses of rRNA genes present a distorted representation of relative cellular abundance; (ii) PCR primer pools are often degenerate or the primers are anticipated to anneal to template sequences containing mismatches with the primers, thereby producing bias in amplification efficiency among different templates; and (iii) samples are generally heavily amplified (30 cycles or more) leading to the possibility of extensive chimera formation.

Recently, we identified a novel source of PCR bias – namely, the simultaneous operation of linear copying and exponential amplification during the early cycles of PCR with degenerate primers [18]. We hypothesized that primer-genomic DNA template annealing operates at a different, and likely lower, efficiency compared to primer-amplicon annealing. These primer-template interactions, operating at different efficiencies, both contribute to distortion of the underlying template community, particularly in the early cycles of PCR. To address this source of bias, we developed the polymerase-exonuclease (PEX) PCR method to separate PCR into two distinct stages of linear copying and exponential amplification. Furthermore, the PEX PCR method prevents the locus-specific primers. Although effective, the PEX PCR method requires an enzymatic step (exonuclease), which lengthens the workflow. We sought to improve upon the prior protocol and remove the effort associated with exonuclease treatment. Nonetheless, the

PEX PCR method – and the separation of linear copying and exponential amplification – serves as the conceptual foundation for the new method. In PEX PCR, after two cycles of linear amplification with locus-specific primers containing 5' non-degenerate linker sequences, the initial stage of the reaction is terminated, primers are removed with exonuclease I treatment, and the linear copies subsequently amplified using non-degenerate primers targeting the 5' linker sequences **(Figure 8**). Here, we present a method that replaces exonuclease treatment with size-selective bead-based purification (*e.g.* AMPure XP beads) but achieves substantial savings in overall labor and sample manipulation by a pooling of all samples prior to purification.

The primary objective of this study was to develop an improved pipeline for utilizing the PEX PCR concept, while retaining the ability to reduce PCR bias. To demonstrate the effectiveness of the updated workflow, we replicated a temperature-gradient analysis of a single complex environmental genomic DNA sample using both standard PCR and DePCR workflows. Data were interrogated to examine the observed microbial community structure by method and reaction annealing temperature. In addition, primer utilization profiles (PUPs) were analyzed to assess the effects of annealing temperature on the relative utilization of each primer within a degenerate pool of primers. Subsequently, we examined the behavior of the amplification system with varying input gDNA. A final experiment examined the ability of each unique primer within a degenerate primer pool to amplify a complex environmental sample using both the standard PCR and DePCR methodologies.

Materials and Methods

DNA Templates

A single microbial genomic DNA (gDNA) sample obtained from chinchilla feces was used for this study. The fecal sample was extracted using the PowerSoil DNA extraction kit (Mo Bio Laboratories, Carlsbad, CA).

Primer Synthesis

The primers used for this study are 341F (CCTACGGGAGGCAGCAG) [10, 56] and 806R (GGACTAC**HV**GGGT**W**TCTAAT) [56, 73]. The 806R primer pool is 18-fold degenerate, with theoretical melting temperatures ranging from 54.7°C to 61°C. Melting temperatures of the primers were calculated using the OligoAnalyzer3.1 tool [74], assuming 250 nM primer concentration, 2 mM Mg^{2+}, and 0.2mM dNTPs. Synthesis of the primers was performed either as single degenerate primer pools (standard approach), or as individual primers without degeneracies by Integrated DNA Technologies (IDT; Coralville, IA). Primers were synthesized as LabReady and ordered at a fixed concentration of 100 micromolar. Primers contained common sequence linkers (CS1 and CS2) at the 5' ends, as shown in **Table 4**. Linker sequences are required for the later incorporation of Illumina sequencing adapters and sample-specific barcodes.

Table 4: Primers used in this study

341F Primer	Primer Sequence	Linker (CS1) Sequence	Final Sequence Name	Final Sequence Ordered
341F	CCTACGGGAGGCAGCAG	**ACACTGACGACATGGTTCTACA**	>CS1_515F	**ACACTGACGACATGGTTCTACA**CCTACGGGAGGCAGCAG

806R Primer and Variants	Primer Sequence	Linker (CS2) Sequence	Final Sequence Name	Final Sequence Ordered
806R	GGACTAC**HV**GGGT**W**TCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R	**TACGGTAGCAGAGACTTGGTCT**GGACTACHVGGGTWTCTAAT
806R-RPV1	GGACTACTAGGGTATCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R_V1	**TACGGTAGCAGAGACTTGGTCT**GGACTACTAGGGTATCTAAT
806R-RPV2	GGACTACTAGGGTTTCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R_V2	**TACGGTAGCAGAGACTTGGTCT**GGACTACTAGGGTTTCTAAT
806R-RPV3	GGACTACAAGGGTATCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R_V3	**TACGGTAGCAGAGACTTGGTCT**GGACTACAAGGGTATCTAAT
806R-RPV4	GGACTACAAGGGTTTCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R_V4	**TACGGTAGCAGAGACTTGGTCT**GGACTACAAGGGTTTCTAAT
806R-RPV5	GGACTACCAGGGTATCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R_V5	**TACGGTAGCAGAGACTTGGTCT**GGACTACCAGGGTATCTAAT
806R-RPV6	GGACTACAGGGGTATCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R_V6	**TACGGTAGCAGAGACTTGGTCT**GGACTACAGGGGTATCTAAT
806R-RPV7	GGACTACTGGGGTATCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R_V7	**TACGGTAGCAGAGACTTGGTCT**GGACTACTGGGGTATCTAAT
806R-RPV8	GGACTACTCGGGTATCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R_V8	**TACGGTAGCAGAGACTTGGTCT**GGACTACTCGGGTATCTAAT
806R-RPV9	GGACTACACGGGTATCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R_V9	**TACGGTAGCAGAGACTTGGTCT**GGACTACACGGGTATCTAAT
806R-RPV10	GGACTACTCGGGTTTCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R_V10	**TACGGTAGCAGAGACTTGGTCT**GGACTACTCGGGTTTCTAAT
806R-RPV11	GGACTACCAGGGTTTCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R_V11	**TACGGTAGCAGAGACTTGGTCT**GGACTACCAGGGTTTCTAAT
806R-RPV12	GGACTACAGGGGTTTCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R_V12	**TACGGTAGCAGAGACTTGGTCT**GGACTACAGGGGTTTCTAAT
806R-RPV13	GGACTACTGGGGTTTCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R_V13	**TACGGTAGCAGAGACTTGGTCT**GGACTACTGGGGTTTCTAAT
806R-RPV14	GGACTACACGGGTTTCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R_V14	**TACGGTAGCAGAGACTTGGTCT**GGACTACACGGGTTTCTAAT
806R-RPV15	GGACTACCGGGGTATCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R_V15	**TACGGTAGCAGAGACTTGGTCT**GGACTACCGGGGTATCTAAT
806R-RPV16	GGACTACCCGGGTATCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R_V16	**TACGGTAGCAGAGACTTGGTCT**GGACTACCCGGGTATCTAAT
806R-RPV17	GGACTACCGGGGTTTCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R_V17	**TACGGTAGCAGAGACTTGGTCT**GGACTACCGGGGTTTCTAAT
806R-RPV18	GGACTACCCGGGTTTCTAAT	**TACGGTAGCAGAGACTTGGTCT**	>CS2_806R_V18	**TACGGTAGCAGAGACTTGGTCT**GGACTACCCGGGTTTCTAAT

Illumina Primers				Final Sequence Ordered
P5				AATGATACGGCGACCACCGA
P7				CAAGCAGAAGACGGCATACGA

Standard PCR Protocol

The standard PCR protocol or targeted amplicon sequencing (TAS) protocol is a two-stage NGS library preparation protocol for generating barcoded amplicons ready for Illumina sequencing, and was performed as described previously [38] (**Figure 8A**). Briefly, gDNA was PCR amplified with primers CS1_341F and CS2_806R. The first stage PCR reaction was conducted in a total reaction volume of 10 µl. Each reaction contained 5 µl of MyTaq HS master mix (Bioline, Taunton, MA), 0.5 µl of each primer or degenerate primer at a concentration of 5 µM (*e.g.*, CS1_341F and CS2_806R; leading to a 250 nM working concentration), 10 ng of gDNA template, and water up to 10 µl total volume. The first stage of the PCR was conducted using the following thermocycling conditions: 95°C for 5 minutes, followed by 28 cycles of 95°C for 30 seconds, annealing temperature (from 40°C to 60°C) for 30 seconds, 72°C for 30 seconds; and a final elongation step at 72°C for 7 minutes. Subsequently, a second PCR amplification was performed in 10 µl reactions in 96-well plates to incorporate Illumina sequencing adapters and a sample-specific barcode. A mastermix for the entire plate was made using the MyTaq HS 2X mastermix. Each well received a separate primer pair with a unique 10-base barcode, obtained from the Access Array Barcode Library for Illumina (Fluidigm, South San Francisco, CA; Item# 100-4876). These Access Array primers contained the CS1 and CS2 linkers at the 3′ ends of the oligonucleotides. One µl of reaction mixture from the first stage amplification was used as input template for the second stage reaction, without cleanup. Cycling conditions were as follows: 95 °C for 5 minutes, followed by 8 cycles of 95 °C for 30”, 60 °C for 30” and 72 °C for 30”. A final, 7-minute elongation step was performed at 72 °C. Samples were pooled and sequenced on an Illumina MiSeq employing V2 chemistry and 2x250 base reads.

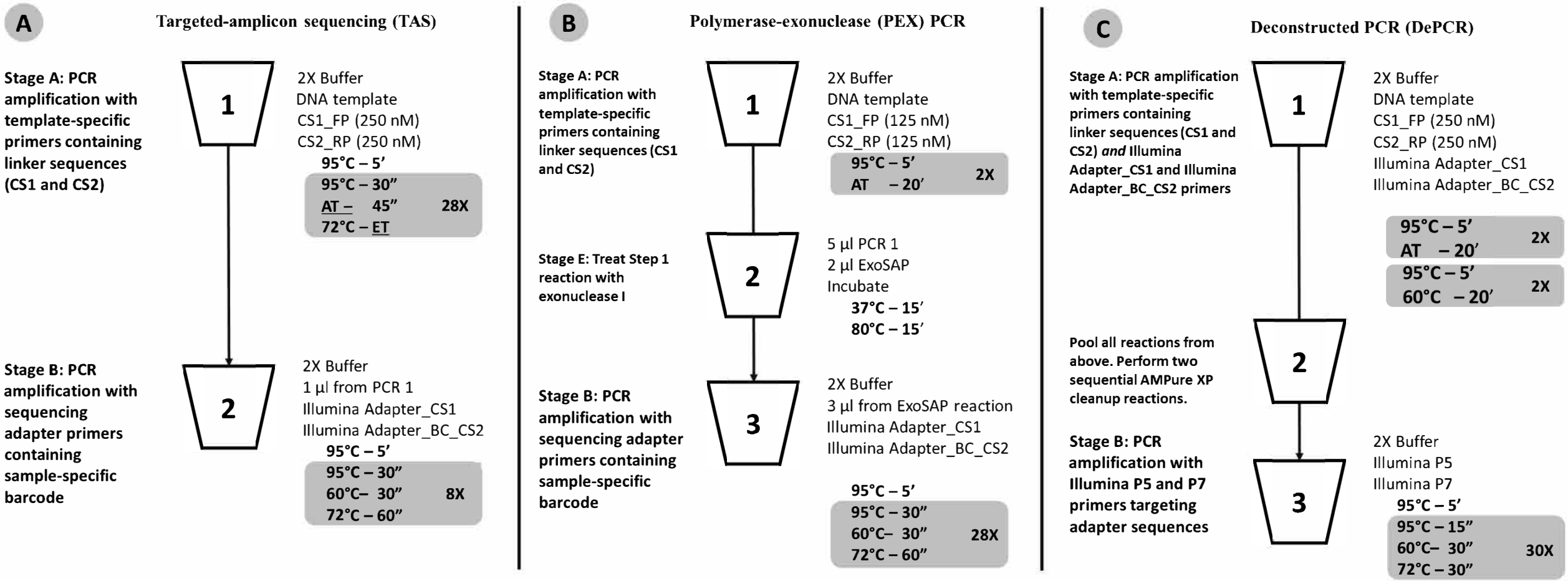


Figure 8: Schematic of (A) standard (TAS), (B) polymerase-exonuclease (PEX) PCR, and (C) Deconstructed PCR (DePCR) workflows

AT = annealing temperature; ET = Elongation time. CS1 = common sequence 1 adapter. CS2 = common sequence 2 adapter. BC = barcode. FP = Forward primer. RP = Reverse primer. Primer sequences are shown in **Figure 9** and **Table 4.**

Deconstructed PCR (DePCR) Protocol

As with the TAS method, the DePCR method is also a two-stage PCR process (**Figure 8C)** and is a modification of the previously described PEX PCR method **(Figure 8B)**. For each sample, the first stage reaction was conducted in a 96-well plate with each well containing 5 µl of MyTaq master mix, 0.5 µl of each primer or degenerate primer at a concentration of 5 µM (*e.g.*, CS1_341F and CS2_806R; leading to a 250 nM working concentration), 10 ng of template, 1 µl Access Array Barcode Library containing a unique sample-specific barcode, and water up to 10 µl. The thermocycler conditions for first stage were composed of two cycles of denaturation at 95°C for 5 minutes and annealing (40°C-60°C, depending on experiment) for 20 minutes, followed by two cycles of denaturation for 5 minutes at 95°C and annealing at 60°C for 20 minutes, and a final extension temperature of 72°C for 10 minutes. For temperature gradient experiments, annealing temperatures of 40°C, 45°C, 50°C, 55°C, and 60°C were tested. For single reverse primer variant (RPV) analyses, an annealing temperature of 50°C was used for both TAS and DePCR amplification reactions. Subsequently, a pool composed of 5 µl from the first reaction of each sample was collected and processed for cleanup using AMPure XP beads (Beckman-Coulter) at 0.7X per the manufacturer's recommendations. The cleaning step was performed twice, sequentially. A final elution volume of 20 µl was used to concentrate the sample prior to the second stage of the DePCR reaction. The second stage reactions were conducted in a final volume of 20 µl; the reaction contained 10 µl of MyTaq HS master mix, 1 µl of Illumina P5 (AATGATACGGCGACCACCGA) and P7 (CAAGCAGAAGACGGCATACGA) primers, 2 µl of purified template from pooled first stage PCR, and water up to 20 µl. The thermocycler conditions were: 95°C for 3 minutes, 30 cycles at 95°C for 15 seconds, 60°C for 30 seconds and 72°C for 30 seconds. Prior to sequencing the pool

libraries were purified using a Pippin Prep DNA Size Selection System (Sage Science), employing a 2% agarose gel cassette and selecting for fragment sizes from 450-600 bp. Sequencing of the amplified pool was performed on an Illumina MiSeq employing V2 chemistry and 2x250 base reads. Library preparation and sequencing were performed at the UIC Sequencing Core (UICSQC).

Sequence Data Analysis

Raw sequence FASTQ files were merged using the software package PEAR [75], with default parameters. Merged sequences were trimmed using the software package trimmomatic [12]. Sequences shorter than 400 bases and longer than 500 bases were removed. Sequences were then screened for chimeras using the USEARCH61 algorithm [76], and putative chimeric sequences were removed from the dataset. Subsequently, sequences were pooled and clustered into operational taxonomic units (OTUs) at a threshold of 97% similarity (QIIME v1.8.0) [1]. Representative sequences from all OTUs were annotated using the UCLUST algorithm and the Greengenes 13_8 reference database [65], and a biological observational matrix (BIOM) was generated this annotation pipeline [77]. The BIOM file was analyzed and visualized using the software package Primer7 [78] and the R environment [79]. The R package 'vegan' [80] was employed to generate alpha diversity indices (Shannon, richness, and evenness indices) and to perform rarefaction of BIOM files. Bray-Curtis dissimilarity indices were calculated within the R package 'vegan' and these indices were used to evaluate differences in composition between samples. Analysis of similarity (ANOSIM) calculations were performed at the taxonomic level of genus, using square root transformed data. Initial analysis and processing of the samples was performed using QIIME (v1.8.0) package scripts. Metric multi-dimensional scaling (mMDS)

plots were generated using the cmdscale and ggplot2 functions [81] within the R programming environment. Ellipses, representing a 95% confidence interval around group centroids, were drawn assuming a multivariate t-distribution. Some visualizations were performed using the software package OriginPro 2018 (OriginLab, Northampton, Mass).

Data Archive

Raw sequence data files were submitted in the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI). The BioProject identifier of the samples is PRJNA506229. Full metadata for each sample are provided in **Table S4**.

Results

Theory

The Deconstructed PCR (DePCR) method is based on the polymerase-exonuclease (PEX) PCR method described previously [18]. We previously noted that the first two cycles of PCR are unique in that no amplification of the template is performed. Rather, linear copying of the template nucleic acid prepares the reaction for exponential amplification, starting in the third cycle. In the prior manuscript, linear copying of the original gDNA template was separated from exponential amplification of target copies using exonuclease I (**Figure 8B**). Locus-specific primers containing 5' linker sequences anneal to genomic DNA during two cycles of amplification. Subsequently, exonuclease I was used to remove unused primers from reaction mixtures. Finally, the copied templates were exponentially amplified using primers targeting the 5' linker sequences but not the

source genomic DNA. This approach is viable, but cumbersome due to the need for endonuclease treatment of each sample, and for individual amplification of each sample with primers containing Illumina sequencing adapters and sample-specific barcodes.

We modified the original protocol by including both locus-specific primers containing 5' linkers and primers with Illumina sequencing adapters, sample-specific barcodes, and 3' linkers together in the first linear stage of the reaction (**Figure 8C).** Thus, this approach combines primer sets used in both stage A and B of the PEX PCR method in the same reaction. In addition, four cycles of linear copying are performed, instead of two as in the PEX PCR method (**Figures 8 and 9**). The resulting products are target copies containing Illumina sequencing adapter sequences, sample specific barcodes, linker sequences, and the region of interest. The four cycles of copying serve to prepare the templates for exponential amplification but also (unlike PEX PCR) incorporate a sample-specific barcode so that samples can be pooled and amplified exponentially simultaneously. As with PEX PCR, the linear amplification stage – if operating at 100% efficiency – does not increase the total number of targets from that present in the source template DNA.

After linear copying during the first four cycles, the reactions are pooled and purified to remove unincorporated primers. It is essential for the proper functioning of the method that the primers from the initial stage of the reaction are completely removed; otherwise these locus-specific primers continue to interact with template and amplicons during exponential amplification cycles. We observed that a single cleanup using AMPure XP beads (0.7X) was not sufficient to fully remove all primers; therefore, a double cleanup (*i.e.*, two sequential AMPure XP 0.7X cleanups of the pooled reactions) is performed. The final purified DNA includes a range of DNA types, but only the fragments that contain Illumina sequencing adapters at both ends of the molecule have been generated only through linear copying steps and are available for amplification using Illumina

P5 and P7 primers (**Figure 9**). The entire pool is then used as input template for subsequent amplification using primers consisting of Illumina P5 and P7 sequences. Linear-copied DNA fragments from all samples within the pool, each now containing a sample-specific barcode, are thus subject to exponential amplification simultaneously. One useful feature of this approach is that hundreds of samples can be amplified simultaneously within a single reaction.

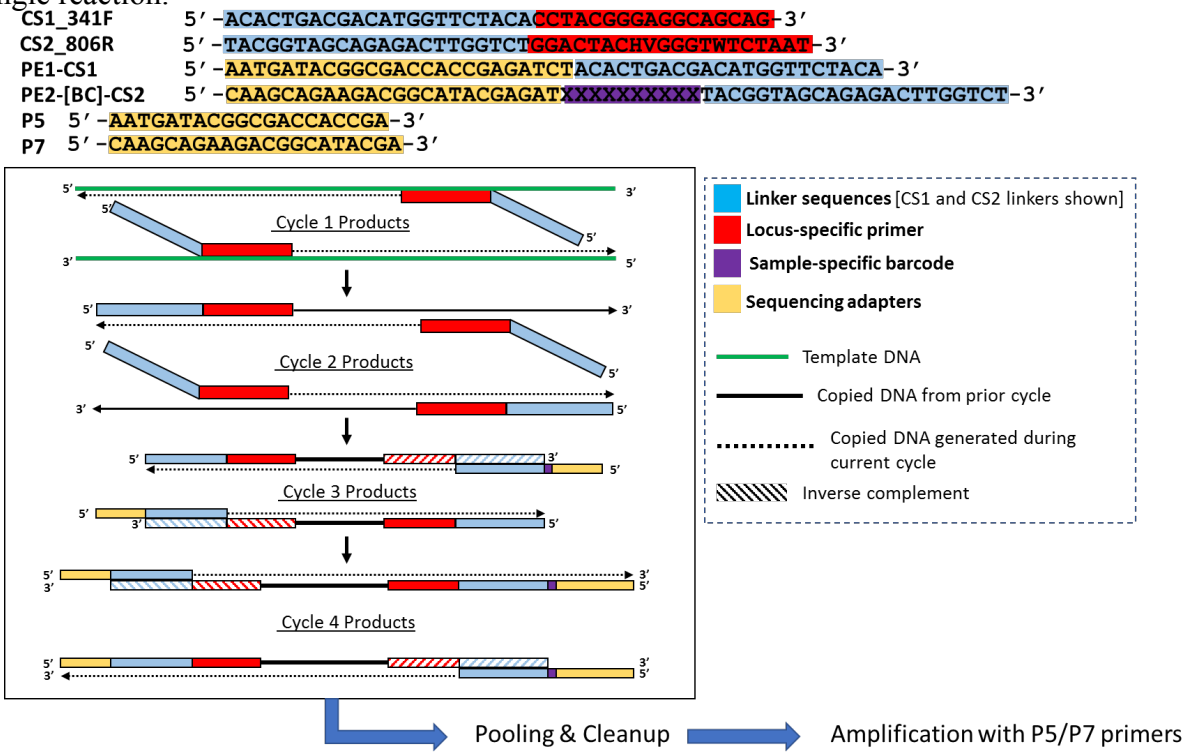


Figure 9: Polymerase-generated intermediates in the first stage of the DePCR workflow

The theoretical advantages of this novel workflow include: (1) the elimination of a separate exonuclease step for each sample, (2) the rapid reduction of many reactions into a single reaction for purification and exponential amplification, and (3) all associated benefits of the prior PEX PCR, in which linear and exponential amplification stages of PCR are isolated from each other and where locus-specific primers are only active for two linear cycles of copying.

Validation of the DePCR method

To assess the effects of amplification method (TAS vs DePCR) and annealing temperature on observed microbial community structure, a single genomic DNA sample was amplified across multiple annealing temperatures using both amplification strategies. Five technical replicates for each condition were performed, and amplicons were sequenced together. The data were analyzed to determine if there were significant differences in sequence metrics (chimera formation), alpha diversity (richness and Shannon index), and observed community structure (beta diversity analyses performed using multi-dimensional scaling and analysis of similarity (ANOSIM). Rates of detectable chimera formation were several orders of magnitude lower with the DePCR pipeline relative to the TAS pipeline, regardless of annealing temperature (**Table 5**). Average chimera detection rate for TAS-processed samples range from 5.16 to 6.53%, while that for DePCR-processed samples ranged from 0.03-0.1%; this difference was significant at all annealing temperatures tested (ANOVA, $P<0.001$). Low rates of detectable chimeras were found in all experiments conducted with DePCR, with averages in the range of 0.01-0.1% (**Table 5**). Conversely, alpha diversity metrics (genus-level richness and Shannon index), were slightly and significantly higher in TAS-based analyses relative to DePCR. Genus-level richness was on average from 1.06-1.21X higher in TAS analyses relative to DePCR, across annealing temperatures from 40˚C to 60˚C (one-way ANOVA; p values ranged from 1.9E-5 to 1.3E-1; **Table 6**). Shannon indices were from 1.03-1.06X higher in TAS analyses relative to DePCR across annealing temperatures from 40˚C to 60˚C (ANOVA; $p<8.13E-4$; **Table 6**).

Table 5: Rates of detectable chimeras in sequence data

Experiment	PCR Method	Annealing Temp. (°C)	Input concentration (ng/reaction)	Chimera detection rate (Average (SD))	ANOVA
Annealing temperature	TAS	40	10	5.16% (0.37%)	1.41E-09
	DePCR	40	10	0.05% (0.03%)	
	TAS	45	10	6.49% (0.29%)	4.05E-11
	DePCR	45	10	0.10% (0.07%)	
	TAS	50	10	6.53% (0.21%)	2.02E-12
	DePCR	50	10	0.04% (0.02%)	
	TAS	55	10	5.69% (0.39%)	9.66E-10
	DePCR	55	10	0.05% (0.02%)	
	TAS	60	10	5.46% (0.49%)	7.56E-09
	DePCR	60	10	0.03% (0.02%)	
Input gDNA concentration	DePCR	50	20	0.05% (0.02%)	5.20E-01
	DePCR	50	10	0.03% (0.03%)	
	DePCR	50	5	0.03% (0.01%)	
	DePCR	50	2.5	0.02% (0.01%)	
	DePCR	50	1.25	0.03% (0.03%)	
Reverse primer variants	TAS	50	10	11.98% (3.85%)	0.00
	DePCR	50	10	0.06% (0.08%)	

Average rates of detectable chimeras are shown for each experiment performed in this study. Significantly lower rates of chimera formation were observed for DePCR-amplified gDNA samples relative to TAS-amplified samples, across multiple annealing temperatures. No significant difference in chimera formation was observed with DePCR methodology with varying gDNA input levels. Significantly higher chimera formation was also observed with TAS relative to DePCR when individual primer variants (RPVs) were utilized. SD = standard deviation.

Table 6: Alpha diversity indices of observed microbial communities

PCR Method	Annealing Temp. (°C)	Shannon Index (Average (SD))	ANOVA	Richness (Average (SD))	ANOVA
TAS	40	2.69 (0.02)	4.76E-05	61.20 (1.92)	1.92E-05
DePCR	40	2.55 (0.03)		50.60 (1.82)	
TAS	45	2.72 (0.03)	5.86E-05	60.60 (2.70)	1.32E-01
DePCR	45	2.59 (0.03)		57.20 (3.63)	
TAS	50	2.74 (0.03)	2.58E-04	64.00 (2.65)	6.56E-02
DePCR	50	2.66 (0.01)		59.60 (3.78)	
TAS	55	2.72 (0.02)	8.13E-04	62.00 (1.87)	2.98E-02
DePCR	55	2.64 (0.03)		58.60 (2.19)	
TAS	60	2.72 (0.01)	6.16E-04	60.60 (2.70)	3.31E-02
DePCR	60	2.63 (0.03)		56.60 (2.19)	

Shannon indices were calculated at the taxonomic levels of genus for all samples amplified using TAS and DePCR methodologies across five annealing temperatures of 40°, 45°, 50°, 55° and 60°C. Datasets were rarefied to 4,500 sequences/sample. For each methodology and annealing temperature, an average and standard deviation of five technical replicates is shown. At all temperatures, TAS-amplified samples had higher Shannon indices relative to DePCR-amplified samples. SD = standard deviation.

A strong, significant effect of annealing temperature on the observed microbial community structure was seen in both TAS and DePCR amplification methods (**Figure 10A**).

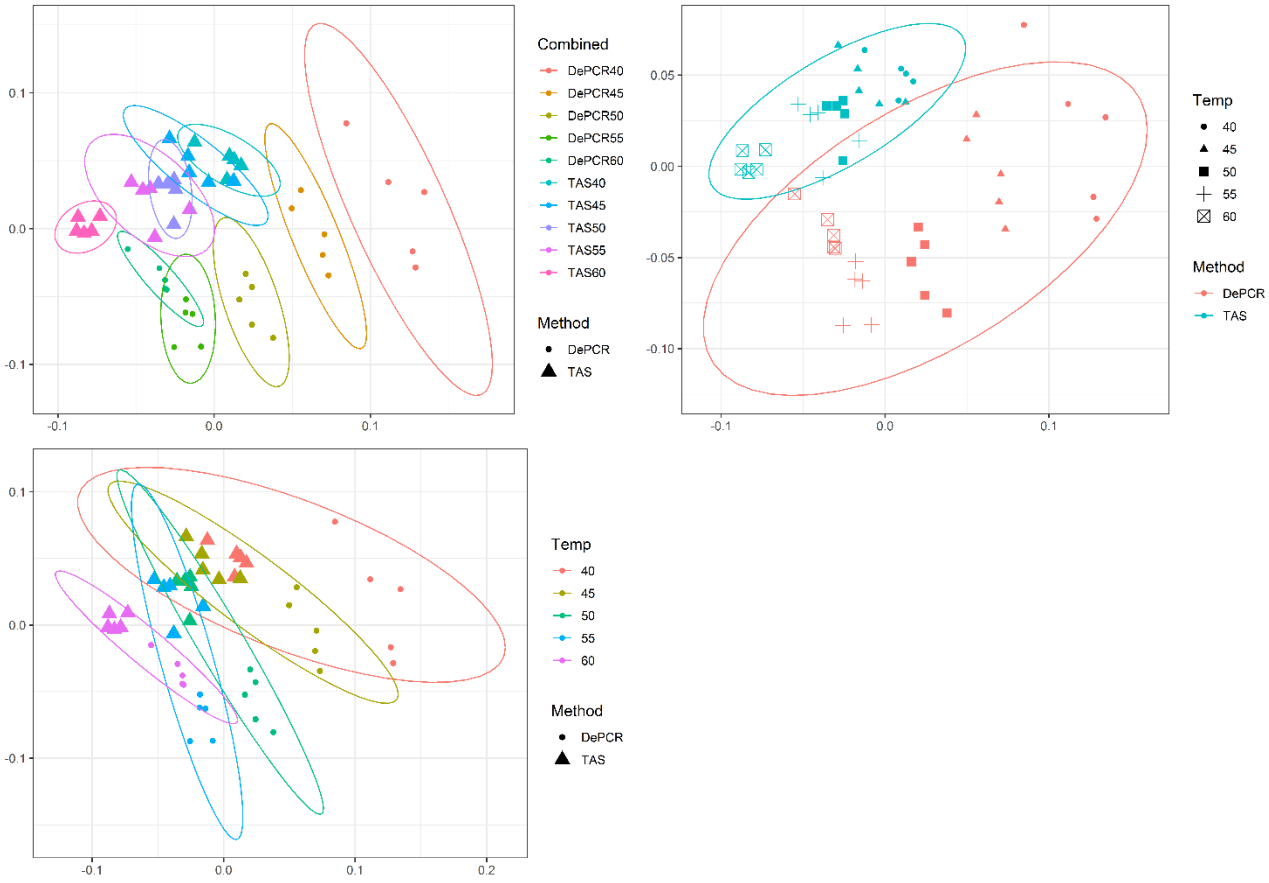


Figure 10: Effect of PCR methodology and annealing temperature on observed microbial communities

Genus-level abundance data were visualized using metric MDS (mMDS) ordination employing a distance matrix based on Bray-Curtis similarity. For each PCR condition (TAS or DePCR), five technical replicates were analyzed using annealing temperatures of 40°, 45°, 50°, 55° or 60° Celsius. Ellipses represent 95% confidence intervals around centroids. Rarefaction was performed to a depth of 4,500 sequences per sample. Observed community structure was significantly different across **(A)** all combinations of temperature and method (one-way ANOSIM Global R=0.713; P=0.0001); **(B)** temperature (two-way ANOSIM R=0.832; p=0.0001), and (**C**) amplification method (two-way ANOSIM R=0.988; P=0.0001).

Although the overall scale of difference between TAS and DePCR was modest (maximum Bray-Curtis dissimilarity between samples = 0.23 between a TAS sample with 60˚C annealing temperature and a DePCR sample with 40˚C), there was a significant effect of amplification method on observed microbial community at all temperatures. Two-way ANOSIM analyses indicated significant differences by temperature across methods (R=0.832; p=0.0001; **Figure 10B**), and by amplification method across temperatures (R=0.988; p=0.0001; **Figure 10C**). Similar trends were observed for increases in annealing temperature in both methods, with temperature loading primarily on MDS axis 1. As previously noted [18], greater variability in observed microbial community structure was noted with DePCR with low annealing temperature, particularly at 40˚C (**Figure 10A**).

One key feature of the DePCR methodology is the ability to determine which primers in a degenerate pool are interacting with the source genomic DNA. This is achieved as the exponential amplification of the template is performed using primers targeting Illumina sequencing adapters and not the locus-specific primers (**Figures 8, 9**). Locus-specific primers only interact with the gDNA and the first linear copies of gDNA during the first two cycles of the DePCR method. These primer sequences are retained during exponential amplification with primers targeting linker sequences. Conversely, in standard PCR, the locus-specific primers interact with both the genomic DNA template and with copies made from the genomic DNA during exponential amplification; thus, information regarding primer-gDNA template interactions are lost [18]. We thus examined the so-called "primer utilization profiles" (PUPs) for these reactions (**Figure 11**). The relative frequency of each of the 18 unique primer variants is shown for each replicate at each PCR condition (temperature x method). Standard PCR amplification protocol (TAS) removes primer-template interaction information as primer-amplicon interactions throughout the amplification

reaction tolerate mismatches; all 18 primer variants are used at similar frequencies, regardless of annealing temperature (**Figure 11A**).

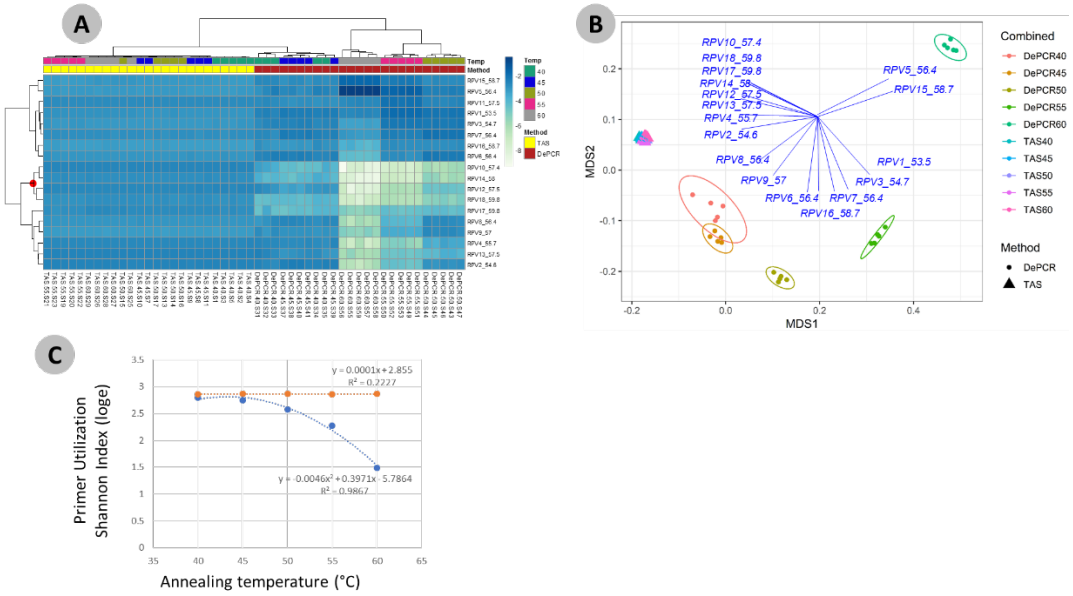


Figure 11: Effect of annealing temperature and amplification methodology on primer utilization profiles (PUPs)

(A) Two-way clustered heatmap of log-transformed primer variant utilization during amplification of fecal genomic DNA. Samples (columns) are color-coded by amplification method (TAS or DePCR) and amplification annealing temperature (40°, 45°, 50°, 55° and 60°C), with five technical replicates per condition and rarefaction to 1,800 sequences/sample. Primers (rows) are clustered by profile similarity across all samples and represent all 18 primer variants (RPV1 – RPV18) present in the 806R degenerate primer pool. Theoretical melting temperatures for each primer are shown adjacent to primer name. **(B)** mMDS ordination of PUPs based on Bray-Curtis similarity. Vectors represent Pearson correlations (>0.9) for each primer variant. Ellipses represent 95% confidence intervals around centroids for DePCR amplification reactions. Five technical replicates per condition were generated and for each sample, rarefaction was

performed to 1,800 sequences. (**C**) Regression analysis was performed was performed on average Shannon index values for primer utilization for each methodology (TAS and DePCR) across annealing temperature. A very small effect of annealing temperature on primer utilization evenness was observed in TAS (orange line). A negative quadratic relationship was observed between annealing temperature and primer utilization evenness in DePCR (blue line). Analyses were based on five technical replicates rarefied to 4,500 sequences per sample.

Some patterning is observed in the TAS method, but overall diversity of primer utilization is extremely high and only small differences were observed between temperatures of 40-60˚C (**Figure 11B**). The average Shannon index for PUP profiles of TAS samples across all annealing temperatures was 2.859-2.864; the maximum possible natural log Shannon index for 18 features is 2.890. This PUP diversity profiling demonstrates that for standard TAS PCR, the primers used in copying throughout the amplification reaction are not dependent on annealing temperature.

Conversely, a strong effect of annealing temperature is observed on the PUP of samples amplified using the DePCR protocol (**Figure 11A, B**). A shift in PUP patterning is observed with increasing annealing temperature, and at 60˚C two primer variants (RPV5 and RPV15) dominate. At lower annealing temperatures, a broader range of primers are utilized in the initial stages of gDNA copying. The relationship between annealing temperature and primer utilization richness (here represented as the Shannon index) was best fit with a polynomial equation and is shown in **Figure 11C**. As annealing temperature increases, fewer and fewer primer variants interact with the gDNA template. Conversely, at the lowest tested annealing temperature of 40˚C, the Shannon index of the DePCR amplicons nearly matched that of the TAS. Several primer variants, however, including RPVs 10, 12, 14 and 18, were poorly utilized in DePCR amplifications regardless of

annealing temperature (**Figure 11A**). These four variants included variants with high melting temperatures (57.4, 57.5, 58 and 59.8˚C), while the two most utilized RPVs at PCR annealing temperatures of 60˚C had moderate to high annealing temperatures (56.4 and 58.7˚C). Thus, the melting temperature of the primer did not directly correlate with utilization at different PCR annealing temperatures in this system. The observed primer utilization profiles represent a template-specific phenomenon, and different PUPs would be recovered with different DNA templates.

Determination of linearity in DePCR amplification

In the DePCR protocol, after four initial cycles of linear copying during the first stage of DePCR, samples are pooled prior to purification and second stage amplification with Illumina P5 and P7 primers. The pooling of samples can only be performed because of the incorporation of a sample-specific unique barcode for each sample during the first stage. During the second stage amplification, primers target the Illumina adapters are used for amplification, and all templates from all samples are amplified simultaneously (**Figure 8C**). Since there is no opportunity for primer-template bias during the second stage (*i.e.,* **Stage B of Figure 8C**) of amplification (all amplifiable template molecules contain Illumina sequencing adapters) and primers are non-degenerate, the relative abundance of template molecules from a single sample within the pool should be maintained during amplification. To determine if the relative abundance of template molecules from each sample was maintained in the DePCR protocol, we performed an experiment in which input gDNA (feces) was varied from 1.25 ng to 20 ng per 10 µl reaction. All input levels were performed with five technical replicates. After the first stage (4 cycles) of the DePCR, all replicates from all gDNA input levels were pooled in equal volume and purified.

The purified product was then amplified with P5 and P7 primers, and the final pool sequenced. We first assessed whether the input DNA concentration was correlated with the total number of reads generated using this approach (**Figure S6**). Since all samples were amplified together, and low input DNA samples should theoretically provide fewer molecules to the combined pool, we hypothesized that a linear relationship should exist between input DNA in the first stage and the number of reads generated per sample. A significant positive correlation between input gDNA concentration and absolute number of reads recovered from each sample was observed, though substantial variability at each input concentration was observed (R^2=0.58, **Figure S6C**). We also sought to determine if the input gDNA concentration from the same sample had a significant effect on the observed microbial community structure. Although there was a positive correlation between input gDNA and total number of sequences recovered, we observed no significant effect of input gDNA on the microbial community structure (**Figure S6A;** Global ANOSIM R=-0.034; p=0.79**)**. Similarly, no significant difference in primer utilization was observed with different gDNA input concentrations (**Figure S6B**). Thus, increasing input gDNA concentration alters the number of molecules passing to the second stage of the DePCR reaction, but within the observed concentration range does not affect the primer utilization profile or final observed microbial community structure.

Assessing the effect of individual primers in a degenerate primer pool

Degenerate primer pools are generally used to amplify genomic DNA, although not all primers actively interact with the source gDNA (**Figure 11A**). This degenerate mixture of primers is employed to target a broad range of taxa, and the presence of additional primer variants in pools has been shown to improve detection of known microbial lineages [46, 52, 82, 83]. In

standard PCR, all primers do eventually interact with amplified copies of gDNA during the many cycles of exponential amplification; however, many primers do not interact with the source genomic DNA due to preferential annealing of other primers (**Figure 11A**). We sought, therefore, to determine how much microbial diversity could be detected using each primer variant independently in PCR reactions using both the TAS and DePCR methods. In addition, we sought to determine how the observed microbial community structure differed by single primer variant usage. We hypothesized that the single primer variant PCR would better approximate degenerate primer pools when using the DePCR method relative to the TAS method, as our prior work showed that a deconstructed PCR approach was more tolerant of mismatches between primer and gDNA template than TAS [18]. The tolerance of mismatches may lead to better capture of microbial community diversity when a greater number of mismatches between primer and template are present, as is expected in a single primer PCR. To explore this, we PCR-amplified a single gDNA template (feces) with the 18 unique reverse primer variants (RPVs) from the degenerate primer pool. Each reaction was performed in technical duplicates, and each reaction was performed using the DePCR and the TAS method. Three RPVs from the TAS method were removed from the analysis due to pipetting error, as determined by primer utilization profiles. These included one replicate of RPV5 and both replicates of RPV15 (**Table S4**). We compared alpha and beta diversity analyses of the PCRs employing 15-18 unique RPVs to those generated with the fully degenerate primer set. All alpha and beta diversity analyses were performed on data rarefied to a depth of 1800 sequences/sample (**Table S4 – experiment 3**).

When employing fully degenerate primer pools, observed alpha diversity (Shannon index) of the fecal sample was slightly, but significantly higher when analyzed using the TAS protocol relative to the DePCR protocol (average Shannon index, five replicates, 2.71 to 2.66;

ANOVA P<0.001; **Table 7**). We then calculated average Shannon indices for analyses of the same gDNA sample with individual RPVs, employing TAS and DePCR protocols. The average Shannon index for the TAS reactions with unique RPVs (2.40) was significantly lower than that measured for the DePCR reactions (2.58) (ANOVA P<0.001; **Table 7**).

Table 7: Effects of amplification method and reverse primer variants on observed microbial community alpha diversity

Comparison	# replicates analyzed	Average Shannon Index [43]SD), TAS	Average Shannon Index [43], DePCR	ANOVA
Amplification with 18-fold degenerate primer pools	5	2.71 (0.03)	2.66 (0.04)	3.14E-05
Amplification with each RPV independently	33 (TAS) or 36 (DePCR)	2.4 (0.01)	2.58 (0.21)	5.95E-05
Summation of independent RPVs and re-rarefaction to 1800 sequences (5x)	5	2.48 (0.03)	2.69 (0.02)	7.40E-07
	ANOVA	3.69E-08	3.77E-01	

Fecal gDNA was PCR amplified with 18-fold degenerate reverse primer pools (5 technical replicates), and with each unique reverse primer variant (RPV; 2 technical replicates). Data sets were rarefied to 1,800 sequences per sample, and Shannon indices (loge) were calculated. When using fully degenerate primer pools, average Shannon index was significantly higher for TAS methodology relative to DePCR methodology. When data from all reactions with individual RPVs were analyzed, average Shannon index was significantly lower for TAS methodology relative to DePCR methodology. Data from RPVs (1,800 sequences/sample) were pooled and re-rarefied to 1,800 sequences (5 repetitions), and the resulting average Shannon index was

significantly lower for the TAS methodology relative to DePCR methodology. Different approaches with the DePCR method did not generate significantly different Shannon indices (ANOVA P=0.377), while the same approaches generated significantly different Shannon indices (ANOVA P<0.001).

Finally, all RPV data, rarefied to 1800 sequences per sample, was pooled together for TAS and DePCR approaches, independently. These combined datasets were then randomly sub-sampled to 1800 sequences. These rarefactions were performed five times, and the average Shannon index for the combined RPVs was calculated. In this approach, average Shannon index from TAS (2.48) was significantly lower than for DePCR (2.69) (ANOVA P<0.001; **Table 7**). Across all three methods of calculating observed diversity, there was no significant different in measured Shannon index for the DePCR method (ANOVA, P=0.377), while a significant decrease with each RPV independently was observed with the TAS method (ANOVA, P=3.69e-8). When each RPV is used independently in the TAS protocol, the overall captured diversity is lower than with reactions with degenerate pools (**Table 7)** due to the greater number of potential mismatch interactions that can occur when a complex template is amplified with a single, non-degenerate primer. As the DePCR method is more tolerant of mismatches, no significant decrease in average Shannon index was observed. However, the observed variance in Shannon index among the individual RPVs was greater for the DePCR than for the TAS method **(Table 7**).

We next examined the structure of the observed fecal microbial communities in standard TAS and DePCR with degenerate primer pools, and with reactions conducted using RPVs (**Figure 12**).

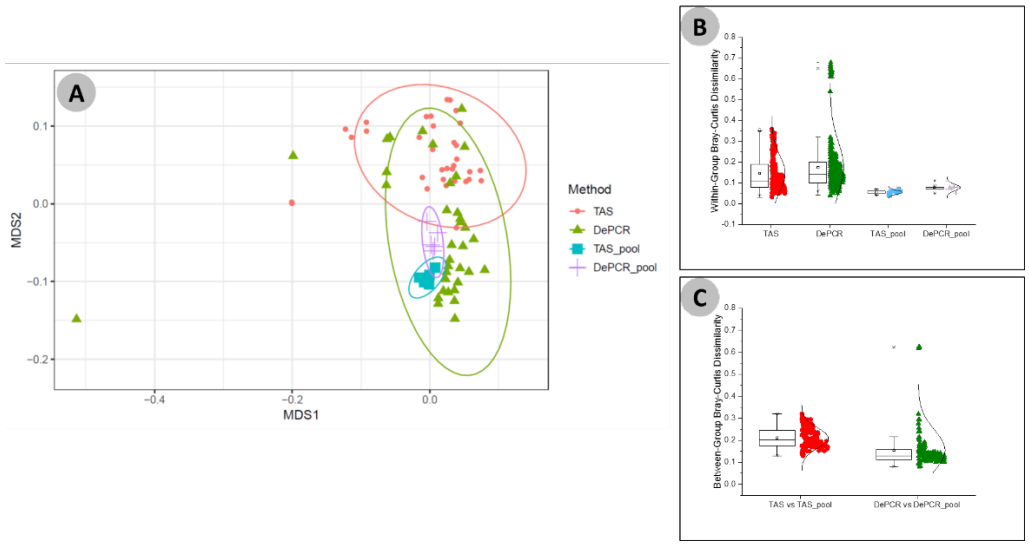


Figure 12: Microbial community structure revealed using individual primer variants with TAS and DePCR amplification methodologies

(**A**) Fecal gDNA was amplified using the 341F primer with18 unique 806R reverse primer variants (RPVs) under standard PCR (TAS) and DePCR workflows. Three RPVs were removed from the TAS analysis due to pipetting error, as described in the text. Genus-level biological observation matrices (BIOMs) were visualized using mMDS. Each amplification with a unique RPV was performed in technical duplicate, and five technical replicates were generated using degenerate primer pools (TAS_pool or DePCR_pool). All samples were rarefied to 1,800 sequences. Ellipses represent 95% confidence intervals around centroids. TAS profiles generated with RPVs were significantly distinct from TAS profiles generated with degenerate primer pools (ANOSIM R=0.487; P=0.003). DePCR profiles generated with RPVs were not significantly distinct from DePCR profiles generated with degenerate primer pools (ANOSIM R=-0.306; P=0.99). **(B)** Within-group Bray-Curtis dissimilarity distributions for profiles generated with RPVs and with degenerate pools. **(C)** Between-group Bray-Curtis dissimilarity distributions for observed microbial community structure generated with RPVs and with degenerate primer pools.

Average dissimilarity among TAS_pool and TAS RPV profiles (0.211) was greater than for DePCR_pool and DePCR RPV profiles (0.154) (ANOVA P<0.001).

We observed high reproducibility for five replicates using TAS (*i.e.*, 'TAS_pool') or DePCR (*i.e.*, 'DePCR_pool') with degenerate primer pools (**Figure 12A, B**) and observed microbial community structure was significantly different between TAS and DePCR employing the degenerate primer pools (ANOSIM, R=0.401, p=0.001). Compared to amplifications with degenerate pools of primers, within-group variability was much greater for the analyses of RPVs individually with either amplification protocol (**Figure 12A, B, 'TAS' and 'DePCR'**). Within-group Bray-Curtis dissimilarity (BCD) of amplicons from the 15 (TAS) to 18 (DePCR) RPVs ranged from 0.03 to 0.36 for the TAS method and from 0.04 to 0.68 for the DePCR method (ANOVA P<0.001; **Figure 12B**). Conversely, the within-group BCD for five technical replicates generated with degenerate primer pools were 0.04 to 0.07 for TAS and 0.05 to 0.11 for DePCR (ANOVA P<0.001). Profiles of the individual RPVs from DePCR analyses could be divided into two groups: (a) RPVs with profiles highly similar to degenerate primer pool analysis with either DePCR or TAS; and (b) RPVs with profiles divergent from the degenerate pool communities, and more similar to RPVs from TAS amplification reactions. Overall, the observed microbial community structure generated using the DePCR method with RPVs and with degenerate pools was not significantly different (ANOSIM R=-0.306, p=0.99). Conversely, the observed microbial community structure generated using RPVs was significantly different that that observed with degenerate primer pools for the TAS method (ANOSIM R=0.487; p=0.003). Average BCD between TAS_pool and TAS RPV profiles (0.211) was significantly greater than for DePCR_pool and DePCR RPV (0.154) (ANOVA P<0.001; **Figure 12C**). DePCR BCD profiles

were heavily weighted toward low dissimilarity, with a long tail of high dissimilarity comparisons. The long tail is a result of some primers generating highly divergent observed microbial communities with the DePCR protocol. Many of the primers which showed the poorest utilization within the degenerate pool (*e.g.*, RPV10, 12, 14, and 18; **node with red dot in Figure 11A**), generated the most divergent single RPV profiles. This suggests that these primers do not closely match the most dominant taxa within this particular gDNA sample.

Discussion

We demonstrate here an updated protocol for the Deconstructed PCR methodology [18] which reduces the overall complexity of the workflow and increases the throughput. Complete removal of 1st stage (or "Stage A") primers (locus-specific primers containing 5' overhanging linkers) is essential for the effectiveness of the DePCR protocol, and we have replaced the exonuclease step with a bead-based magnetic cleanup. The new method improves throughput by generating barcoded DNA fragments through 4 cycles of linear amplification; thus, all samples can be pooled before bead-cleanup. This reduces workflow complexity and cost, while retaining the essential features of the DePCR reaction. Complete removal of primers is difficult to measure directly, however; thus, the primer utilization profiles (PUPs) are the clearest indication of successful removal of locus-specific degenerate primers from the first stage of the reaction. With standard PCR, no true signal is obtained from the PUPs, as primer-amplicon interactions during late cycles generates a 'scrambled' signal due to mismatch interaction with amplicons present at high abundance. In DePCR, a PUP signal can be obtained as locus-specific primers only interact with the gDNA template and linear copies during the first two cycles of PCR. Subsequently, all exponential amplification is performed using conserved sequences that are not present in the

source gDNA. In this way, the primer sequences interacting with the source gDNA are 'fossilized' and can be interrogated directly. When using this approach, we observed strong effects of annealing temperature on primer-gDNA template interactions, with a negative quadratic correlation between annealing temperature and evenness of primer utilization. At highest annealing temperatures, very few primers from the primer pool anneal to the gDNA template, and this leads to a shift in the sequences that are amplified by PCR with a result of significantly different observed microbial communities. We note that the elevated annealing temperature by itself does not select for primer variants with the highest theoretical melting temperature. Rather, primer variants, presumably template-specific, are favored regardless of their melting temperature.

A surprising benefit to the DePCR methodology is the reduced rate of chimera formation. Chimeras are artifactual hybrid sequences generated from two or more templates due to incomplete polymerase extension during PCR, and their presence can be difficult to detect and lead to overestimation of diversity and alteration of observed microbial community structure [36, 84, 85]. Input genomic DNA concentration and target microbial community complexity have been identified as contributors [86, 87]. We previously observed that chimera formation was correlated with total number of PCR cycles in both first and second stages of PCR [39], and this has been reported elsewhere in many studies [34, 85, 86]. As many factors can contribute to chimera formation, various solutions have been proposed, including reducing input gDNA concentration [26], reducing PCR cycles [28, 88], and employing highly processive enzymes [86], among others. In this study, we have observed that the use of the DePCR methodology can dramatically and significantly lower rates of observed chimeras resulting in rates that were generally below 0.1%. These low rates of chimera formation were observed across all annealing

temperatures and input template concentrations tested. The reasons for the dramatic decrease in chimera formation rate with the DePCR method are likely a result of: (a) reduction in input DNA concentration for exponential amplification due to the double-purification step, (b) higher annealing temperature for the exponential amplification due to targeting of P5/P7 Illumina adapters –potentially reducing the re-annealing of PCR products to other products, and (c) long elongation times during the first cycles, reducing the formation of incomplete molecules during the first stages of PCR. Conceivably, chimera formation with DePCR could be reduced further; we performed 30 cycles of amplification to generate robust PCR yields for sequencing. However, the amplification of the pool of amplicons during the second stage PCR could be titrated across different numbers of cycles, and the reaction with the fewest numbers of cycles yielding sufficient DNA for sequencing could be employed. It is critical to remember that the rate of chimera formation represents only the rate of *detectable* chimera formation, and that chimeras generated from closely related sequences are not only likely to occur at higher rates [34] but are also essentially undetectable by chimera detection software. We note that in this study, amplification of fecal gDNA with degenerate primer pools resulted in higher observed diversity with the TAS method relative to the same sample amplified with the DePCR protocol (**Table 7**), and this could represent the residual presence of chimeras that were not removed.

Suzuki and Giovannoni [28] previously modeled PCR reactions with mixed templates by incorporating efficiency parameters into equations estimating molarity of amplicon yield. They further estimated second-order kinetics wherein changes in the concentration of specific PCR products alter efficiencies during the amplification, including through inhibition of amplification by competition between primers and amplicons for annealing locations. With increasing cycle number, reaction efficiency dropped dramatically. The DePCR method theoretically circumvents

at least some of these issues. First, since locus-specific primers interact with template only during two cycles of copying (linear only), any differences in amplification efficiency of templates are limited to those two cycles. Subsequently, all templates are amplified with primers targeting sequences common to all amplifiable templates. Suzuki and Giovannoni [28] showed that even a relatively high amplification efficiency could lead to dramatic distortion of the underlying template ratios within 10-15 cycles. In DePCR approaches, amplification efficiency is expected to be lowest during the first two cycles – when primers anneal to gDNA templates with varying numbers of mismatches – and then higher during the remaining cycles as amplification is performed with perfectly matching primers. We also note that in PCRs with degenerate primers, each primer variant is present at a low concentration (total primer concentration / number of variants); in the 2nd stage of the DePCR protocol, a non-degenerate primer at a high concentration relative to each variant is used for amplification. Thus, DePCR limits the number of cycles operating at low primer efficiency and uses high-efficiency reactions to perform exponential amplification. Degenerate locus-specific primer interactions with PCR amplicons are also removed, thereby removing additional variable efficiency annealing steps from the PCR.

We previously demonstrated that a deconstructed PCR approach could help overcome PCR distortions due to mismatches between primers and templates in a mock community [18], and we believe this is in part due to the circumventing of multiple cycles with low amplification efficiency. Single mismatches between templates and primers can substantially alter observed microbial community structure, and indeed, many modifications to degenerate primer pools are performed to increase degeneracy by adding single variants targeting specific microbial taxa [36]. In this study, we independently used each primer variant in a degenerate primer pool both

to examine the potential for each primer to amplify a complex microbial gDNA template and to assess the ability of the DePCR protocol to enable single non-degenerate primers to broadly amplify microbial taxa with mismatches. We observed that while the observed microbial community structure varied widely using non-degenerate primer variants (both TAS and DePCR), many single non-degenerate primer variants were able to generate reasonable approximations of the microbial community structure as revealed through amplification reactions with degenerate primer pools, thus indicating that the DePCR approach can be used with complex microbial samples to improve tolerance of mismatches. This suggests that a more empirical approach to primer design can be taken by using the DePCR method to reduce the complexity of degenerate primer pools or enable broader target range of highly degenerate primer pools targeting functional genes. Primer utilization profiling can in turn be used to provide empirical evidence demonstrating which primers within the degenerate primer pool are interacting with unknown templates. The inclusion of non-essential variants decreases the concentration of all other primers in a primer pool, and removal of unneeded primer variants may be beneficial. However, when using the same primer set for a broad range of complex genomic DNA samples from different environments, we expect that the 'essential' primers will vary from system to system.

We can recommend the DePCR protocol for reactions where degenerate primer pools are used or for primer-template systems where mismatches are possible or expected. Several caveats, however, should be considered. First, the method is not recommended for reactions requiring stringent PCR conditions. Second, since reactions are pooled together after the first linear cycles and then amplified, the reactions are sensitive to the relative number of copies within each sample. As observed in **Figure S6C**, there is a linear response between input gDNA and number

of sequences generated. Thus, input gDNA concentration of similar samples should be carefully controlled to avoid large variance in number of sequences generated per sample. Furthermore, different sample types should be amplified independently, as different samples may have a different density of targets per ng of DNA, leading to further variance in sequence reads generated. Third, in the updated DePCR protocol where Illumina P5 and P7 primer are used, polymerase extension copies through the DNA region containing the sample-specific barcode and can introduce errors. In this study, we employed Fluidigm Access Array primers which contain 10-base barcodes with a Hamming distance of 3 (each barcode has at least 3 mismatches with all other barcodes), and this large Hamming distance should limit mis-assignment of reads. However, with other barcoding systems, or with very high PCR cycle or error-prone polymerases, this source of error could lead to cross-signaling between samples or loss of reads. Finally, we note that when assessing if a DePCR protocol is functioning properly, it is important to employ an analysis of primer utilization across a temperature gradient analysis with standard (TAS) and DePCR workflows. In standard PCR, a small or no effect of temperature should be observed on the PUPs, while a strong shift in primer utilization should be observed with the DePCR protocol. Since primer utilization with DePCR can be extremely broad at low annealing temperatures, it can be difficult to differentiate between a properly operating or failed DePCR protocol without the temperature gradient analysis.

Supporting Information

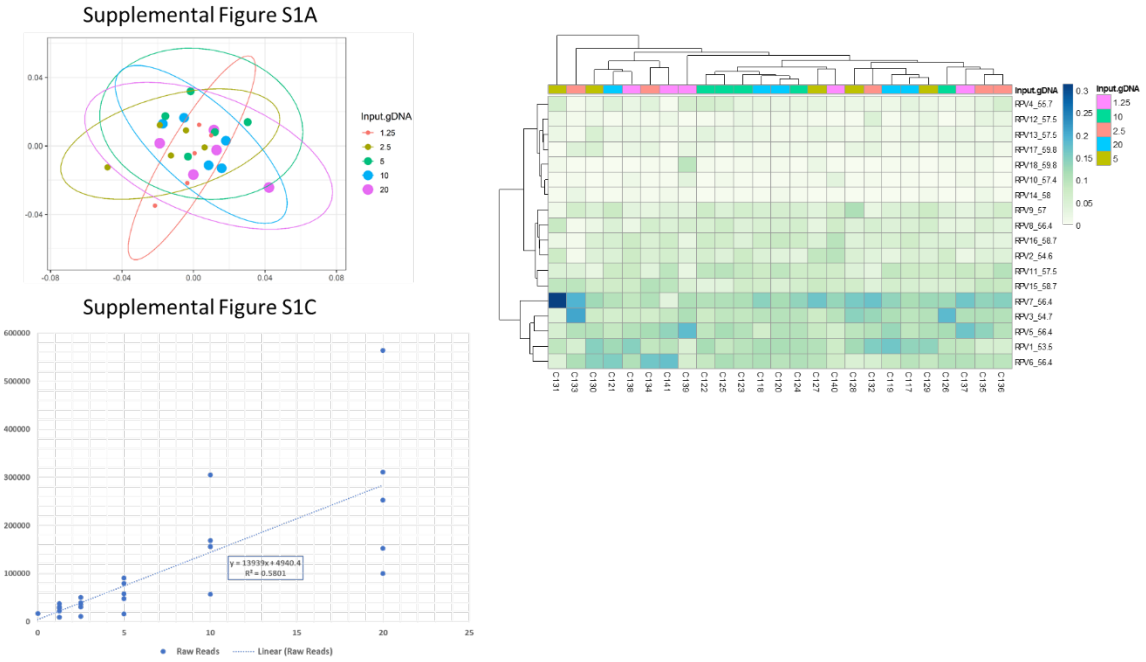


Figure S6. Effect of input gDNA template concentration on microbial community composition and PUPs using DePCR.

Analyses were performed on rarefied data sets (8,000 sequences per sample), with five technical replicates for each DNA input level (1.25, 2.5, 5, 10 or 20 ng/µl). (**A**) Genus-level mMDS ordination of microbial community structure using a distance matrix based on Bray-Curtis similarity. No significant differences were observed between all the concentrations (Global ANOSIM: R=-0.03376, p=0.79). Ellipses represent a 95% confidence interval around the centroid. (**B**) Primer utilization profiles for all primer variants (RPV1 – RPV18), visualized as a heatmap. (**C**) A positive correlation between input gDNA (1.25, 2.5, 5, 10, 20 ng/µl) and sequence yield was observed. For all input levels, the same gDNA template was used with five technical replicates. All samples were pooled after stage A of DePCR and amplified together using Illumina P5 and P7 primers. Data were rarefied to 8,000 sequences per sample.

Table S4: Mapping file metadata associated with all samples used in this study

Exp	Sample	PCR Method	Primer Variant	Annealing Temp	Barcode Name	Barcode Sequence	Input DNA	gDNA Conc (ng/ul)	Raw Reads	Merged	Trimmed	Post Chimera Removal	Chimeras Removed	Chimera Removed (%)	Shannon (Genus)	Simpson (Genus)	Richness (Genus)	Evenness (Genus)	Rarefaction	Reference Figures / Tables
1	C1.40.S1	TAS	Pool	40	FLD0001	GTATCGTCGT	Chin. Feces	10	103388	101213	87769	82726	5043	5.75%	2.70	0.88	59	0.66	4500	10, 11, T7, T5
1	C2.40.S2	TAS	Pool	40	FLD0002	GTGTATGCGT	Chin. Feces	10	172406	168166	145598	138284	7314	5.02%	2.66	0.87	62	0.65	4500	10, 11, T7, T5
1	C3.40.S3	TAS	Pool	40	FLD0003	TGCTCGTAGT	Chin. Feces	10	124807	122221	106435	100786	5649	5.31%	2.73	0.89	60	0.67	4500	10, 11, T7, T5
1	C4.40.S4	TAS	Pool	40	FLD0004	GTCGTCGTCT	Chin. Feces	10	115690	113027	97821	93034	4787	4.89%	2.68	0.88	64	0.64	4500	10, 11, T7, T5
1	C5.40.S5	TAS	Pool	40	FLD0005	GTGCGTGTGT	Chin. Feces	10	112168	109502	95164	90554	4610	4.84%	2.70	0.88	61	0.66	4500	10, 11, T7, T5
1	C7.45.S7	TAS	Pool	45	FLD0007	GTCGTGTACT	Chin. Feces	10	140264	137108	118939	110739	8200	6.89%	2.68	0.87	64	0.64	4500	10, 11, T7, T5
1	C8.45.S8	TAS	Pool	45	FLD0008	GATGTAGCGT	Chin. Feces	10	149816	146093	126440	118066	8374	6.62%	2.74	0.88	57	0.68	4500	10, 11, T7, T5
1	C9.45.S9	TAS	Pool	45	FLD0009	GAGTGATCGT	Chin. Feces	10	135480	132419	115597	108290	7307	6.32%	2.70	0.88	61	0.66	4500	10, 11, T7, T5
1	C10.45.S10	TAS	Pool	45	FLD0010	CGCTATCAGT	Chin. Feces	10	143235	138419	117398	109798	7600	6.47%	2.73	0.88	62	0.66	4500	10, 11, T7, T5
1	C11.45.S11	TAS	Pool	45	FLD0011	CGCTGTAGTC	Chin. Feces	10	131977	127713	109495	102777	6718	6.14%	2.73	0.89	59	0.67	4500	10, 11, T7, T5
1	C13.50.S13	TAS	Pool	50	FLD0013	GAGCTAGTGA	Chin. Feces	10	103192	100268	86023	80162	5861	6.81%	2.77	0.89	63	0.67	(4500), [1800]	(10, 11, T7, T5), [12, T6]
1	C14.50.S14	TAS	Pool	50	FLD0014	CGTGCTGTCA	Chin. Feces	10	124293	120456	103252	96391	6861	6.64%	2.74	0.88	60	0.67	(4500), [1800]	(10, 11, T7, T5), [12, T6]
1	C15.50.S15	TAS	Pool	50	FLD0015	GATCGTCTCT	Chin. Feces	10	122686	119885	103189	96495	6694	6.49%	2.71	0.88	65	0.65	(4500), [1800]	(10, 11, T7, T5), [12, T6]
1	C16.50.S16	TAS	Pool	50	FLD0016	GTGCTGTCGT	Chin. Feces	10	129109	125996	109380	102335	7045	6.44%	2.76	0.88	65	0.66	(4500), [1800]	(10, 11, T7, T5), [12, T6]
1	C17.50.S17	TAS	Pool	50	FLD0017	TGAGCGTGCT	Chin. Feces	10	128078	125392	110177	103273	6904	6.27%	2.73	0.88	67	0.65	(4500), [1800]	(10, 11, T7, T5), [12, T6]
1	C19.55.S19	TAS	Pool	55	FLD0019	TCAGTGTCTC	Chin. Feces	10	134517	131823	114976	108364	6612	5.75%	2.69	0.87	63	0.65	4500	10, 11, T7, T5
1	C20.55.S20	TAS	Pool	55	FLD0020	GTGCTCATGT	Chin. Feces	10	169782	165945	143739	134636	9103	6.33%	2.75	0.88	62	0.67	4500	10, 11, T7, T5
1	C21.55.S21	TAS	Pool	55	FLD0021	CGTATCTCGA	Chin. Feces	10	145938	141502	120776	114277	6499	5.38%	2.71	0.88	59	0.66	4500	10, 11, T7, T5
1	C22.55.S22	TAS	Pool	55	FLD0022	GTCATGCGTC	Chin. Feces	10	137078	133709	115719	109243	6476	5.60%	2.73	0.88	62	0.66	4500	10, 11, T7, T5
1	C23.55.S23	TAS	Pool	55	FLD0023	CTATGCGATC	Chin. Feces	10	121829	118565	101340	95884	5456	5.38%	2.74	0.88	64	0.66	4500	10, 11, T7, T5
1	C25.60.S25	TAS	Pool	60	FLD0025	TGTGTGCATG	Chin. Feces	10	144519	141449	122705	115500	7205	5.87%	2.72	0.88	59	0.67	4500	10, 11, T7, T5
1	C26.60.S26	TAS	Pool	60	FLD0026	GAGTGTCACT	Chin. Feces	10	131595	128331	111540	105606	5934	5.32%	2.72	0.88	58	0.67	4500	10, 11, T7, T5
1	C27.60.S27	TAS	Pool	60	FLD0027	CTAGTCTCGT	Chin. Feces	10	134892	131539	112464	106098	6366	5.66%	2.71	0.88	59	0.67	4500	10, 11, T7, T5
1	C28.60.S28	TAS	Pool	60	FLD0028	GAGTGCATCT	Chin. Feces	10	157453	153935	133200	125493	7707	5.79%	2.75	0.88	64	0.66	4500	10, 11, T7, T5
1	C29.60.S29	TAS	Pool	60	FLD0029	TGCGTAGTCG	Chin. Feces	10	124015	121021	104802	99909	4893	4.67%	2.72	0.88	63	0.66	4500	10, 11, T7, T5
1	C31.40.S31	DePCR	Pool	40	FLD0031	CTGTAGTGCG	Chin. Feces	10	114122	71766	4653	4652	1	0.02%	2.53	0.86	51	0.64	4500	10, 11, T7, T5
1	C32.40.S32	DePCR	Pool	40	FLD0032	GTGCGCTAGT	Chin. Feces	10	151979	96133	9612	9602	10	0.10%	2.54	0.86	51	0.65	4500	10, 11, T7, T5
1	C33.40.S33	DePCR	Pool	40	FLD0033	TGTGCTCGCA	Chin. Feces	10	200714	121186	5265	5263	2	0.04%	2.54	0.87	48	0.66	4500	10, 11, T7, T5
1	C34.40.S34	DePCR	Pool	40	FLD0034	GATGCGAGCT	Chin. Feces	10	127039	81670	8339	8335	4	0.05%	2.52	0.86	50	0.64	4500	10, 11, T7, T5
1	C35.40.S35	DePCR	Pool	40	FLD0035	CTGTACGTGA	Chin. Feces	10	190573	119628	9367	9364	3	0.03%	2.60	0.87	53	0.66	4500	10, 11, T7, T5
1	C37.45.S37	DePCR	Pool	45	FLD0037	TGTCGAGTCA	Chin. Feces	10	52927	40428	21314	21294	20	0.09%	2.55	0.85	59	0.63	4500	10, 11, T7, T5
1	C38.45.S38	DePCR	Pool	45	FLD0038	GTCTACTGTC	Chin. Feces	10	55533	37992	19429	19425	4	0.02%	2.60	0.86	59	0.64	4500	10, 11, T7, T5
1	C39.45.S39	DePCR	Pool	45	FLD0039	CAGTCAGAGT	Chin. Feces	10	51984	38396	17668	17664	4	0.02%	2.60	0.87	55	0.65	4500	10, 11, T7, T5
1	C40.45.S40	DePCR	Pool	45	FLD0040	CGCAGTCTAT	Chin. Feces	10	47578	36101	17542	17513	29	0.17%	2.56	0.86	52	0.65	4500	10, 11, T7, T5
1	C41.45.S41	DePCR	Pool	45	FLD0041	GTATGAGCAC	Chin. Feces	10	59059	46008	21208	21171	37	0.17%	2.62	0.87	61	0.64	4500	10, 11, T7, T5
1	C43.50.S43	DePCR	Pool	50	FLD0043	TATAGCACGC	Chin. Feces	10	48121	46185	38368	38346	22	0.06%	2.66	0.87	58	0.65	(4500), [1800]	(10, 11, T7, T5), [12, T6]
1	C44.50.S44	DePCR	Pool	50	FLD0044	TCATGCGCGA	Chin. Feces	10	50657	47927	38755	38732	23	0.06%	2.66	0.88	56	0.66	(4500), [1800]	(10, 11, T7, T5), [12, T6]
1	C45.50.S45	DePCR	Pool	50	FLD0045	TATGCGCTGC	Chin. Feces	10	42712	40960	34233	34226	7	0.02%	2.67	0.87	59	0.66	(4500), [1800]	(10, 11, T7, T5), [12, T6]
1	C46.50.S46	DePCR	Pool	50	FLD0046	TCTCTGTGCA	Chin. Feces	10	55328	52123	42270	42254	16	0.04%	2.68	0.87	66	0.64	(4500), [1800]	(10, 11, T7, T5), [12, T6]
1	C47.50.S47	DePCR	Pool	50	FLD0047	CTATCGCGTG	Chin. Feces	10	47576	44772	36513	36500	13	0.04%	2.65	0.87	59	0.65	(4500), [1800]	(10, 11, T7, T5), [12, T6]
1	C49.55.S49	DePCR	Pool	55	FLD0049	CTGCATGATC	Chin. Feces	10	73613	71480	61024	60995	29	0.05%	2.67	0.87	59	0.65	4500	10, 11, T7, T5
1	C50.55.S50	DePCR	Pool	55	FLD0050	CGCGTATCAT	Chin. Feces	10	59831	57618	48634	48594	40	0.08%	2.63	0.87	61	0.64	4500	10, 11, T7, T5
1	C51.55.S51	DePCR	Pool	55	FLD0051	GTATCTCTCG	Chin. Feces	10	77419	75569	65315	65279	36	0.06%	2.61	0.86	55	0.65	4500	10, 11, T7, T5
1	C52.55.S52	DePCR	Pool	55	FLD0052	GCTCATATGC	Chin. Feces	10	67895	65982	56098	56084	14	0.02%	2.67	0.87	59	0.65	4500	10, 11, T7, T5
1	C53.55.S53	DePCR	Pool	55	FLD0053	CACTATGTCG	Chin. Feces	10	56623	54431	45771	45749	22	0.05%	2.62	0.87	59	0.64	4500	10, 11, T7, T5
1	C55.60.S55	DePCR	Pool	60	FLD0055	CGTCACAGTA	Chin. Feces	10	21333	20566	16972	16970	2	0.01%	2.66	0.88	58	0.65	4500	10, 11, T7, T5
1	C56.60.S56	DePCR	Pool	60	FLD0056	TCGCGTGAGA	Chin. Feces	10	27482	26666	22530	22521	9	0.04%	2.63	0.88	56	0.65	4500	10, 11, T7, T5
1	C57.60.S57	DePCR	Pool	60	FLD0057	TACATCGCTG	Chin. Feces	10	22266	21670	18212	18204	8	0.04%	2.58	0.86	58	0.63	4500	10, 11, T7, T5
1	C58.60.S58	DePCR	Pool	60	FLD0058	GTGAGAGACA	Chin. Feces	10	40704	39657	33921	33909	12	0.04%	2.66	0.87	58	0.65	4500	10, 11, T7, T5
1	C59.60.S59	DePCR	Pool	60	FLD0059	GACTGTACGT	Chin. Feces	10	25664	24939	21048	21046	2	0.01%	2.65	0.88	53	0.67	4500	10, 11, T7, T5
2	C117_CACAGTGATG	DePCR	Pool	50	FLD0118	CACAGTGATG	Chin. Feces	20	252743	247066	229042	228950	92	0.04%	2.69	0.87	66	0.64	8000	12, T7
2	C118_CGAGCTAGCA	DePCR	Pool	50	FLD0119	CGAGCTAGCA	Chin. Feces	20	152773	149474	139195	139172	23	0.02%	2.69	0.88	67	0.64	8000	12, T7
2	C119_GAGACTATGC	DePCR	Pool	50	FLD0120	GAGACTATGC	Chin. Feces	20	564011	556769	528053	527730	323	0.06%	2.64	0.87	66	0.63	8000	12, T7
2	C120_CAGAGCTAGT	DePCR	Pool	50	FLD0121	CAGAGCTAGT	Chin. Feces	20	311297	305120	285609	285384	225	0.08%	2.68	0.87	68	0.64	8000	12, T7
2	C121_CGCAGAGCAT	DePCR	Pool	50	FLD0122	CGCAGAGCAT	Chin. Feces	20	100316	97814	91440	91403	37	0.04%	2.69	0.87	69	0.63	8000	12, T7
2	C122_TGTACAGCGA	DePCR	Pool	50	FLD0123	TGTACAGCGA	Chin. Feces	10	155895	154408	144673	144659	14	0.01%	2.69	0.88	63	0.65	8000	12, T7
2	C123_ACGTCAGTAT	DePCR	Pool	50	FLD0124	ACGTCAGTAT	Chin. Feces	10	156965	153013	143410	143300	110	0.08%	2.69	0.88	65	0.64	8000	12, T7
2	C124_TCACAGCATA	DePCR	Pool	50	FLD0125	TCACAGCATA	Chin. Feces	10	304500	299555	280936	280803	133	0.05%	2.66	0.87	65	0.64	8000	12, T7
2	C125_ACTGCGTGTC	DePCR	Pool	50	FLD0126	ACTGCGTGTC	Chin. Feces	10	168252	165075	154436	154382	54	0.03%	2.71	0.88	66	0.65	8000	12, T7
2	C126_CGATCGACTG	DePCR	Pool	50	FLD0127	CGATCGACTG	Chin. Feces	10	57229	55829	52064	52061	3	0.01%	2.66	0.87	64	0.64	8000	12, T7
2	C127_GCGAGATGTA	DePCR	Pool	50	FLD0128	GCGAGATGTA	Chin. Feces	5	90420	88868	83583	83552	31	0.04%	2.69	0.88	65	0.64	8000	12, T7
2	C128_CTGATGCAGA	DePCR	Pool	50	FLD0129	CTGATGCAGA	Chin. Feces	5	47463	46372	42865	42849	16	0.04%	2.69	0.88	63	0.65	8000	12, T7
2	C129_GTGACGTACG	DePCR	Pool	50	FLD0130	GTGACGTACG	Chin. Feces	5	79433	78588	73963	73932	31	0.04%	2.66	0.87	62	0.64	8000	12, T7
2	C130_CGACGCTGAT	DePCR	Pool	50	FLD0131	CGACGCTGAT	Chin. Feces	5	58366	57004	51664	51654	10	0.02%	2.68	0.88	63	0.65	8000	12, T7
2	C131_CTACGATCAG	DePCR	Pool	50	FLD0132	CTACGATCAG	Chin. Feces	5	16138	15778	14339	14339	0	0.00%	2.70	0.88	64	0.65	8000	12, T7
2	C132_GCACTAGACA	DePCR	Pool	50	FLD0133	GCACTAGACA	Chin. Feces	2.5	30696	30016	27759	27756	3	0.01%	2.68	0.87	62	0.65	8000	12, T7
2	C133_CTAGCAGATG	DePCR	Pool	50	FLD0134	CTAGCAGATG	Chin. Feces	2.5	11147	10874	10067	10063	4	0.04%	2.68	0.87	56	0.67	8000	12, T7

Exp	Sample	PCR Method	Primer Variant	Annealing Temp	Barcode Name	Barcode Sequence	Input DNA	gDNA Conc (ng/ul)	Raw Reads	Merged	Trimmed	Post Chimera Removal	Chimeras Removed	Chimera Removed (%)	Shannon (Genus)	Simpson (Genus)	Richness (Genus)	Evenness (Genus)	Rarefaction	Reference Figures / Tables
2	C134_CATGATACGC	DePCR	Pool	50	FLD0135	CATGATACGC	Chin. Feces	2.5	36464	35614	33040	33031	9	0.03%	2.69	0.88	66	0.64	8000	12, T7
2	C135_GCAGCTGTCA	DePCR	Pool	50	FLD0136	GCAGCTGTCA	Chin. Feces	2.5	50445	49572	46618	46609	9	0.02%	2.68	0.87	67	0.64	8000	12, T7
2	C136_ACGTATCATC	DePCR	Pool	50	FLD0137	ACGTATCATC	Chin. Feces	2.5	39358	38371	35836	35831	5	0.01%	2.64	0.87	60	0.65	8000	12, T7
2	C137_AGTATCGTAC	DePCR	Pool	50	FLD0138	AGTATCGTAC	Chin. Feces	1.25	28198	27912	26590	26575	15	0.06%	2.64	0.87	62	0.64	8000	12, T7
2	C138_GATACACTGA	DePCR	Pool	50	FLD0139	GATACACTGA	Chin. Feces	1.25	37941	37379	35198	35192	6	0.02%	2.66	0.87	61	0.65	8000	12, T7
2	C139_GACTAGTCAG	DePCR	Pool	50	FLD0140	GACTAGTCAG	Chin. Feces	1.25	8852	8681	8106	8105	1	0.01%	2.64	0.86	63	0.64	8000	12, T7
2	C140_GATGACTACG	DePCR	Pool	50	FLD0141	GATGACTACG	Chin. Feces	1.25	22351	22118	20909	20907	2	0.01%	2.65	0.87	61	0.65	8000	12, T7
2	C141_CAGAGAGTCA	DePCR	Pool	50	FLD0142	CAGAGAGTCA	Chin. Feces	1.25	30685	29983	28045	28024	21	0.07%	2.68	0.88	60	0.66	8000	12, T7
3	C233_GCCATGTCAT	TAS	RPV1	50	FLD0238	GCCATGTCAT	Chin. Feces	10	25323	24827	22407	19426	2981	13.30%	2.39	0.84	47	0.62	1800	12, T7, T6
3	C234_TCTGCCTATA	TAS	RPV1	50	FLD0239	TCTGCCTATA	Chin. Feces	10	23815	23535	21521	17893	3628	16.86%	2.38	0.84	50	0.61	1800	12, T7, T6
3	C235_CTTAGTTCGC	TAS	RPV2	50	FLD0240	CTTAGTTCGC	Chin. Feces	10	31645	31166	28150	24302	3848	13.67%	2.33	0.82	42	0.62	1800	12, T7, T6
3	C236_CGTAATGAGC	TAS	RPV2	50	FLD0241	CGTAATGAGC	Chin. Feces	10	40115	39306	35073	30544	4529	12.91%	2.35	0.82	46	0.61	1800	12, T7, T6
3	C237_TTGCTTAGTC	TAS	RPV3	50	FLD0242	TTGCTTAGTC	Chin. Feces	10	298471	295760	274907	230709	44198	16.08%	2.52	0.86	50	0.64	1800	12, T7, T6
3	C238_TCTTGTTCAC	TAS	RPV3	50	FLD0243	TCTTGTTCAC	Chin. Feces	10	30498	30240	27952	23660	4292	15.35%	2.33	0.82	47	0.61	1800	12, T7, T6
3	C239_GTGGCTTCGT	TAS	RPV4	50	FLD0244	GTGGCTTCGT	Chin. Feces	10	41374	40961	37803	32460	5343	14.13%	2.22	0.80	47	0.58	1800	12, T7, T6
3	C240_TGTTCGATAG	TAS	RPV4	50	FLD0245	TGTTCGATAG	Chin. Feces	10	25852	25649	24002	21188	2814	11.72%	2.23	0.80	45	0.59	1800	12, T7, T6
3	C241_TCATTCAGTG	TAS	RPV5	50	FLD0246	TCATTCAGTG	Chin. Feces	10	28622	28224	25617	22465	3152	12.30%	2.39	0.83	50	0.61	1800	12, T7, T6
3	C243_GTAGAAGTGG	TAS	RPV6	50	FLD0248	GTAGAAGTGG	Chin. Feces	10	27757	27410	25169	22423	2746	10.91%	2.49	0.85	46	0.65	1800	12, T7, T6
3	C244_TGGAGCATGT	TAS	RPV6	50	FLD0249	TGGAGCATGT	Chin. Feces	10	38451	38111	35336	28986	6350	17.97%	2.45	0.84	51	0.62	1800	12, T7, T6
3	C245_GAAGGAGATA	TAS	RPV7	50	FLD0250	GAAGGAGATA	Chin. Feces	10	35859	35195	31942	28497	3445	10.79%	2.45	0.84	54	0.61	1800	12, T7, T6
3	C246_CGAATGTATG	TAS	RPV7	50	FLD0251	CGAATGTATG	Chin. Feces	10	36804	36049	32257	27408	4849	15.03%	2.49	0.85	51	0.63	1800	12, T7, T6
3	C247_TCGTGAATGA	TAS	RPV8	50	FLD0252	TCGTGAATGA	Chin. Feces	10	40014	39370	35612	30546	5066	14.23%	2.39	0.83	53	0.60	1800	12, T7, T6
3	C248_GAATAGCTGA	TAS	RPV8	50	FLD0253	GAATAGCTGA	Chin. Feces	10	28376	27873	25128	21327	3801	15.13%	2.41	0.83	60	0.59	1800	12, T7, T6
3	C249_TTGTCACATC	TAS	RPV9	50	FLD0254	TTGTCACATC	Chin. Feces	10	31182	30901	28727	25836	2891	10.06%	2.48	0.84	61	0.60	1800	12, T7, T6
3	C250_CTGGAGGCTA	TAS	RPV9	50	FLD0255	CTGGAGGCTA	Chin. Feces	10	37611	36855	33180	28017	5163	15.56%	2.35	0.82	47	0.61	1800	12, T7, T6
3	C251_TGTCAGCTTA	TAS	RPV10	50	FLD0256	TGTCAGCTTA	Chin. Feces	10	27558	27323	25357	24059	1298	5.12%	2.45	0.83	51	0.62	1800	12, T7, T6
3	C252_GTTCTTCGTA	TAS	RPV10	50	FLD0257	GTTCTTCGTA	Chin. Feces	10	33771	33377	30489	28638	1851	6.07%	2.50	0.85	55	0.62	1800	12, T7, T6
3	C253_TTACACGTTC	TAS	RPV11	50	FLD0258	TTACACGTTC	Chin. Feces	10	34670	34405	32205	27148	5057	15.70%	2.28	0.83	43	0.61	1800	12, T7, T6
3	C254_GTAGCCAGTA	TAS	RPV11	50	FLD0259	GTAGCCAGTA	Chin. Feces	10	31863	31427	28645	24163	4482	15.65%	2.33	0.83	46	0.61	1800	12, T7, T6
3	C255_TGAGAAGGTA	TAS	RPV12	50	FLD0260	TGAGAAGGTA	Chin. Feces	10	33361	32959	30636	27542	3094	10.10%	2.57	0.86	53	0.65	1800	12, T7, T6
3	C256_CCATATGATC	TAS	RPV12	50	FLD0261	CCATATGATC	Chin. Feces	10	45388	44360	39330	35221	4109	10.45%	2.61	0.87	51	0.66	1800	12, T7, T6
3	C257_CGATCCTATA	TAS	RPV13	50	FLD0262	CGATCCTATA	Chin. Feces	10	30251	29593	26516	23537	2979	11.23%	2.50	0.86	47	0.65	1800	12, T7, T6
3	C258_TGACTAGCTT	TAS	RPV13	50	FLD0263	TGACTAGCTT	Chin. Feces	10	33869	33516	30918	27676	3242	10.49%	2.51	0.86	53	0.63	1800	12, T7, T6
3	C259_TAACTCTGCT	TAS	RPV14	50	FLD0264	TAACTCTGCT	Chin. Feces	10	39100	38714	35657	33801	1856	5.21%	2.37	0.82	52	0.60	1800	12, T7, T6
3	C260_TCGAATGTGC	TAS	RPV14	50	FLD0265	TCGAATGTGC	Chin. Feces	10	33548	33143	30194	28612	1582	5.24%	2.39	0.82	52	0.61	1800	12, T7, T6
3	C263_GAACTATCAC	TAS	RPV16	50	FLD0268	GAACTATCAC	Chin. Feces	10	21642	21290	19106	16642	2464	12.90%	2.45	0.84	57	0.61	1800	12, T7, T6
3	C264_TCGAGGTACT	TAS	RPV16	50	FLD0269	TCGAGGTACT	Chin. Feces	10	12922	12782	11723	10063	1660	14.16%	2.36	0.83	61	0.57	1800	12, T7, T6
3	C265_TGCGGATGGT	TAS	RPV17	50	FLD0270	TGCGGATGGT	Chin. Feces	10	26736	26468	24507	21618	2889	11.79%	2.43	0.84	52	0.62	1800	12, T7, T6
3	C266_TTCGAGCTAT	TAS	RPV17	50	FLD0271	TTCGAGCTAT	Chin. Feces	10	25453	25235	23578	21144	2434	10.32%	2.48	0.85	57	0.61	1800	12, T7, T6
3	C267_GGTCTGGTGT	TAS	RPV18	50	FLD0272	GGTCTGGTGT	Chin. Feces	10	32478	32152	29696	26850	2846	9.58%	2.26	0.80	49	0.58	1800	12, T7, T6
3	C268_CTAAGTCATG	TAS	RPV18	50	FLD0273	CTAAGTCATG	Chin. Feces	10	33815	33187	29612	25621	3991	13.48%	2.26	0.80	47	0.59	1800	12, T7, T6
3	C269_TTGCAGATCA	DePCR	RPV1	50	FLD0274	TTGCAGATCA	Chin. Feces	10	57091	56582	51975	51958	17	0.03%	2.65	0.87	52	0.67	1800	12, T7, T6
3	C270_CTGCGAATGT	DePCR	RPV1	50	FLD0275	CTGCGAATGT	Chin. Feces	10	43649	42770	38050	38014	36	0.09%	2.53	0.86	47	0.66	1800	12, T7, T6
3	C271_CTGTTCTAGC	DePCR	RPV2	50	FLD0276	CTGTTCTAGC	Chin. Feces	10	43957	42996	37860	37819	41	0.11%	2.54	0.86	44	0.67	1800	12, T7, T6
3	C272_CACTTGTGTG	DePCR	RPV2	50	FLD0277	CACTTGTGTG	Chin. Feces	10	44722	43601	38312	38208	104	0.27%	2.55	0.86	47	0.66	1800	12, T7, T6
3	C273_TGGATGACAT	DePCR	RPV3	50	FLD0278	TGGATGACAT	Chin. Feces	10	64037	63247	58090	58072	18	0.03%	2.64	0.87	49	0.68	1800	12, T7, T6
3	C274_GATCCTGAGC	DePCR	RPV3	50	FLD0279	GATCCTGAGC	Chin. Feces	10	86323	84606	75864	75838	26	0.03%	2.67	0.88	49	0.69	1800	12, T7, T6
3	C275_GTCGGTCTGA	DePCR	RPV4	50	FLD0280	GTCGGTCTGA	Chin. Feces	10	41753	41237	37661	37656	5	0.01%	2.46	0.85	47	0.64	1800	12, T7, T6
3	C276_TGTTACGATC	DePCR	RPV4	50	FLD0281	TGTTACGATC	Chin. Feces	10	55400	54919	50942	50918	24	0.05%	2.50	0.85	48	0.65	1800	12, T7, T6
3	C277_GTCTTGGCTC	DePCR	RPV5	50	FLD0282	GTCTTGGCTC	Chin. Feces	10	74461	73412	66368	66353	15	0.02%	2.67	0.88	48	0.69	1800	12, T7, T6
3	C278_GGTCGTGCAT	DePCR	RPV5	50	FLD0283	GGTCGTGCAT	Chin. Feces	10	46200	45428	41269	41232	37	0.09%	2.61	0.86	55	0.65	1800	12, T7, T6
3	C279_CAGGCTCAGT	DePCR	RPV6	50	FLD0284	CAGGCTCAGT	Chin. Feces	10	60720	59076	52490	52465	25	0.05%	2.56	0.87	45	0.67	1800	12, T7, T6
3	C280_TAGCTTCACT	DePCR	RPV6	50	FLD0285	TAGCTTCACT	Chin. Feces	10	62842	62241	57602	57590	12	0.02%	2.61	0.87	48	0.67	1800	12, T7, T6
3	C281_CAGATGTCCT	DePCR	RPV7	50	FLD0286	CAGATGTCCT	Chin. Feces	10	33724	32860	29078	29070	8	0.03%	2.56	0.86	44	0.68	1800	12, T7, T6
3	C282_TTACGCAGTG	DePCR	RPV7	50	FLD0287	TTACGCAGTG	Chin. Feces	10	59302	58514	53230	53219	11	0.02%	2.64	0.87	51	0.67	1800	12, T7, T6
3	C283_TTCGTTCCTG	DePCR	RPV8	50	FLD0288	TTCGTTCCTG	Chin. Feces	10	78565	77768	71849	71839	10	0.01%	2.49	0.85	47	0.65	1800	12, T7, T6
3	C284_CACTGCTTGA	DePCR	RPV8	50	FLD0289	CACTGCTTGA	Chin. Feces	10	88732	85851	74806	74731	75	0.10%	2.53	0.86	45	0.67	1800	12, T7, T6
3	C285_TCTAGCGTGG	DePCR	RPV9	50	FLD0290	TCTAGCGTGG	Chin. Feces	10	44913	44244	39268	39264	4	0.01%	2.56	0.86	52	0.65	1800	12, T7, T6
3	C286_GCATAATCGC	DePCR	RPV9	50	FLD0291	GCATAATCGC	Chin. Feces	10	97832	95600	84767	84748	19	0.02%	2.61	0.87	53	0.66	1800	12, T7, T6
3	C287_GTCGTAACAC	DePCR	RPV10	50	FLD0292	GTCGTAACAC	Chin. Feces	10	8884	8694	6919	6915	4	0.06%	2.80	0.88	53	0.71	1800	12, T7, T6
3	C288_GAGATTGCTA	DePCR	RPV10	50	FLD0293	GAGATTGCTA	Chin. Feces	10	25158	24762	21135	21056	79	0.37%	3.70	0.95	116	0.78	1800	12, T7, T6
3	C289_GGACAGATGG	DePCR	RPV11	50	FLD0294	GGACAGATGG	Chin. Feces	10	53963	52979	47319	47288	31	0.07%	2.48	0.85	49	0.64	1800	12, T7, T6
3	C290_CTTACGTTGC	DePCR	RPV11	50	FLD0295	CTTACGTTGC	Chin. Feces	10	29952	29305	25072	25069	3	0.01%	2.53	0.86	46	0.66	1800	12, T7, T6
3	C291_GTGTTCGGTC	DePCR	RPV12	50	FLD0296	GTGTTCGGTC	Chin. Feces	10	25585	25247	22488	22457	31	0.14%	2.65	0.86	49	0.68	1800	12, T7, T6
3	C292_CTCAAGAAGC	DePCR	RPV12	50	FLD0297	CTCAAGAAGC	Chin. Feces	10	23160	22540	19328	19320	8	0.04%	2.49	0.84	43	0.66	1800	12, T7, T6
3	C293_TCTCGGATAG	DePCR	RPV13	50	FLD0298	TCTCGGATAG	Chin. Feces	10	20172	19756	17331	17329	2	0.01%	2.61	0.87	45	0.69	1800	12, T7, T6
3	C294_CTCTGGACGA	DePCR	RPV13	50	FLD0299	CTCTGGACGA	Chin. Feces	10	28219	27463	23845	23837	8	0.03%	2.64	0.87	51	0.67	1800	12, T7, T6

Exp	Sample	PCR Method	Primer Variant	Annealing Temp	Barcode Name	Barcode Sequence	Input DNA	gDNA Conc (ng/ul)	Raw Reads	Merged	Trimmed	Post Chimera Removal	Chimeras Removed	Chimera Removed (%)	Shannon (Genus)	Simpson (Genus)	Richness (Genus)	Evenness (Genus)	Rarefaction	Reference Figures / Tables
3	C295_CGAGCATTGT	DePCR	RPV14	50	FLD0300	CGAGCATTGT	Chin. Feces	10	3920	3802	3024	3024	0	0.00%	2.36	0.83	27	0.72	1800	12, T7, T6
3	C296_CCAAGAAGAA	DePCR	RPV14	50	FLD0301	CCAAGAAGAA	Chin. Feces	10	9370	8896	7239	7239	0	0.00%	2.37	0.81	36	0.66	1800	12, T7, T6
3	C297_TCCTTGTTCT	DePCR	RPV15	50	FLD0302	TCCTTGTTCT	Chin. Feces	10	2454	2347	1829	1828	1	0.05%	2.39	0.85	25	0.74	1800	12, T7, T6
3	C298_GTAACGATGT	DePCR	RPV15	50	FLD0303	GTAACGATGT	Chin. Feces	10	71521	70416	63694	63681	13	0.02%	2.56	0.87	46	0.67	1800	12, T7, T6
3	C299_TGGACTCAGA	DePCR	RPV16	50	FLD0304	TGGACTCAGA	Chin. Feces	10	308400	304388	275824	275695	129	0.05%	2.57	0.86	51	0.65	1800	12, T7, T6
3	C300_GGCATCATGC	DePCR	RPV16	50	FLD0305	GGCATCATGC	Chin. Feces	10	53742	52692	46580	46555	25	0.05%	2.60	0.86	55	0.65	1800	12, T7, T6
3	C301_GTATAACGCT	DePCR	RPV17	50	FLD0306	GTATAACGCT	Chin. Feces	10	15551	15256	13276	13275	1	0.01%	2.58	0.88	38	0.71	1800	12, T7, T6
3	C302_GCAGATAAGT	DePCR	RPV17	50	FLD0307	GCAGATAAGT	Chin. Feces	10	19224	18811	16300	16296	4	0.02%	2.53	0.85	47	0.66	1800	12, T7, T6
3	C303_GTCGGCTCTA	DePCR	RPV18	50	FLD0308	GTCGGCTCTA	Chin. Feces	10	10297	10132	8693	8692	1	0.01%	2.37	0.84	36	0.66	1800	12, T7, T6
3	C304_TTCGATAGCA	DePCR	RPV18	50	FLD0309	TTCGATAGCA	Chin. Feces	10	9927	9757	8542	8536	6	0.07%	2.33	0.81	40	0.63	1800	12, T7, T6

Chapter IV: PCR effects of melting temperature adjustment of individual primers in degenerate primer pools

(Previously published as Naqib A, Jeon T, Kunstman K, Wang W, Shen Y, Sweeney D, Hyde M, Green SJ. 2019. PCR effects of melting temperature adjustment of individual primers in degenerate primer pools. PeerJ 7:e6570.)

Abstract

Deep sequencing of small subunit ribosomal RNA (SSU rRNA) gene amplicons continues to be the most common approach for characterization of complex microbial communities. PCR amplifications of conserved regions of SSU rRNA genes often employ degenerate pools of primers to enable targeting of a broad spectrum of organisms. One little noticed feature of such degenerate primer sets is the potential for a wide range of melting temperatures between the primer variants. The melting temperature variation of primers in a degenerate pool could lead to variable amplification efficiencies and PCR bias. Thus, we sought to adjust the melting temperature of each primer variant individually. Individual primer modifications were used to reduce theoretical melting temperature variation between primers, as well as to introduce inter-cluster nucleotide diversity during Illumina sequencing of primer regions. We demonstrate here the suitability of such primers for microbial community analysis. However, no substantial differences in microbial community structure were revealed when using primers with adjusted melting temperatures, though the optimal annealing temperature decreased.

Introduction

In molecular surveys performed in the field of molecular microbial ecology, genes such as those encoding for the small subunit ribosomal RNA (SSU rRNA), dissimilatory sulfite reductase, nitrite reductase, and more are frequently targeted to survey the total microbial community or members of specific metabolic groups. For broad surveys of microbial community structure, the SSU rRNA gene is most frequently targeted. One common feature of primers used for microbial community surveys is sequence degeneracy; that is, rather than a single primer, a mixture of multiple highly similar primers, targeting multiple variants of priming regions, is employed to cover as broad a taxonomic range of organisms as possible. Over time, the degeneracy of commonly used primers tends to increase due to the availability of additional sequence data demonstrating mismatches between primers and novel sequences. In many cases, introduction of new variants has been highly successful, allowing the detection of specific microbial clades. For example, the commonly used 16S rRNA gene primer 27F does not properly amplify genomic DNA from bacteria of the genus *Bifidobacterium* and other taxa [46, 52]. A series of additional degeneracies were introduced to improve taxonomic coverage [52]. More recently, additional degeneracies were added to the 515F primer to allow targeting of *Crenarchaeota* and *Thaumarchaeota* and to 806R to allow targeting of the SAR11 clade [82, 83].

As new variants are introduced into primer pools, primer melting temperature is rarely considered. Thus, the melting temperature (Tm) of primer variants within a degenerate pool can have a substantial range. For example, the original '806R' degenerate primer employed by the Earth Microbiome Project (EMP) is 18-fold degenerate, with a theoretical Tm range of approximately 7°C. We previously demonstrated that the annealing temperature of PCRs altered the profile of primers in a degenerate pool annealing to genomic DNA (so-called "primer

utilization profiles" or PUP; [18]). The variable melting temperature of the individual primers in a degenerate pool creates the potential for additional bias, as differences in free energy binding between primers and perfectly matching templates will vary. Furthermore, mismatch interactions could be favored when primers with high theoretical Tm values are used in PCR reactions run at low annealing temperatures necessary for low Tm primers. Alternatively, some primers may not have the opportunity to anneal in PCR reactions with elevated annealing temperatures. Such reactions can still yield amplification due to the presence of high Tm primers in the degenerate pool.

The objective of this study was to develop and test an experimental system to determine whether the broad range of primer melting temperatures in degenerate primer pools contributes substantially to observed microbial community profiles generated from amplicon-based sequencing approaches. To do so, we altered each primer independently by removing nucleotides from both 5' and 3' ends of oligonucleotide primers to minimize variance in primer theoretical Tm. These oligonucleotides were synthesized independently and pooled in equimolar concentration. Subsequently, these primer pools were used for PCR amplification of several complex genomic DNA samples, followed by high-throughput sequencing. Sequencing results were compared to amplicon sequence data generated from PCRs employing standard degenerate primers. As a secondary objective, we sought to determine if we could modify primer sequences at the 5' end to introduce nucleotide diversity into amplicons sequenced on next-generation sequencers.

Materials and Methods

Primer Design and Synthesis

The most recent Earth Microbiome Project (EMP) primers, 515F (Parada) ["515F"] and 806R [82] ["806R"], were utilized as the default primer set [73, 82, 83, 89]. These primers are 4-fold degenerate (515F) and 24-fold degenerate (806R), with theoretical melting temperatures ranging from 66.9 to 71.8°C (515F) and from 54.7 to 61.7°C (806R) (**Table S5**). Primer theoretical melting temperatures were calculated using the OligoAnalyzer3.1 calculator [74], assuming 250 nM primer concentration, 2 mM Mg^{2+}, and 0.2 mM dNTPs. Maximum Delta G values for each sequence were calculated using the self-dimer option in the OligoAnalyzer software. Primers were synthesized either as single degenerate primer pools (standard approach), or as individual primers without degeneracies by Integrated DNA Technologies (IDT; Coralville, IA). All primers were synthesized as LabReady and delivered at a fixed concentration of 100 micromolar. Most primers contained common sequence linkers (CS1 and CS2) at the 5' ends, as shown in **Table S6**. These linker sequences are necessary for the later incorporation of Illumina sequencing adapters and sample-specific barcodes. For new primer pools containing shortened primers, a combination of either 4 ("515F") or 24 ("806R") non-degenerate primers were combined in equal volume to generate degenerate pools at 100 micromolar concentration. These primer pools were named "515F_pool" and "806R_pool", and when these primers were used together, the primer set was named "ShortEMP". One experiment, described below, employed ShortEMP primer pools without common sequences ("NoLinker_ShortEMP") to assess the effect of the linker sequences on analysis of microbial community structure. A final set of primers, employing variable length spacers, was generated and named the "LongEMP" primer set. These primers included variable length spacers between the common sequence linkers and

the gene-specific EMP primer regions, as described previously [90]. Sequencing with these primers was performed on both MiniSeq and MiSeq instruments, but data from the MiniSeq using the "LongEMP" primer set did not properly merge due to read-length limitations, and was not further analyzed.

Genomic DNA Templates

Four microbial genomic DNA (gDNA) samples were employed in this study. These include the ZymoBIOMICS Microbial Community DNA standard (D6306; Zymo Research, Irvine, CA; 'Zymo'), as well as three environmental samples derived from Lake Michigan sediment ('LMC'), garden soil ('Soil') and mammalian [32] feces ('Feces'). Sediment, soil and fecal samples were extracted using a PowerSoil DNA extraction kit (Qiagen, Hilden, Germany), following the manufacturer's protocol.

Amplicon library preparation and sequencing

A two-stage PCR amplification strategy was used to generate sequencer-ready amplicons [38]. Genomic DNA was first PCR amplified with primer set EMP (CS1_806R and CS2_515F), ShortEMP (CS1_806R_pool and CS2_515F_pool), or LongEMP (CS1_806R_long_pool and CS2_806R_long_pool) (**Table S6**). All primers contained 5' common sequence tags (known as common sequence 1 and 2, CS1 and CS2) as described previously [61]. First stage PCR amplifications were performed in 10 microliter reactions in 96-well plates, using MyTaq HS 2X mastermix (Bioline, Taunton, MA). PCR conditions were 95°C for 5 minutes, followed by 28 cycles of 95°C for 30", variable annealing temperature for 60" and 72°C for 90". For

temperature gradients, annealing temperatures of 40°, 45° and 50°C were employed. Subsequently, annealing temperatures of 45°C were used for ShortEMP reactions and 50°C for EMP reactions.

Second stage PCR amplifications were performed in 10 microliter reactions in 96-well plates. A mastermix for the entire plate was made using the MyTaq HS 2X mastermix. Each well received a separate primer pair with a unique 10-base barcode, obtained from the Access Array Barcode Library for Illumina (Fluidigm, South San Francisco, CA; Item# 100-4876). These AccessArray primers contained the CS1 and CS2 linkers at the 3' ends of the oligonucleotides. Cycling conditions were as follows: 95°C for 5 minutes, followed by 8 cycles of 95°C for 30", 60°C for 30" and 72°C for 30".

In an experiment using the NoLinker_ShortEMP primers, the above protocol was modified slightly. To assess the effect of common sequence linkers on the observed microbial community structure, we performed the first stage PCR amplification with NoLinker_ShortEMP primers. PCR conditions were identical to those described above. During the second stage PCR amplification, both Fluidigm Access Array Barcode and the ShortEMP primers were included, and 12 cycles of amplification were performed in place of 8. In this approach, common sequences are incorporated during the second stage PCR as the ShortEMP primers *with* linkers amplify the NoLinker_ShortEMP amplicons; subsequently, the Fluidigm Access Array barcode primers amplify amplicons containing the common sequence linkers.

In all experiments, samples were pooled using an EpMotion5075 liquid handling robot (Eppendorf, Hamburg, Germany). Pooled libraries were purified using an AMPure XP cleanup protocol (0.6X, vol/vol; Agencourt, Beckmann-Coulter) to remove fragments smaller than 300 bp. The pooled libraries, with either a 1% or 20% phiX spike-in, were loaded onto either an

Illumina MiniSeq mid-output kit (2x153 paired-end reads) or an Illumina MiSeq V2 kit (2x250 paired-end reads), as indicated in **Table S7**. Fluidigm sequencing primers, targeting the CS1 and CS2 linker regions, were used to initiate sequencing. De-multiplexing of reads was performed on instrument. Library preparation, pooling, and sequencing were performed at the University of Illinois at Chicago Sequencing Core (UICSQC).

Bioinformatic Analysis of Sequence Data

Raw FASTQ files were downloaded from Illumina Basespace. Sequence reads were merged using PEAR (Paired-End Read Merger) [75] with default parameters. Merged reads were quality trimmed (<Q20 discarded) and length trimmed (<250 bases were removed), and primer sequences were removed using the 'Trim reads' algorithm within the software package CLC Genomics Workbench (v11; Qiagen; Hilden, Germany). Trimmed sequences were reverse complemented using a QIIME script. Chimeras were removed using the USEARCH81 algorithm [76]. Subsequently, for environmental samples (LMC, Soil, Feces), sequences were pooled, renamed and clustered into operational taxonomic units (OTUs) at a threshold of 97% similarity (QIIME v1.8.0; [1]). Each OTU was annotated taxonomically based on the representative sequences using the UCLUST algorithm and the greengenes 13_8 reference database [65]. A biological observational matrix (BIOM) was generated from the clustered OTU data and the taxonomy data [77]. For subsequent analyses, data were rarefied using the *vegan* package within the R programming language. Rarefaction depths were adjusted by analysis; depth of rarefaction for each figure is shown in **Table S7**. The BIOMs were analyzed and visualized using the software package Primer7 [66] and in the R programming environment [79]. Dendrogram creation and SIMPROF tests were conducted within Primer7. The *vegan* package [80] was used

to generate alpha diversity indices and to calculate pairwise Bray-Curtis dissimilarity scores. Analysis of similarity (ANOSIM) calculations were performed at the taxonomic level of genus, using square root transformed data. Metric multi-dimensional scaling (mMDS) plots were created using the cmdscale and ggplot2 functions within R. Ellipses, representing a 95% confidence interval around group centroids, were created assuming a multivariate t-distribution [91]. Taxon-level differential abundances between sample groups were identified using the software package STAMP [92] employing White's non-parametric t-test [93]. P values were adjusted using the Benjamini-Hochberg False Discovery Rate correction [94].

Sequence data from 'Zymo' samples were processed using the same pipeline, but data were not clustered. Instead, merged, trimmed and chimera-cleaned data were mapped against reference gene sequences for the eight bacterial reference organisms using the software package CLC genomics workbench v10 (Qiagen, Aarhus, Denmark). The eight reference organisms included: *Lactobacillus fermentum* (AJ575812), *Bacillus subtilis* (DQ993674), *Escherichia coli* (J01859), *Enterococcus faecalis* (EU887827), *Salmonella enterica* (JQ694167), *Listeria monocytogenes* (M58822), *Pseudomonas aeruginosa* (LN874213), and *Staphylococcus aureus* (L37597). Mapping data were then converted to biological observation matrices for use in visualization and statistical analyses, as described above. Ideal scores were calculated according to the formula described previously [18]. Briefly, the Ideal score is a summation of the absolute difference between the expected relative abundance and the observed relative abundance for each feature in a multi-feature dataset.

Data sharing

Raw sequence data files were submitted in the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI). The BioProject identifier for the samples is PRJNA492144. Full metadata for each sample are provided in **Table S7**.

Results

Primer design

The design of new primers with lower variation in melting temperature was performed with two constraints, including: (a) no primer shorter than 16 bases, and (b) all degenerate positions are retained. Outside of these constraints, we aimed to minimize theoretical melting temperature variation within each degenerate primer pool. The high overall GC content and 3' run of A or T bases in the 515F primer limited our ability to adjust Tm of this primer while meeting the above constraints. An underlying assumption of the primer modification was that shortening the primers could only increase the range of potential targets [44]. To adjust primer Tm, bases were sequentially removed from either or both 5' and 3' ends of each oligonucleotide (**Figure 13a; Table S5**). The final primer design ("ShortEMP") yielded pools of primers with overall lower average theoretical melting temperatures, and with smaller range in Tm between all primers within the pool compared to the original pool ("EMP") (**Figure 13b; Table S5**). The 515F primer pool, comprised of 4 primers, was only modestly affected by the primer alterations. The Tm range prior to modification was 66.9 to 71.8°C; after modification, the Tm range was 66.9 to 69.1°C. The Tm range of the modified primer pool was not significantly different from that of the original pool (Wilcoxon-Mann-Whitney (WMW) test; p=0.37). The 806R primer pool,

comprised of 24 primers was more strongly affected by the alterations. The Tm range prior to modification was 54.7 to 61.7°C; after modification, the Tm range was 53.6 to 56.6°C. The Tm range of the modified primer pool was significantly different from that of the original pool (WMW test; $p<0.00001$).

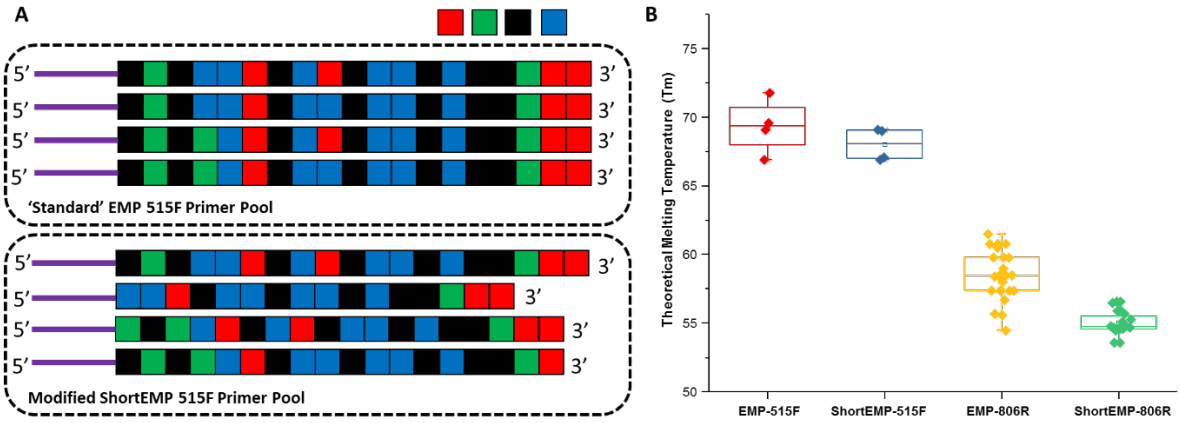


Figure 13: Schematic of primer design and primer theoretical melting temperature distribution

(A) Standard degenerate primer pools are synchronized and have low nucleotide diversity when sequenced on Illumina sequencers. Nucleotides were removed from the 5' ends of locus-specific portions of oligonucleotide primers to adjust melting temperature and to introduce nucleotide diversity. Nucleotides were removed from the 3' ends of locus-specific portions of primers to adjust melting temperature only. Sequencing reactions are initiated using the 'common sequences' (purple lines) adjacent to the locus-specific regions of the PCR primers. **(B)** Distribution of theoretical melting temperatures (Tm) for primer pools using standard EMP primers and modified ShortEMP primers. Modified 515F primer pool Tm distribution is not significantly different from the standard 515F primer pool (Wilcoxon-Mann-Whitney test;

p=0.37), while modified 806R primer pool Tm distribution is significantly different from the standard 806R primer pool (WMW test; $p<0.00001$).

A secondary aim of design was to introduce nucleotide diversity at the 5' end of the combined oligonucleotides by preferential removal of bases as the 5' end of the gene-specific portion of each primer (**Figure 13a; Tables S1, S4, S5**). Illumina's technical notes indicate that having relatively even proportions of all four nucleotides during each sequencing cycle is necessary for proper sequencing. This is particularly true during the first cycles of sequencing, when the sequencer is still training itself to identify clusters. To improve sequence quality of so-called "low diversity" libraries (*e.g.*, 16S rRNA gene amplicon libraries), exogenous spike-ins of shotgun DNA libraries derived from the virus phiX174 are typically employed. Elsewhere, Lundberg et al. [90] introduced nucleotide diversity through the synthesis of primers containing frameshift nucleotides. Briefly, a mixture of six different forward and six different reverse primers were pooled together [90]. Each primer variant contained an identical gene-specific degenerate primer at the 3' end of the oligonucleotides, but a variable number of nucleotides upstream of the gene-specific portion of the primers. This variable number of upstream nucleotides introduces artificial diversity into the sequencing reaction due to an effective frame shift by offsetting the start of highly conserved regions of the amplicon by 1-5 bases. We aimed to achieve the same effect by removing bases from the 5' end of the gene-specific portion of the primers, thereby introducing the frame-shift effect (**Figure 13a**). To assess the effect of 5' nucleotide removals on the overall nucleotide diversity at each position, we generated two *in silico* calculations: (a) number of different nucleotides present at any given position across the first 16 nucleotides of each primer pool, and (b) the Shannon index for nucleotide diversity at

each of the first 16 nucleotides of each primer pool (**Tables S8 and S9**). The original primer pool design has very low nucleotide diversity, as the primers are perfectly synchronized and have only two or three degenerate positions. Thus, for the original 515F primer pool, all primer positions other than the 4th and 9th positions have only a single nucleotide represented. The average number of nucleotides present at any position is 1.125, and the average Shannon index (natural log) across the first 16 positions is 0.09 (**Table S8**). Conversely, after modifying the individual primers, the average nucleotide diversity across the 16 positions was 2.44, with an average Shannon index of 0.77. The maximum possible Shannon index value for 4 different nucleotides is 1.39. A similar effect was observed for the 806R primer pool (**Table S9**). Prior to redesign, the average nucleotide diversity across the first 16 positions was 1.38, compared to 3.19 after redesign. Likewise, Shannon index increased from 0.2 to 1.0 after redesign.

Microbiome profiling using EMP and ShortEMP primers

To assess the effects of primer modifications, initially a single complex genomic DNA template (LMC) was profiled across an annealing temperature gradient using EMP and ShortEMP primer sets. We previously observed a strong and consistent shift in microbial community structure associated with increasing annealing temperatures from 40°C to 55°C, with specific taxa such as Prevotellaceae increasing in relative abundance with increasing annealing temperature in mammalian fecal samples [18]. Therefore, annealing temperatures of 40°C, 45°C and 50°C were tested in this study. The single gDNA template was PCR amplified at each of the temperatures and primer sets with six technical replicates. The observed microbial community from the source gDNA template was significantly affected by primer set (EMP or ShortEMP) and by annealing temperature (**Figure 14**). Variation in microbial community structure by temperature was

described primarily by the MDS axis 1, while that of primer set was described primarily by MDS axis 2. The magnitude of the shift was small, with overall Bray-Curtis similarity of all comparisons between EMP and ShortEMP replicates >0.84 **(Figure 15A).** Within each primer set, replicates had Bray-Curtis similarity values >0.89, with ShortEMP replicates having slightly and significantly greater similarity (**Figure S7, LMC**). No significant effect of primer set and annealing temperature on alpha diversity indices (*e.g.*, richness and Shannon index) at the genus- or OTU-level were observed for LMC (**Table 8**).

Table 8: Alpha diversity indices of observed microbial communities

Sample	OTU Shannon	OTU Richness	Sample	Genus Shannon	Genus Richness
EMP_LMC_40	6.88 ± 0.05	3787.60 ± 97.70	EMP_LMC_40	4.76 ± 0.02	545.60 ± 11.35
EMP_LMC_45	6.81 ± 0.09	3738.80 ± 81.14	EMP_LMC_45	4.74 ± 0.06	546.00 ± 13.21
EMP_LMC_50	6.86 ± 0.02	3810.80 ± 57.22	EMP_LMC_50	4.77 ± 0.02	545.00 ± 10.89
ShortEMP_LMC_40	6.81 ± 0.10	3708.60 ± 77.73	ShortEMP_LMC_40	4.76 ± 0.04	534.60 ± 10.78
ShortEMP_LMC_45	6.82 ± 0.02	3780.00 ± 42.27	ShortEMP_LMC_45	4.73 ± 0.01	533.00 ± 4.18
ShortEMP_LMC_50	6.86 ± 0.06	3862.80 ± 107.70	ShortEMP_LMC_50	4.77 ± 0.03	542.20 ± 5.81
KW	p=0.134	p=0.139	KW	p=0.107	p=0.161
EMP_Soil_50	7.16 ± 0.03	2907.50 ± 99.02	EMP_Soil_50	4.82 ± 0.02	414.50 ± 10.45
ShortEMP_Soil_45	7.09 ± 0.04	2832.67 ± 65.56	ShortEMP_Soil_45	4.80 ± 0.03	418.17 ± 24.17
KW	**p=0.010**	p=0.200	KW	p=0.262	p=0.873
EMP_Feces_50	4.94 ± 0.04	563.17 ± 25.34	EMP_Feces_50	2.41 ± 0.00	50.00 ± 2.37
ShortEMP_Feces_45	4.90 ± 0.08	605.17 ± 65.49	ShortEMP_Feces_45	2.39 ± 0.05	52.50 ± 2.51
KW	p=0.200	p=0.055	KW	p=0.055	p=0.089

Shannon indices and richness were calculated at the taxonomic levels of genus and at the operational taxonomic unit (OTU97) level. For each primer set (EMP or ShortEMP), an average and standard deviation of six technical replicates is shown. Kruskal-Wallis tests [95] were performed to determine if observed diversity was significantly different between EMP and ShortEMP analyses. Diversity indices were not significantly different between primer sets with the exception of OTU-level Shannon index for the Soil sample.

At each annealing temperature, relatively few taxa (16-28 genus-level taxa) were significantly differently abundant between the EMP and ShortEMP analyses (**Figures 14B, S9**).

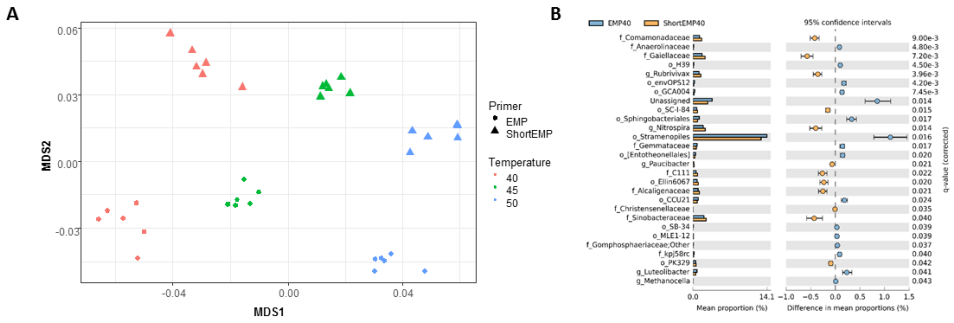


Figure 14: Effect of annealing temperature and primer set on observed microbial community in Lake Michigan sediment

A composite sample of gDNA from Lake Michigan sediment was amplified using EMP or ShortEMP primer pools at annealing temperatures of 40, 45 and 50°C and sequenced on an Illumina MiniSeq instrument. Sequence data were rarefied to a depth of 17,500 sequences per sample. (**A)** Genus-level annotations of sequence data were visualized using metric multidimensional scaling (mMDS) employing a distance matrix based on Bray-Curtis similarity. Across all temperatures, the use of EMP and ShortEMP primers resulted in slightly, but significantly, different observed microbial communities (ANOSIM R=0.473; $p<0.001$; MDS axis 2). Increasing annealing temperature also led to significant changes in observed microbial community structure (ANOSIM R=0.694, $p<0.001$; MDS axis 1). Observed microbial community structures at all temperature and primer set combinations were significantly different (ANOSIM R values 0.748-1.0; $p<0.002$). (**B**) A taxon-by-taxon analysis was performed to identify taxa with significantly different relative abundance between EMP and ShortEMP primer sets. Shown is the comparison between EMP and ShortEMP primers at 40°C annealing temperature; comparisons at 45°C and 50°C are shown in **Figure S9**. Genus-level annotations

are shown (when available), and the mean relative abundance (six technical replicates) for each primer set is shown, together with the difference in mean proportions. For each comparison a q-value, calculated in the software package STAMP using White's non-parametric t-test along with a Benjamini-Hochberg FDR correction, is shown. Only significantly differently abundant taxa ($q<0.05$) are shown. Sequences annotated as Stramenopiles are derived from SSU rRNA genes of chloroplasts from these organisms.

Based on the results of the temperature gradient analysis (above) and testing of a mock community standard (below), an annealing temperature of 45°C was chosen for additional analyses employing the ShortEMP primers and 50°C for EMP primers. Two additional complex microbial samples were analyzed at these temperatures, with six technical replicates for each. These samples include 'feces' and 'soil'. In each of these analyses, a significant effect of primer set was observed (**Figure 15B, C**; ANOSIM $R=1$, $p<0.008$), and in each case, the magnitude of the effect was similar, with overall Bray-Curtis similarity between all replicates of the same sample >0.82 (**Figure 15B, C**). Small effects of primer set on alpha diversity indices were observed, and only the OTU-level Shannon index for the soil sample was significantly different between ShortEMP primers relative to EMP primers (7.09 vs 7.16) (**Table 8)**.

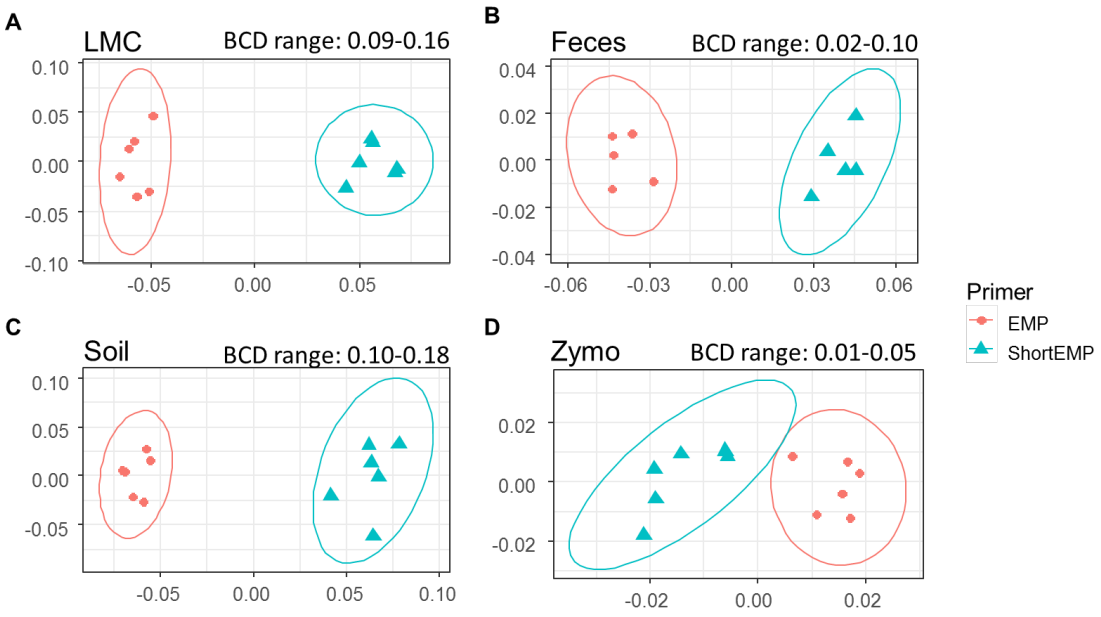


Figure 15: Effect of primer set on observed microbial communities in complex microbial samples

Genus-level annotations of sequence data sequenced on an Illumina MiSeq instrument were visualized using mMDS ordination employing a distance matrix based on Bray-Curtis similarity. X-axes represent MDS axis 1 and y-axes represent MDS axis 2 for all sample types. Sequence data were rarefied to different depths for each sample type (**Table S7**). For each sample, six technical replicates were performed at optimal annealing temperatures of 45°C (ShortEMP) and 50°C (EMP). Small, but significant, shifts in microbial communities were observed between EMP and ShortEMP primers for (a) Lake Michigan sediment (LMC), ANOSIM R=1, P=0.0021, (b) rat feces (Feces), ANOSIM R=1, P=0.0027, (c) garden soil (Soil), ANOSIM R=1, P=0.0033, and ZymoBIOMICS Microbial Community DNA standard (Zymo), ANOSIM R=1, P=0.0019. The range of Bray-Curtis dissimilarity (BCD) between EMP and ShortEMP technical replicates is shown above each figure. Ellipses represent a 95% confidence interval around the centroid. Two outliers from the fecal analysis were removed (see **Figure S8**).

Interrogation of a mock community with EMP, ShortEMP, LongEMP and NoLinker_ShortEMP primer sets

The ZymoBIOMICS Microbial Community DNA standard ("Zymo") was used as a gDNA template to assess the capability of the various primer sets to characterize microbial communities. The Zymo standard is composed of 8 bacterial taxa and 2 fungal taxa at varying levels of abundance (fungi are not amplified by the 515F/806R primer set). When analyzed using 16S rRNA gene amplicon sequencing, the relative abundance of each strain should range from 4.2% to 18.4%. Both primer sets (EMP and ShortEMP) generated highly similar but significantly distinct results (**Figure 15D**; ANOSIM R=1, p=0.0019). For each replicate, an Ideal Score [18] was calculated based on the expected relative abundance of each taxon. The Ideal Score represents a summation of the difference in relative abundance of each taxon in a mock community. A perfect representation of the expected relative abundance of each taxon will yield an ideal score of zero, with values greater than zero representing increasing discordance from the expected results. Ideal score results were generated for each of the six technical replicates and compared at annealing temperatures of 40°C, 45°C and 50°C for EMP and ShortEMP primers (**Table 9**). When employing the EMP primers for analysis of the Zymo standard, annealing temperatures below 50°C produced progressively worse results (*i.e.*, higher ideal scores). When employing the ShortEMP primers for analysis of the Zymo standard, annealing temperatures of 45°C and 50°C yielded similar results, with low ideal scores relative to 40°C (**Table 9**).

Ideal scores indicated that when used at their optimal annealing temperature EMP primers were slightly, but significantly, better than ShortEMP primers in recovering the expected distribution of the Zymo bacterial taxa. Although the EMP primers slightly outperformed the ShortEMP primers when compared to the expected distribution of eight bacterial taxa, the ShortEMP

primers better tolerated a broader range of annealing temperatures. ShortEMP primers produced nearly identical results at 45°C and 50°C (**Table 9**), demonstrating that the lower overall Tm of the primer pool can contribute to a shift in optimum annealing temperature.

Table 9: Ideal score calculations for analysis of the 'Zymo' Standard

Annealing Temp.	EMP (SD)	ShortEMP (SD)	KW
40°C	30.33 (1.75)	25.10 (6.29)	0.054
45°C	16.03 (1.54)	12.95 (0.98)	0.006
50°C	8.95 (1.52)	13.33 (1.49)	0.006
KW	0.001	0.003	
ShortEMP 45°C vs EMP 50°C	**EMP50 (SD)**	**ShortEMP45 (SD)**	**KW**
	8.95 (1.52)	12.95 (0.98)	0.006

The Ideal score is a summation of the absolute difference between the expected relative abundance and the observed relative abundance for each feature in a multi-feature dataset. Lower Ideal scores indicate a better representation of the expected community. Kruskal-Wallis test [95] p-values are shown for comparisons across temperatures or between EMP and ShortEMP primers at single temperatures. SD = standard deviation.

We sought to assess whether the addition of spacer regions, as described previously by Lundberg et al. [90], would alter the efficiency of amplification in this system. We synthesized primers containing common sequence linkers, a 0-5 base frameshift sequence, a two-base 'linker', and the EMP primer sequence (**Table S6;** 'LongEMP'). A pool of 6 forward primers and 6 reverse primers were used in the standard two-stage PCR protocol to amplify the mock community gDNA. The observed mock community generated from LongEMP analyses was most similar, and not significantly different, from that generated using the EMP primer set (**Figure 16;**

ANOSIM R=-0.179, p=0.965). EMP and LongEMP primer sets generated more similar observed microbial communities, relative to the ShortEMP primer set (ANOSIM R=0.715-0.722, p=0.002). Although significant, the difference was not large, and EMP, ShortEMP and LongEMP generated highly similar results when applied to the mock community DNA standard (**Figure 16**).

We also sought to assess whether the common sequencer linkers themselves contributed to distortion of the underlying microbial community structure. Conceptually, the 3' ends of the *linker* sequences could interact with genomic DNA templates, leading to preferential amplification of some templates. To avoid initial interaction of linker sequences with gDNA templates, we performed the first stage PCR amplification with ShortEMP primers *without* common sequence linkers (NoLinker_ShortEMP).

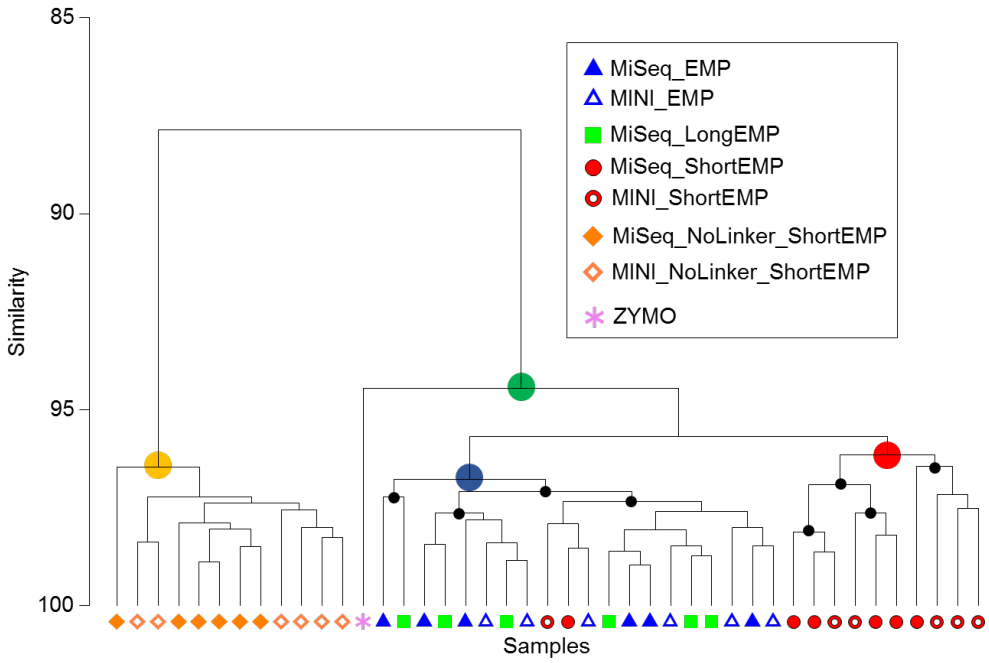


Figure 16: Comparison of representation of mock community standard using all primer sets and sequencing platforms

A single mock community standard ('Zymo') consisting of 8 bacterial taxa was amplified using EMP, ShortEMP, LongEMP and NoLinker_ShortEMP primers using both MiniSeq and MiSeq platforms and a dendrogram representing Bray-Curtis similarity is shown. Rarefaction was performed to a depth of 4,000 sequences per sample. The similarity profile routine (SIMPROF) test was performed to identify clusters with non-random structure. Nodes with a non-random structure (significance level of 5%) are indicated with black or colored circles. The ideal representation of the standard is shown as a pink asterisk. The observed community generated using NoLinker_ShortEMP primers (node with yellow circle) was significantly divergent from those observed with all other primer sets (ANOSIM R=1, p=0.002) and from the Zymo standard, and a significant effect of sequencing platform was observed with these primers (R=0.416, p=0.002, 462 permutations) The microbial communities observed with EMP and LongEMP primers were most similar to each other (node with blue circle). No significant effect of sequencing platform was observed with EMP primers (R=0.166, p=0.091) or ShortEMP primers (node with red circle; ANOSIM R=0.05, p=0.281). The microbial communities observed with EMP, LongEMP and ShortEMP clustered together with the Zymo standard (node with green circle). No data are shown for LongEMP with the MiniSeq platform, due to incomplete merging of paired-end reads.

After 28 cycles of amplification, the generated PCR amplicons were transferred to the second stage PCR amplification, as performed for all other reactions. To allow these amplicons without linkers to be prepared for Illumina sequencing, both ShortEMP primers (with common sequence linkers) and Fluidigm Access Array barcoding primers were added to the second stage reaction. Final amplicons were sequenced and the resulting sequence data analyzed together with EMP,

ShortEMP and LongEMP primer sets (**Figure 16**). The observed community generated by the NoLinker_ShortEMP primer set was significantly different than those generated by EMP, ShortEMP and LongEMP primer sets (ANOSIM R=1, p=0.002), and the observed community structure was also more divergent from the expected structure than those generated with the other primer sets (**Figure 16**). Thus, the common sequence linkers do not appear to substantially alter the observed mock microbial community structure, and removing the linkers leads to a more complex workflow and a poorer representation of the mock community.

We should note that the Zymo standard is not an ideal mock community for assessing the action of a degenerate primer pool. We sought to identify the sequences of the eight bacterial DNA templates in the region of the 515F and 806R primers, and despite the moderately broad range of bacterial taxa included (Proteobacteria, Actinobacteria, and Firmicutes), all sequences were identical at the primer annealing locations. This mock community is, therefore, appropriate for determining that an overall PCR and sequencing workflow is successful but is not a good approximation of a highly complex natural microbial community, where even conserved sites such as the 515F and 806R primer sites contain considerable sequence heterogeneity.

We also used the amplicons generated from the Zymo standard to assess variability introduced by sequencing on the Illumina MiSeq or MiniSeq platforms. For the EMP, ShortEMP, and NoLinker_ShortEMP, identical amplicons were sequenced and analyzed on both platforms. The LongEMP amplicon reads generated on the Illumina MiniSeq did not consistently merge, and were not included in the analysis. This is due to the additional ‘spacer’ bases (up to 6 on each read) and the short overlap of paired-end reads generated with 2x153 bases reads on the MiniSeq. MiSeq and MiniSeq results from the same amplicons clustered together, regardless of primer set (**Figure 16**). For the EMP and ShortEMP primer sets, the effect of sequencing platform

was not significant (ANOSIM R=0.166, p=0.091 and R=0.05, p=0.281, respectively). For the NoLinker_ShortEMP primer set, a significant effect of sequencing platform was observed (ANOSIM R=0.416, p=0.002). Thus, a slight effect of sequencing platform was observed, but the choice of platform did not alter biological conclusions.

Assessing the need for phiX spike-in with ShortEMP primers

We sequenced ShortEMP amplicons on an Illumina MiniSeq run without substantial phiX spike-in to determine if ShortEMP amplicons by themselves produced sufficient nucleotide diversity to allow proper clustering. These amplicons, generated from Lake Michigan sediment, at annealing temperatures of 40, 45 and 50, were analyzed together with the same amplicons generated on a MiniSeq run with a 20% phiX spike-in. The overall quality of the run was extremely high (>96% pass-filter, >97% Q30 with approximately 1.75% phiX spike-in), and results were similar to those generated on a 20% phiX run (**Figure 17**). Slight trends towards small, significant differences were observed between the two sequencing runs, but the overall magnitude of the difference was small.

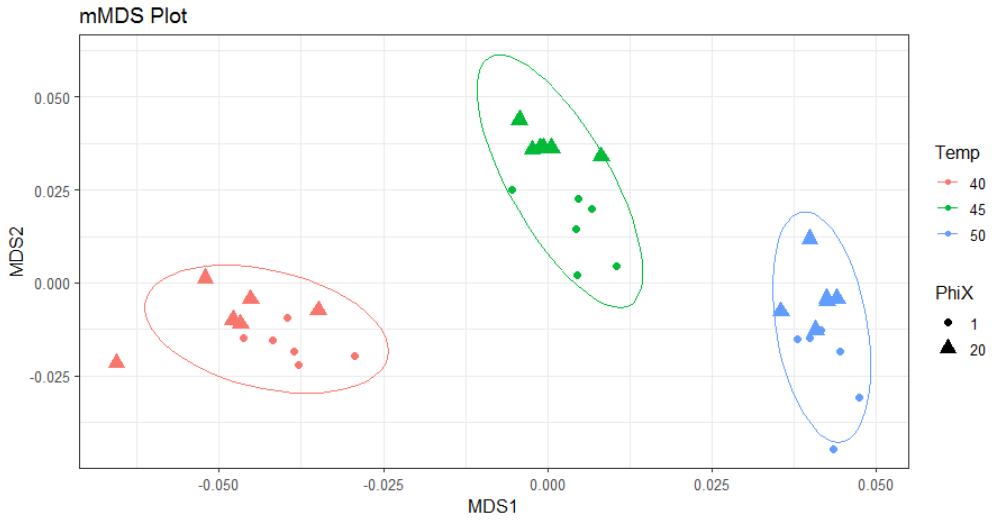


Figure 17: Effect of phiX spike-in on observed microbial communities with amplicons generated using ShortEMP primers

Genus-level annotations of sequence data generated on an Illumina MiniSeq instrument were visualized using mMDS ordination employing a distance matrix based on Bray-Curtis similarity. Rarefaction was performed to a depth of 30,000 sequences per sample. For each ShortEMP PCR condition, six technical replicates were analyzed using either ~20% or ~1% phiX spike-in. Small differences, consistent with run-to-run variation, were observed between the two sequencing runs (ANOSIM: 40°C R=0.146, p=0.052; 45°C R=0.394, p=0.011; 50°C R=0.178, p=0.100). Ellipses represent a 95% confidence interval around the centroid.

Discussion

We demonstrate here that the strategy to remove nucleotides from the 5' ends of individual primers is effective for introducing nucleotide diversity into a primer pool, and for reducing the Tm range of the primers within a degenerate primer pool. Selective removal of bases at the 3' end can be used together with 5' base removal to adjust overall primer melting temperature. Base removal at the 3' end, when using the Fluidigm sequencing protocol, does not impact nucleotide diversity for the purposes of Illumina sequencing. In this study, to ensure that the greatest nucleotide diversity was present during the initial cycles of the first sequencing reaction, the 'CS1' linker sequence was attached to the 806R primer. This approach is inverted compared to standard EMP workflows, where the CS1 linker is attached to the 515F primer [38]. For future designs, the choice of linker can be adjusted as needed to ensure the highest nucleotide diversity during the initial cycles of the first read of Illumina sequencers, when cluster identification is

performed. When amplicons generated with the modified primers were sequenced on an Illumina MiniSeq sequencer with <2% phiX, the data quality and community analyses were consistent with the same amplicons generated on a run with 20% phiX. Therefore, we conclude that the nucleotide diversity created by removing bases at the 5' ends of primers to effect the frameshift was sufficient to allow for sequencing without substantial phiX inclusion.

We sought to determine whether the broad range of primer melting temperatures in degenerate primer pools contributes substantially to observed microbial community profiles. Individual primer modifications were used to reduce theoretical melting temperature variation between primers with a degenerate pool, and the modified primers were used to amplify mock and environmental samples. When compared to amplicons generated using standard primer sets, the reducing melting temperature variability in the modified primer pools did not substantially alter observed microbial community structure in the tested samples. The modified primers had limited or no effect on measured alpha diversity in complex microbial samples and shifts in microbial community structure associated with the modified primers relative to the standard primers were small in scale. The shortened primers appear to have greater tolerance for lower annealing temperatures than the standard primers. We considered the possibility that high Tm primers could dominate primer-template interactions, leading to variable amplification efficiencies in PCR and reduced taxonomic coverage. Our results, however, do not support any substantial modification in the target range for the modified primers, even at low annealing temperatures. We do note, however, that the 515F and 806R primers themselves have highly divergent theoretical melting temperatures, and that this discrepancy could contribute to the difficulty in expanding the targeted taxonomic range of PCR-based microbiome sequencing.

Conclusions

We demonstrate here a novel method to introduce nucleotide diversity into PCR amplicons for sequencing on Illumina sequencers. Through selective removal of bases at the 5' end of oligonucleotide primers, nucleotide diversity can be introduced without substantial effect on the activity of the primers themselves. When employing this strategy with a commonly-used primer set targeting microbial SSU rRNA genes, no substantial effects on observed microbial community structures were observed.

Acknowledgements

We kindly thank Maryam Elfeki and Brian Murphy for the supply of lake sediment genomic DNA and Katherine Mak for the supply of rat fecal genomic DNA.

Supporting Information

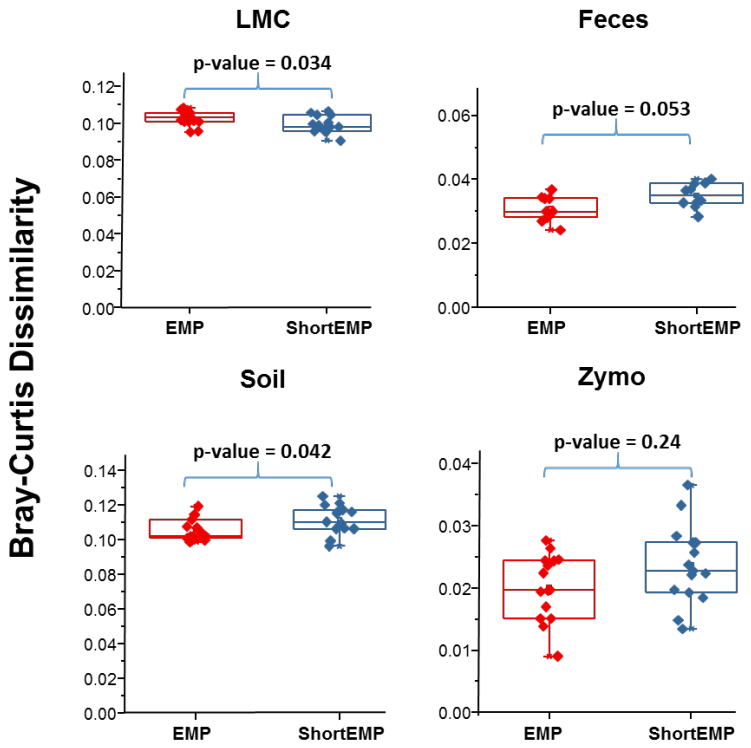


Figure S7: Box plots of within-group Bray-Curtis dissimilarity scores for microbiome analyses conducted with EMP and ShortEMP primer sets.

For each sample, Bray-Curtis dissimilarity was calculated for 6 technical replicates with EMP primers (15 comparisons), and 6 technical replicates with ShortEMP primers (15 comparisons). A comparison of median within-sample similarity for replicates from EMP and ShortEMP amplifications was performed, and were significantly different for LMC, Feces and Soil (Mann-Whitney test, $P<0.053$). An outlier of one replicate from both EMP and ShortEMP fecal analyses was removed (see **Figure S8**).

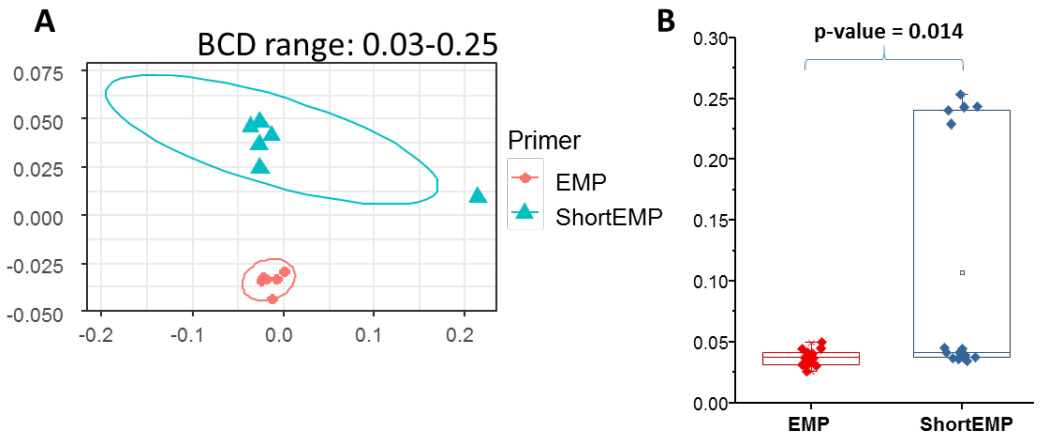
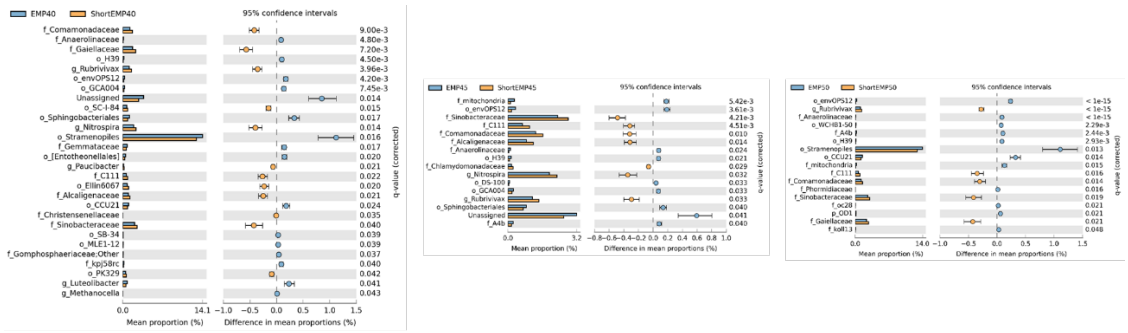


Figure S8: Analysis of fecal sample replicates including outliers.

A single technical replicate, representing an outlier, was removed from both EMP and ShortEMP analyses (**Figures 15 and S7**). Analyses in this figure are shown including the deep outliers. Inclusion of the outliers does not modify the conclusions of the analysis. However, the ShortEMP outlier is greatly different from all other technical replicates from all samples in the study. (**A**) Genus-level annotations of sequence data were visualized using mMDS ordination employing a distance matrix based on Bray-Curtis similarity. Six technical replicates were performed at optimal annealing temperatures of 45°C (ShortEMP) and 50°C (EMP). Small, but significant, shifts in microbial communities were observed between EMP and ShortEMP primers for Feces (ANOSIM R=0.68, P=0.0025). Bray-Curtis dissimilarity (BCD) between EMP and ShortEMP technical replicates is shown above the figure. Ellipses represent a 95% confidence interval around the centroid. **(B)** Box plot of within-group Bray-Curtis dissimilarity scores for microbiome analyses conducted with EMP and ShortEMP primer sets on fecal DNA.



40°C Annealing Temperature

45°C Annealing Temperature

50°C Annealing Temperature

Figure S9: Significant differences in taxon relative abundance between EMP and ShortEMP analyses of Lake Michigan Sediment at varying annealing temperatures.

Genus-level annotations are shown (when available), and the mean relative abundance (six technical replicates) for each primer set is shown, together with the difference in mean proportions. For each comparison a q-value, calculated in the software package STAMP using White's non-parametric t-test along with a Benjamini-Hochberg FDR correction, is shown. Only significantly differently abundant taxa ($q<0.05$) are shown. Sequences annotated as Stramenopiles are derived from SSU rRNA genes of chloroplasts from these organisms.

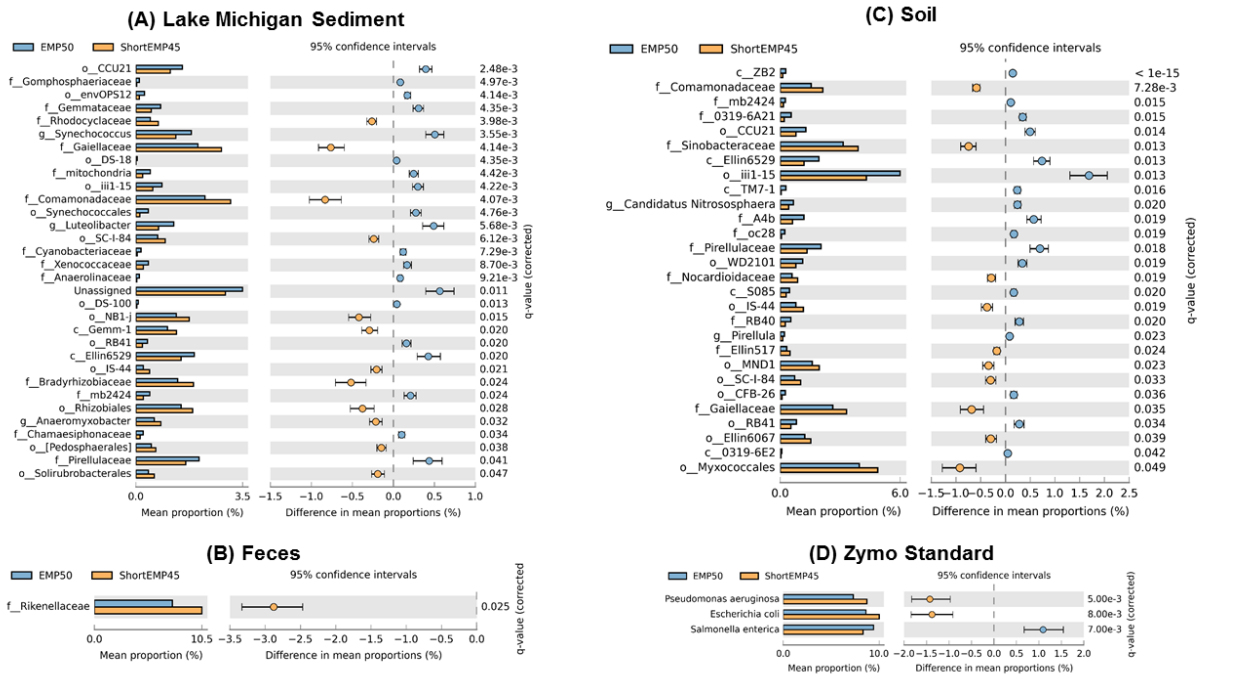


Figure S10: Significant differences in taxon relative abundance between EMP and ShortEMP analyses for LMC, Feces, Soil and Zymo samples.

Genus-level annotations (except Zymo) are shown (when available), and the mean relative abundance (six technical replicates) for each primer set is shown, together with the difference in mean proportions. For each comparison a q-value, calculated in the software package STAMP using White's non-parametric t-test along with a Benjamini-Hochberg FDR correction, is shown. Only significantly differently abundant taxa ($q<0.05$) are shown. Zymo sequences were annotated to the taxonomic level of species by mapping to known references.

Table S5: Original (EMP) and modified (ShortEMP) primer variants for 515F and 806R primers

	ORIGINAL				VARIANTS																								
515F Variants	**Tm (°C)**	**dG (kcal/mole)**	**Sequence**	**Length**	**New_Tm (°C)**	**New_dG (kcal/mole)**				**T/C**						**A/C**													**Degeneracies**
515Fb	**67 - 72**	**-43.55**	**GTGYCAGCMGCCGCGGTAA**	**19**	**67 - 69**	**-41.12**	**G**	**T**	**G**	**Y**	**C**	**A**	**G**	**C**	**M**	**G**	**C**	**C**	**G**	**C**	**G**	**G**	**T**	**A**	**A**			**Y M**
515Fb_V1	69.6	-43.64	GTGTCAGCAGCCGCGGTAA	**19**	66.9	-40.34	G	T	G	T	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A			T A
515Fb_V2	71.8	-46.76	CCAGCCGCCGCGGTAA	**16**	67.1	-40.33				C	C	A	G	C	C	G	C	C	G	C	G	G	T	A	A			C C
515Fb_V3	66.9	-40.34	TGCCAGCAGCCGCGGTAA	**18**	69.1	-42.29		T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A			C A
515Fb_V4	69.1	-43.47	GTGTCAGCCGCCGCGGTA	**18**	69	-41.53	G	T	G	T	C	A	G	C	C	G	C	C	G	C	G	G	T	A				T C
806R Variants														**A/C/G/T**	**A/C/G**						**A/T**							
806Rb	**54.7 - 61.7**	**-35.55**	**GGACTACNVGGGTWTCTAAT**	**20**	**53.6 - 56.6**	**-31.25**	**G**	**G**	**A**	**C**	**T**	**A**	**C**	**N**	**V**	**G**	**G**	**G**	**T**	**W**	**T**	**C**	**T**	**A**	**A**	**T**		**N V W**
806Rb_V1	55.7	-32.85	GGACTACAAGGGTATCTAA	**19**	54.9	-31.38	G	G	A	C	T	A	C	A	A	G	G	G	T	A	T	C	T	A	A			A A A
806Rb_V2	57.4	-33.98	GACTACCAGGGTATCTAAT	**19**	54.7	-30.91		G	A	C	T	A	C	C	A	G	G	G	T	A	T	C	T	A	A	T		C A A
806Rb_V3	57.4	-34.14	GGACTACGAGGGTATCTA	**18**	55.9	-30.72	G	G	A	C	T	A	C	G	A	G	G	G	T	A	T	C	T	A				G A A
806Rb_V4	54.5	-31.52	GGACTACTAGGGTATCTAAT	**20**	54.5	-31.52	G	G	A	C	T	A	C	T	A	G	G	G	T	A	T	C	T	A	A	T		T A A
806Rb_V5	58	-34.26	GACTACACGGGTATCTAAT	**19**	55.3	-31.2		G	A	C	T	A	C	A	C	G	G	G	T	A	T	C	T	A	A	T		A C A
806Rb_V6	59.8	-37.11	CTACCCGGGTATCTAAT	**17**	53.6	-31.12				C	T	A	C	C	C	G	G	G	T	A	T	C	T	A	A	T		C C A
806Rb_V7	60.5	-37.72	CTACGCGGGTATCTAAT	**17**	54.7	-31.73				C	T	A	C	G	C	G	G	G	T	A	T	C	T	A	A	T		G C A
806Rb_V8	57.4	-34.14	GGACTACTCGGGTATCTA	**18**	55.9	-30.72	G	G	A	C	T	A	C	T	C	G	G	G	T	A	T	C	T	A				T C A
806Rb_V9	57.4	-33.98	GACTACAGGGGTATCTAAT	**19**	54.7	-30.91		G	A	C	T	A	C	A	G	G	G	G	T	A	T	C	T	A	A	T		A G A
806Rb_V10	59.8	-37.11	GGACTACCGGGGTATC	**16**	56.6	-31.13	G	G	A	C	T	A	C	C	G	G	G	G	T	A	T	C						C G A
806Rb_V11	59.8	-37.11	CTACGGGGGTATCTAAT	**17**	53.6	-31.12				C	T	A	C	G	G	G	G	G	T	A	T	C	T	A	A	T		G G A
806Rb_V12	57.4	-33.98	GACTACTGGGGTATCTAAT	**19**	54.7	-30.91		G	A	C	T	A	C	T	G	G	G	G	T	A	T	C	T	A	A	T		T G A
806Rb_V13	56.7	-34.31	GGACTACAAGGGTTTCTA	**18**	55.1	-30.89	G	G	A	C	T	A	C	A	A	G	G	G	T	T	T	C	T	A				A A T
806Rb_V14	58.5	-35.43	ACTACCAGGGTTTCTAAT	**18**	54.6	-30.79			A	C	T	A	C	C	A	G	G	G	T	T	T	C	T	A	A	T		C A T
806Rb_V15	58.4	-35.6	ACTACGAGGGTTTCTAAT	**18**	54.6	-30.95			A	C	T	A	C	G	A	G	G	G	T	T	T	C	T	A	A	T		G A T
806Rb_V16	55.6	-32.97	GGACTACTAGGGTTTCTAA	**19**	54.8	-31.49	G	G	A	C	T	A	C	T	A	G	G	G	T	T	T	C	T	A	A			T A T
806Rb_V17	59	-35.72	GACTACACGGGTTTCTAA	**18**	55.7	-31.17		G	A	C	T	A	C	A	C	G	G	G	T	T	T	C	T	A	A			A C T
806Rb_V18	60.8	-38.56	ACTACCCGGGTTTCTAA	**17**	56.5	-32.44			A	C	T	A	C	C	C	G	G	G	T	T	T	C	T	A	A			C C T
806Rb_V19	61.5	-39.17	ACTACGCGGGTTTCTA	**16**	56.6	-31.11			A	C	T	A	C	G	C	G	G	G	T	T	T	C	T	A				G C T
806Rb_V20	58.4	-35.6	ACTACTCGGGTTTCTAAT	**18**	54.6	-30.95			A	C	T	A	C	T	C	G	G	G	T	T	T	C	T	A	A	T		T C T
806Rb_V21	58.5	-35.43	ACTACAGGGGTTTCTAAT	**18**	54.6	-30.79			A	C	T	A	C	A	G	G	G	G	T	T	T	C	T	A	A	T		A G T
806Rb_V22	60.8	-38.56	CTACCGGGGTTTCTAAT	**17**	54.8	-32.57				C	T	A	C	C	G	G	G	G	T	T	T	C	T	A	A	T		C G T
806Rb_V23	60.8	-38.56	CTACGGGGGTTTCTAAT	**17**	54.8	-32.57				C	T	A	C	G	G	G	G	G	T	T	T	C	T	A	A	T		G G T
806Rb_V24	58.5	-35.43	ACTACTGGGGTTTCTAAT	**18**	54.6	-30.79			A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T		T G T

Table S6: Primer sequences used in this study

ShortEMP Pool Primers (CS1_806R_pool and CS2_515F_pool)

515Fb variants	Modified Gene-Specific Portion of Primer Sequence	Length	Linker (CS2) Sequence	Final Sequence Name	Final Sequence Ordered from IDT	Final Length
515Fb_V1	GTGTCAGCAGCCGCGGTAA	19	**TACGGTAGCAGAGACTTGGTCT**	>CS2_515Fb_V1	**TACGGTAGCAGAGACTTGGTCT**GTGTCAGCAGCCGCGGTAA	41
515Fb_V2	CCAGCCGCCGCGGTAA	16	**TACGGTAGCAGAGACTTGGTCT**	>CS2_515Fb_V2	**TACGGTAGCAGAGACTTGGTCT**CCAGCCGCCGCGGTAA	38
515Fb_V3	TGCCAGCAGCCGCGGTAA	18	**TACGGTAGCAGAGACTTGGTCT**	>CS2_515Fb_V3	**TACGGTAGCAGAGACTTGGTCT**TGCCAGCAGCCGCGGTAA	40
515Fb_V4	GTGTCAGCCGCCGCGGTA	18	**TACGGTAGCAGAGACTTGGTCT**	>CS2_515Fb_V4	**TACGGTAGCAGAGACTTGGTCT**GTGTCAGCCGCCGCGGTA	40
806Rb variants	**Modified Gene-Specific Portion of Primer Sequence**	**Length**	**Linker (CS1) Sequence**	**Final Sequence Name**	**Final Sequence Ordered from IDT**	**Final Length**
806Rb_V1	GGACTACAAGGGTATCTAA	19	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V1	**ACACTGACGACATGGTTCTACA**GGACTACAAGGGTATCTAA	41
806Rb_V2	GACTACCAGGGTATCTAAT	19	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V2	**ACACTGACGACATGGTTCTACA**GACTACCAGGGTATCTAAT	41
806Rb_V3	GGACTACGAGGGTATCTA	18	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V3	**ACACTGACGACATGGTTCTACA**GGACTACGAGGGTATCTA	40
806Rb_V4	GGACTACTAGGGTATCTAAT	20	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V4	**ACACTGACGACATGGTTCTACA**GGACTACTAGGGTATCTAAT	42
806Rb_V5	GACTACACGGGTATCTAAT	19	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V5	**ACACTGACGACATGGTTCTACA**GACTACACGGGTATCTAAT	41
806Rb_V6	CTACCCGGGTATCTAAT	17	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V6	**ACACTGACGACATGGTTCTACA**CTACCCGGGTATCTAAT	39
806Rb_V7	CTACGCGGGTATCTAAT	17	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V7	**ACACTGACGACATGGTTCTACA**CTACGCGGGTATCTAAT	39
806Rb_V8	GGACTACTCGGGTATCTA	18	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V8	**ACACTGACGACATGGTTCTACA**GGACTACTCGGGTATCTA	40
806Rb_V9	GACTACAGGGGTATCTAAT	19	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V9	**ACACTGACGACATGGTTCTACA**GACTACAGGGGTATCTAAT	41
806Rb_V10	GGACTACCGGGGTATC	16	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V10	**ACACTGACGACATGGTTCTACA**GGACTACCGGGGTATC	38
806Rb_V11	CTACGGGGGTATCTAAT	17	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V11	**ACACTGACGACATGGTTCTACA**CTACGGGGGTATCTAAT	39
806Rb_V12	GACTACTGGGGTATCTAAT	19	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V12	**ACACTGACGACATGGTTCTACA**GACTACTGGGGTATCTAAT	41
806Rb_V13	GGACTACAAGGGTTTCTA	18	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V13	**ACACTGACGACATGGTTCTACA**GGACTACAAGGGTTTCTA	40
806Rb_V14	ACTACCAGGGTTTCTAAT	18	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V14	**ACACTGACGACATGGTTCTACA**ACTACCAGGGTTTCTAAT	40
806Rb_V15	ACTACGAGGGTTTCTAAT	18	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V15	**ACACTGACGACATGGTTCTACA**ACTACGAGGGTTTCTAAT	40
806Rb_V16	GGACTACTAGGGTTTCTAA	19	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V16	**ACACTGACGACATGGTTCTACA**GGACTACTAGGGTTTCTAA	41
806Rb_V17	GACTACACGGGTTTCTAA	18	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V17	**ACACTGACGACATGGTTCTACA**GACTACACGGGTTTCTAA	40
806Rb_V18	ACTACCCGGGTTTCTAA	17	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V18	**ACACTGACGACATGGTTCTACA**ACTACCCGGGTTTCTAA	39
806Rb_V19	ACTACGCGGGTTTCTA	16	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V19	**ACACTGACGACATGGTTCTACA**ACTACGCGGGTTTCTA	38
806Rb_V20	ACTACTCGGGTTTCTAAT	18	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V20	**ACACTGACGACATGGTTCTACA**ACTACTCGGGTTTCTAAT	40
806Rb_V21	ACTACAGGGGTTTCTAAT	18	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V21	**ACACTGACGACATGGTTCTACA**ACTACAGGGGTTTCTAAT	40
806Rb_V22	CTACCGGGGTTTCTAAT	17	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V22	**ACACTGACGACATGGTTCTACA**CTACCGGGGTTTCTAAT	39
806Rb_V23	CTACGGGGGTTTCTAAT	17	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V23	**ACACTGACGACATGGTTCTACA**CTACGGGGGTTTCTAAT	39
806Rb_V24	ACTACTGGGGTTTCTAAT	18	**ACACTGACGACATGGTTCTACA**	>CS1_806Rb_V24	**ACACTGACGACATGGTTCTACA**ACTACTGGGGTTTCTAAT	40

EMP Primers (CS1_806R and CS2_515F)

Original EMP	Original Gene-Specific Primer Sequence	Length	Linker (CS1 or CS2) Sequence	Final Sequence Name	Final Sequence Ordered from IDT	Final Length
515F_EMP	GTGYCAGCMGCCGCGGTAA	19	**TACGGTAGCAGAGACTTGGTCT**	>CS2_515F_EMP	**TACGGTAGCAGAGACTTGGTCT**GTGYCAGCMGCCGCGGTAA	41
806R_EMP	GGACTACNVGGGTWTCTAAT	20	**ACACTGACGACATGGTTCTACA**	>CS1_806R_EMP	**ACACTGACGACATGGTTCTACA**GGACTACNVGGGTWTCTAAT	42

LongEMP Primers (modified from Lunderberg et al. 2013) - CS1_806_long_pool and CS2_806R_long_pool

Long EMP	Original Gene-Specific Primer Sequence	Length	Linker (CS1 or CS2) Sequence	Final Sequence Name	Final Sequence Ordered from IDT	Final Length
Long_515Fb_V1	GAGTGYCAGCMGCCGCGGTAA	21	**TACGGTAGCAGAGACTTGGTCT**	>CS2_515F_LongEMP_V1	TACGGTAGCAGAGACTTGGTCTGAGTGYCAGCMGCCGCGGTAA	43
Long_515Fb_V2	TGAGTGYCAGCMGCCGCGGTAA	22	**TACGGTAGCAGAGACTTGGTCT**	>CS2_515F_LongEMP_V2	TACGGTAGCAGAGACTTGGTCTTGAGTGYCAGCMGCCGCGGTAA	44
Long_515Fb_V3	CTGAGTGYCAGCMGCCGCGGTAA	23	**TACGGTAGCAGAGACTTGGTCT**	>CS2_515F_LongEMP_V3	TACGGTAGCAGAGACTTGGTCTCTGAGTGYCAGCMGCCGCGGTAA	45
Long_515Fb_V4	ACTGAGTGYCAGCMGCCGCGGTAA	24	**TACGGTAGCAGAGACTTGGTCT**	>CS2_515F_LongEMP_V4	TACGGTAGCAGAGACTTGGTCTACTGAGTGYCAGCMGCCGCGGTAA	46
Long_515Fb_V5	GACTGAGTGYCAGCMGCCGCGGTAA	25	**TACGGTAGCAGAGACTTGGTCT**	>CS2_515F_LongEMP_V5	TACGGTAGCAGAGACTTGGTCTGACTGAGTGYCAGCMGCCGCGGTAA	47
Long_515Fb_V6	TGACTGAGTGYCAGCMGCCGCGGTAA	26	**TACGGTAGCAGAGACTTGGTCT**	>CS2_515F_LongEMP_V6	TACGGTAGCAGAGACTTGGTCTTGACTGAGTGYCAGCMGCCGCGGTAA	48
Long_806Rb_V1	ACGGACTACNVGGGTWTCTAAT	22	**ACACTGACGACATGGTTCTACA**	>CS1_806R_LongEMP_V1	ACACTGACGACATGGTTCTACAACGGACTACNVGGGTWTCTAAT	44
Long_806Rb_V2	TACGGACTACNVGGGTWTCTAAT	23	**ACACTGACGACATGGTTCTACA**	>CS1_806R_LongEMP_V2	ACACTGACGACATGGTTCTACATACGGACTACNVGGGTWTCTAAT	45
Long_806Rb_V3	CTACGGACTACNVGGGTWTCTAAT	24	**ACACTGACGACATGGTTCTACA**	>CS1_806R_LongEMP_V3	ACACTGACGACATGGTTCTACACTACGGACTACNVGGGTWTCTAAT	46
Long_806Rb_V4	ACTACGGACTACNVGGGTWTCTAAT	25	**ACACTGACGACATGGTTCTACA**	>CS1_806R_LongEMP_V4	ACACTGACGACATGGTTCTACAACTACGGACTACNVGGGTWTCTAAT	47
Long_806Rb_V5	GACTACGGACTACNVGGGTWTCTAAT	26	**ACACTGACGACATGGTTCTACA**	>CS1_806R_LongEMP_V5	ACACTGACGACATGGTTCTACAGACTACGGACTACNVGGGTWTCTAAT	48
Long_806Rb_V6	TGACTACGGACTACNVGGGTWTCTAAT	27	**ACACTGACGACATGGTTCTACA**	>CS1_806R_LongEMP_V6	ACACTGACGACATGGTTCTACATGACTACGGACTACNVGGGTWTCTAAT	49

NoLinker_ShortEMP Primers

515Fb variants	Modified Gene-Specific Portion of Primer Sequence	Length	Linker (CS2) Sequence	Final Sequence Name	Final Sequence Ordered from IDT	Final Length
515Fb_V1	GTGTCAGCAGCCGCGGTAA	19	**none**	>NoCS_515Fb_V1	GTGTCAGCAGCCGCGGTAA	19
515Fb_V2	CCAGCCGCCGCGGTAA	16	**none**	>NoCS_515Fb_V2	CCAGCCGCCGCGGTAA	16
515Fb_V3	TGCCAGCAGCCGCGGTAA	18	**none**	>NoCS_515Fb_V3	TGCCAGCAGCCGCGGTAA	18
515Fb_V4	GTGTCAGCCGCCGCGGTA	18	**none**	>NoCS_515Fb_V4	GTGTCAGCCGCCGCGGTA	18

806Rb variants	Modified Gene-Specific Portion of Primer Sequence	Length	Linker (CS1) Sequence	Final Sequence Name	Final Sequence Ordered from IDT	Final Length
806Rb_V1	GGACTACAAGGGTATCTAA	19	**none**	>NoCS_806Rb_V1	GGACTACAAGGGTATCTAA	19
806Rb_V2	GACTACCAGGGTATCTAAT	19	**none**	>NoCS_806Rb_V2	GACTACCAGGGTATCTAAT	19
806Rb_V3	GGACTACGAGGGTATCTA	18	**none**	>NoCS_806Rb_V3	GGACTACGAGGGTATCTA	18
806Rb_V4	GGACTACTAGGGTATCTAAT	20	**none**	>NoCS_806Rb_V4	GGACTACTAGGGTATCTAAT	20
806Rb_V5	GACTACACGGGTATCTAAT	19	**none**	>NoCS_806Rb_V5	GACTACACGGGTATCTAAT	19
806Rb_V6	CTACCCGGGTATCTAAT	17	**none**	>NoCS_806Rb_V6	CTACCCGGGTATCTAAT	17
806Rb_V7	CTACGCGGGTATCTAAT	17	**none**	>NoCS_806Rb_V7	CTACGCGGGTATCTAAT	17
806Rb_V8	GGACTACTCGGGTATCTA	18	**none**	>NoCS_806Rb_V8	GGACTACTCGGGTATCTA	18
806Rb_V9	GACTACAGGGGTATCTAAT	19	**none**	>NoCS_806Rb_V9	GACTACAGGGGTATCTAAT	19
806Rb_V10	GGACTACCGGGGTATC	16	**none**	>NoCS_806Rb_V10	GGACTACCGGGGTATC	16
806Rb_V11	CTACGGGGGTATCTAAT	17	**none**	>NoCS_806Rb_V11	CTACGGGGGTATCTAAT	17
806Rb_V12	GACTACTGGGGTATCTAAT	19	**none**	>NoCS_806Rb_V12	GACTACTGGGGTATCTAAT	19
806Rb_V13	GGACTACAAGGGTTTCTA	18	**none**	>NoCS_806Rb_V13	GGACTACAAGGGTTTCTA	18
806Rb_V14	ACTACCAGGGTTTCTAAT	18	**none**	>NoCS_806Rb_V14	ACTACCAGGGTTTCTAAT	18
806Rb_V15	ACTACGAGGGTTTCTAAT	18	**none**	>NoCS_806Rb_V15	ACTACGAGGGTTTCTAAT	18
806Rb_V16	GGACTACTAGGGTTTCTAA	19	**none**	>NoCS_806Rb_V16	GGACTACTAGGGTTTCTAA	19
806Rb_V17	GACTACACGGGTTTCTAA	18	**none**	>NoCS_806Rb_V17	GACTACACGGGTTTCTAA	18

806Rb_V18	ACTACCCGGGTTTCTAA	17	**none**	>NoCS_806Rb_V18	ACTACCCGGGTTTCTAA	17
806Rb_V19	ACTACGCGGGTTTCTA	16	**none**	>NoCS_806Rb_V19	ACTACGCGGGTTTCTA	16
806Rb_V20	ACTACTCGGGTTTCTAAT	18	**none**	>NoCS_806Rb_V20	ACTACTCGGGTTTCTAAT	18
806Rb_V21	ACTACAGGGGTTTCTAAT	18	**none**	>NoCS_806Rb_V21	ACTACAGGGGTTTCTAAT	18
806Rb_V22	CTACCGGGGTTTCTAAT	17	**none**	>NoCS_806Rb_V22	CTACCGGGGTTTCTAAT	17
806Rb_V23	CTACGGGGGTTTCTAAT	17	**none**	>NoCS_806Rb_V23	CTACGGGGGTTTCTAAT	17
806Rb_V24	ACTACTGGGGTTTCTAAT	18	**none**	>NoCS_806Rb_V24	ACTACTGGGGTTTCTAAT	18

Table S7: Mapping file metadata associated with all samples used in this study

Sequencing Run	Sample Name	Primer Set	Template	Sequencer	PhiX (%)	Annealing Temp. (°C)	Ref. Figure / Supp. Table	Rarefaction Depth
1	Ankur19-EMP-LM-40_S19	EMP	LMC	MiniSeq	20%	40	14	17500
1	Ankur20-EMP-LM-40_S20	EMP	LMC	MiniSeq	20%	40	14	17500
1	Ankur21-EMP-LM-40_S21	EMP	LMC	MiniSeq	20%	40	14	17500
1	Ankur22-EMP-LM-40_S22	EMP	LMC	MiniSeq	20%	40	14	17500
1	Ankur23-EMP-LM-40_S23	EMP	LMC	MiniSeq	20%	40	14	17500
1	Ankur24-EMP-LM-40_S24	EMP	LMC	MiniSeq	20%	40	14	17500
1	Ankur25-EMP-LM-45_S25	EMP	LMC	MiniSeq	20%	45	14	17500
1	Ankur26-EMP-LM-45_S26	EMP	LMC	MiniSeq	20%	45	14	17500
1	Ankur27-EMP-LM-45_S27	EMP	LMC	MiniSeq	20%	45	14	17500
1	Ankur28-EMP-LM-45_S28	EMP	LMC	MiniSeq	20%	45	14	17500
1	Ankur29-EMP-LM-45_S29	EMP	LMC	MiniSeq	20%	45	14	17500
1	Ankur30-EMP-LM-45_S30	EMP	LMC	MiniSeq	20%	45	14	17500
1	Ankur31-EMP-LM-50_S31	EMP	LMC	MiniSeq	20%	50	14	17500
1	Ankur32-EMP-LM-50_S32	EMP	LMC	MiniSeq	20%	50	14	17500
1	Ankur33-EMP-LM-50_S33	EMP	LMC	MiniSeq	20%	50	14	17500
1	Ankur34-EMP-LM-50_S34	EMP	LMC	MiniSeq	20%	50	14	17500
1	Ankur35-EMP-LM-50_S35	EMP	LMC	MiniSeq	20%	50	14	17500
1	Ankur36-EMP-LM-50_S36	EMP	LMC	MiniSeq	20%	50	14	17500
1	Ankur55-Shorty-LM-40_S55	ShortEMP	LMC	MiniSeq	20%	40	14, 16	17500, 30000
1	Ankur56-Shorty-LM-40_S56	ShortEMP	LMC	MiniSeq	20%	40	14, 16	17500, 30000
1	Ankur57-Shorty-LM-40_S57	ShortEMP	LMC	MiniSeq	20%	40	14, 16	17500, 30000
1	Ankur58-Shorty-LM-40_S58	ShortEMP	LMC	MiniSeq	20%	40	14, 16	17500, 30000
1	Ankur59-Shorty-LM-40_S59	ShortEMP	LMC	MiniSeq	20%	40	14, 16	17500, 30000
1	Ankur60-Shorty-LM-40_S60	ShortEMP	LMC	MiniSeq	20%	40	14, 16	17500, 30000
1	Ankur61-Shorty-LM-45_S61	ShortEMP	LMC	MiniSeq	20%	45	14, 16	17500, 30000
1	Ankur62-Shorty-LM-45_S62	ShortEMP	LMC	MiniSeq	20%	45	14, 16	17500, 30000
1	Ankur63-Shorty-LM-45_S63	ShortEMP	LMC	MiniSeq	20%	45	14, 16	17500, 30000
1	Ankur64-Shorty-LM-45_S64	ShortEMP	LMC	MiniSeq	20%	45	14, 16	17500, 30000
1	Ankur65-Shorty-LM-45_S65	ShortEMP	LMC	MiniSeq	20%	45	14, 16	17500, 30000
1	Ankur66-Shorty-LM-45_S66	ShortEMP	LMC	MiniSeq	20%	45	14, 16	17500, 30000
1	Ankur67-Shorty-LM-50_S67	ShortEMP	LMC	MiniSeq	20%	50	14, 16	17500, 30000
1	Ankur68-Shorty-LM-50_S68	ShortEMP	LMC	MiniSeq	20%	50	14, 16	17500, 30000
1	Ankur69-Shorty-LM-50_S69	ShortEMP	LMC	MiniSeq	20%	50	14, 16	17500, 30000
1	Ankur70-Shorty-LM-50_S70	ShortEMP	LMC	MiniSeq	20%	50	14, 16	17500, 30000
1	Ankur71-Shorty-LM-50_S71	ShortEMP	LMC	MiniSeq	20%	50	14, 16	17500, 30000
1	Ankur72-Shorty-LM-50_S72	ShortEMP	LMC	MiniSeq	20%	50	14, 16	17500, 30000
2	125146159_Ankur65_S11	ShortEMP	LMC	MiniSeq	1%	40	16	30000
2	125146160_Ankur66_S12	ShortEMP	LMC	MiniSeq	1%	40	16	30000
2	125146161_Ankur70_S16	ShortEMP	LMC	MiniSeq	1%	40	16	30000
2	125148154_Ankur62_S8	ShortEMP	LMC	MiniSeq	1%	40	16	30000
2	125149155_Ankur55_S1	ShortEMP	LMC	MiniSeq	1%	40	16	30000
2	125149157_Ankur63_S9	ShortEMP	LMC	MiniSeq	1%	40	16	30000
2	125150133_Ankur57_S3	ShortEMP	LMC	MiniSeq	1%	45	16	30000
2	125150134_Ankur61_S7	ShortEMP	LMC	MiniSeq	1%	45	16	30000
2	125154154_Ankur58_S4	ShortEMP	LMC	MiniSeq	1%	45	16	30000
2	125155134_Ankur68_S14	ShortEMP	LMC	MiniSeq	1%	45	16	30000
2	125156142_Ankur59_S5	ShortEMP	LMC	MiniSeq	1%	45	16	30000
2	125157143_Ankur64_S10	ShortEMP	LMC	MiniSeq	1%	45	16	30000
2	125164093_Ankur56_S2	ShortEMP	LMC	MiniSeq	1%	50	16	30000

Primer Sets

EMP	Earth Microbiome Primer Set
ShortEMP	Shortened EMP Primer Set
LongEMP	Elongated EMP Primer Set
NoLinker_ShortEMP	Shortened EMP Primer set without Fluidigm linkers

Templates

LMC	Lake Michigan Sediment
Feces	Feces Feces
Soil	Garden Soil
Zymo	ZymoBIOMICS Microbial Community Standard

Sequencing Run	Sample Name	Primer Set	Template	Sequencer	PhiX (%)	Annealing Temp. (°C)	Ref. Figure / Supp. Table	Rarefaction Depth
2	125166096_Ankur72_S18	ShortEMP	LMC	MiniSeq	1%	50	16	30000
2	125167074_Ankur69_S15	ShortEMP	LMC	MiniSeq	1%	50	16	30000
2	125168069_Ankur60_S6	ShortEMP	LMC	MiniSeq	1%	50	16	30000
2	125169062_Ankur67_S13	ShortEMP	LMC	MiniSeq	1%	50	16	30000
2	125169064_Ankur71_S17	ShortEMP	LMC	MiniSeq	1%	50	16	30000
3	149655590_515-806-LMC-5_S158	EMP	LMC	MiSeq	20%	50	15	11553
3	149656577_515-806-LMC-4_S157	EMP	LMC	MiSeq	20%	50	15	11553
3	149658573_515-806-LMC-6_S159	EMP	LMC	MiSeq	20%	50	15	11553
3	149660581_515-806-LMC-1_S154	EMP	LMC	MiSeq	20%	50	15	11553
3	149666566_515-806-LMC-3_S156	EMP	LMC	MiSeq	20%	50	15	11553
3	149667548_515-806-LMC-2_S155	EMP	LMC	MiSeq	20%	50	15	11553
3	149658589_Sort-LMC-2_S219	ShortEMP	LMC	MiSeq	20%	45	15	11553
3	149658590_Sort-LMC-4_S221	ShortEMP	LMC	MiSeq	20%	45	15	11553
3	149668569_Sort-LMC-6_S223	ShortEMP	LMC	MiSeq	20%	45	15	11553
3	149672533_Sort-LMC-1_S218	ShortEMP	LMC	MiSeq	20%	45	15	11553
3	149673528_Sort-LMC-3_S220	ShortEMP	LMC	MiSeq	20%	45	15	11553
3	149673529_Sort-LMC-5_S222	ShortEMP	LMC	MiSeq	20%	45	15	11553
3	149658575_515-806-Feces-3_S172	EMP	Feces	MiSeq	20%	50	15, S11	9201, 6429
3	149658576_515-806-Feces-4_S173	EMP	Feces	MiSeq	20%	50	15, S11	9201, 6429
3	149660583_515-806-Feces-6_S175	EMP	Feces	MiSeq	20%	50	15, S11	9201, 6429
3	149665565_515-806-Feces-2_S171	EMP	Feces	MiSeq	20%	50	S11	6429
3	149667549_515-806-Feces-1_S170	EMP	Feces	MiSeq	20%	50	15, S11	9201, 6429
3	149667551_515-806-Feces-5_S174	EMP	Feces	MiSeq	20%	50	15, S11	9201, 6429
3	149659580_Shot-Feces-1_S234	ShortEMP	Feces	MiSeq	20%	45	15, S11	9201, 6429
3	149659581_Shot-Feces-2_S235	ShortEMP	Feces	MiSeq	20%	45	S11	6429
3	149659582_Shot-Feces-6_S239	ShortEMP	Feces	MiSeq	20%	45	15, S11	9201, 6429
3	149662601_Shot-Feces-5_S238	ShortEMP	Feces	MiSeq	20%	45	15, S11	9201, 6429
3	149669561_Shot-Feces-4_S237	ShortEMP	Feces	MiSeq	20%	45	15, S11	9201, 6429
3	149673527_Shot-Feces-3_S236	ShortEMP	Feces	MiSeq	20%	45	15, S11	9201, 6429
3	149655592_515-806-Soil-4_S165	EMP	Soil	MiSeq	20%	50	15, S11	9201, 6429
3	149656578_515-806-Soil-5_S166	EMP	Soil	MiSeq	20%	50	15	9390
3	149658579_515-806-Soil-6_S167	EMP	Soil	MiSeq	20%	50	15	9390
3	149662585_515-806-Soil-3_S164	EMP	Soil	MiSeq	20%	50	15	9390
3	149665566_515-806-Soil-2_S163	EMP	Soil	MiSeq	20%	50	15	9390
3	149668558_515-806-Soil-1_S162	EMP	Soil	MiSeq	20%	50	15	9390
3	149655601_Short-soil-2_S227	ShortEMP	Soil	MiSeq	20%	45	15	9390
3	149665569_Short-soil-4_S229	ShortEMP	Soil	MiSeq	20%	45	15	9390
3	149666572_Short-soil-5_S230	ShortEMP	Soil	MiSeq	20%	45	15	9390
3	149667556_Short-soil-1_S226	ShortEMP	Soil	MiSeq	20%	45	15	9390
3	149667557_Short-soil-6_S231	ShortEMP	Soil	MiSeq	20%	45	15	9390
3	149672532_Short-soil-3_S228	ShortEMP	Soil	MiSeq	20%	45	15	9390
3	149656579_515-806-Zymo-2_S147	EMP	Zymo	MiSeq	20%	50	15, 17	10861, 4000
3	149660584_515-806-Zymo-4_S149	EMP	Zymo	MiSeq	20%	50	15, 17	10861, 4000
3	149662586_515-806-Zymo-3_S148	EMP	Zymo	MiSeq	20%	50	15, 17	10861, 4000
3	149670563_515-806-Zymo-1_S146	EMP	Zymo	MiSeq	20%	50	15, 17	10861, 4000
3	149671543_515-806-Zymo-6_S151	EMP	Zymo	MiSeq	20%	50	15, 17	10861, 4000
3	149672526_515-806-Zymo-5_S150	EMP	Zymo	MiSeq	20%	50	15, 17	10861, 4000
3	149660592_Short-Zymo-3_S212	ShortEMP	Zymo	MiSeq	20%	45	15, 17	10861, 4000
3	149662598_Short-Zymo-5_S214	ShortEMP	Zymo	MiSeq	20%	45	15, 17	10861, 4000

Sequencing Run	Sample Name	Primer Set	Template	Sequencer	PhiX (%)	Annealing Temp. (°C)	Ref. Figure / Supp. Table	Rarefaction Depth
3	149667555_Short-Zymo-1_S210	ShortEMP	Zymo	MiSeq	20%	45	15, 17	10861, 4000
3	149668567_Short-Zymo-6_S215	ShortEMP	Zymo	MiSeq	20%	45	15, 17	10861, 4000
3	149669559_Short-Zymo-4_S213	ShortEMP	Zymo	MiSeq	20%	45	15, 17	10861, 4000
3	149671545_Short-Zymo-2_S211	ShortEMP	Zymo	MiSeq	20%	45	15, 17	10861, 4000
3	149655595_Long-Zymo-2_S179	LongEMP	Zymo	MiSeq	20%	50	17	4000
3	149660588_Long-Zymo-5_S182	LongEMP	Zymo	MiSeq	20%	50	17	4000
3	149661589_Long-Zymo-4_S181	LongEMP	Zymo	MiSeq	20%	50	17	4000
3	149661590_Long-Zymo-6_S183	LongEMP	Zymo	MiSeq	20%	50	17	4000
3	149668561_Long-Zymo-3_S180	LongEMP	Zymo	MiSeq	20%	50	17	4000
3	149673524_Long-Zymo-1_S178	LongEMP	Zymo	MiSeq	20%	50	17	4000
3	149673526_No-linkers-Zymo-6_S192	NoLinker_ShortEMP	Zymo	MiSeq	20%	45	17	4000
3	149657583_No-linkers-Zymo-1_S152	NoLinker_ShortEMP	Zymo	MiSeq	20%	45	17	4000
3	149654597_No-linkers-Zymo-4_S176	NoLinker_ShortEMP	Zymo	MiSeq	20%	45	17	4000
3	149661592_No-linkers-Zymo-5_S184	NoLinker_ShortEMP	Zymo	MiSeq	20%	45	17	4000
3	149669557_No-linkers-Zymo-2_S160	NoLinker_ShortEMP	Zymo	MiSeq	20%	45	17	4000
3	149672530_No-linkers-Zymo-3_S168	NoLinker_ShortEMP	Zymo	MiSeq	20%	45	17	4000
4	147414288_515-806-Zymo-1_S146	EMP	Zymo	MiniSeq	20%	50	17	4000
4	147414289_515-806-Zymo-2_S147	EMP	Zymo	MiniSeq	20%	50	17	4000
4	147415282_515-806-Zymo-5_S150	EMP	Zymo	MiniSeq	20%	50	17	4000
4	147416283_515-806-Zymo-3_S148	EMP	Zymo	MiniSeq	20%	50	17	4000
4	147418282_515-806-Zymo-4_S149	EMP	Zymo	MiniSeq	20%	50	17	4000
4	147418283_515-806-Zymo-6_S151	EMP	Zymo	MiniSeq	20%	50	17	4000
4	147416297_Short-Zymo-1_S210	ShortEMP	Zymo	MiniSeq	20%	45	17	4000
4	147416298_Short-Zymo-2_S211	ShortEMP	Zymo	MiniSeq	20%	45	17	4000
4	147417289_Short-Zymo-4_S213	ShortEMP	Zymo	MiniSeq	20%	45	17	4000
4	147423299_Short-Zymo-3_S212	ShortEMP	Zymo	MiniSeq	20%	45	17	4000
4	147424294_Short-Zymo-6_S215	ShortEMP	Zymo	MiniSeq	20%	45	17	4000
4	147432295_Short-Zymo-5_S214	ShortEMP	Zymo	MiniSeq	20%	45	17	4000
4	147414304_Long-Zymo-3_S180	LongEMP	Zymo	MiniSeq	20%	50	Did not merge	Did not merge
4	147416295_Long-Zymo-2_S179	LongEMP	Zymo	MiniSeq	20%	50	Did not merge	Did not merge
4	147422286_Long-Zymo-4_S181	LongEMP	Zymo	MiniSeq	20%	50	Did not merge	Did not merge
4	147426297_Long-Zymo-1_S178	LongEMP	Zymo	MiniSeq	20%	50	Did not merge	Did not merge
4	147427297_Long-Zymo-6_S183	LongEMP	Zymo	MiniSeq	20%	50	Did not merge	Did not merge
4	147429292_Long-Zymo-5_S182	LongEMP	Zymo	MiniSeq	20%	50	Did not merge	Did not merge
4	147417288_No-linkers-Zymo-4_S176	NoLinker_ShortEMP	Zymo	MiniSeq	20%	45	17	4000
4	147420292_No-linkers-Zymo-5_S184	NoLinker_ShortEMP	Zymo	MiniSeq	20%	45	17	4000
4	147422288_No-linkers-Zymo-6_S192	NoLinker_ShortEMP	Zymo	MiniSeq	20%	45	17	4000
4	147423297_No-linkers-Zymo-3_S168	NoLinker_ShortEMP	Zymo	MiniSeq	20%	45	17	4000
4	147426299_No-linkers-Zymo-2_S160	NoLinker_ShortEMP	Zymo	MiniSeq	20%	45	17	4000
4	147430297_No-linkers-Zymo-1_S152	NoLinker_ShortEMP	Zymo	MiniSeq	20%	45	17	4000

Table S8: Nucleotide diversity calculations for EMP and ShortEMP 515F primers

EMP 515F	Length/Nuleotide Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
GTGCCAGCAGCCGCGGTAA	19	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G
GTGCCAGCCGCCGCGGTAA	19	G	T	G	C	C	A	G	C	C	G	C	C	G	C	G	G
GTGTCAGCAGCCGCGGTAA	19	G	T	G	T	C	A	G	C	A	G	C	C	G	C	G	G
GTGTCAGCCGCCGCGGTAA	19	G	T	G	T	C	A	G	C	C	G	C	C	G	C	G	G
	A	-	-	-	-	-	4	-	-	2	-	-	-	-	-	-	-
	C	-	-	-	2	4	-	-	4	2	-	4	4	-	4	-	-
	T	-	4	-	2	-	-	-	-	-	-	-	-	-	-	-	-
	G	4	-	4	-	-	-	4	-	-	4	-	-	4	-	4	4
	Total	1	1	1	2	1	1	1	1	2	1	1	1	1	1	1	1
	Shannon Index	0.00	0.00	0.00	0.69	0.00	0.00	0.00	0.00	0.69	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Starting Nucleotide Diversity	1.125																
Average Shannon Index	0.09																

ShortEMP 515F	Length/Nuleotide Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
GTGTCAGCAGCCGCGGTAA	19	G	T	G	T	C	A	G	C	A	G	C	C	G	C	G	G
CCAGCCGCCGCGGTAA	16	C	C	A	G	C	C	G	C	C	G	C	G	G	T	A	A
TGCCAGCAGCCGCGGTAA	18	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T
GTGTCAGCCGCCGCGGTA	18	G	T	G	T	C	A	G	C	C	G	C	C	G	C	G	G
	A	-	-	1	-	1	2	-	1	1	-	-	-	-	-	1	1
	C	1	1	1	1	3	1	1	3	2	1	4	2	1	2	-	-
	T	1	2	-	2	-	-	-	-	-	-	-	-	-	1	-	1
	G	2	1	2	1	-	1	3	-	1	3	-	2	3	1	3	2
	Total	3	3	3	3	2	3	2	2	3	2	1	2	2	3	2	3
	Shannon Index	1.04	1.04	1.04	1.04	0.56	1.04	0.56	0.56	1.04	0.56	0.00	0.69	0.56	1.04	0.56	1.04
Starting Nucleotide Diversity	2.4375																
Average Shannon Index	0.77																

Highest Possible Starting Nucleotide Diversity	**4.00**
Highest Possible Shannon Index	**1.39**

Table S9: Nucleotide diversity calculations for EMP and ShortEMP 806R primers

EMP 806R	Length/Nuleotide Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
GGACTACAAGGGTATCTAAT	20	G	G	A	C	T	A	C	A	A	G	G	G	T	A	T	C
GGACTACCAGGGTATCTAAT	20	G	G	A	C	T	A	C	C	A	G	G	G	T	A	T	C
GGACTACGAGGGTATCTAAT	20	G	G	A	C	T	A	C	G	A	G	G	G	T	A	T	C
GGACTACTAGGGTATCTAAT	20	G	G	A	C	T	A	C	T	A	G	G	G	T	A	T	C
GGACTACACGGGTATCTAAT	20	G	G	A	C	T	A	C	A	C	G	G	G	T	A	T	C
GGACTACCCGGGTATCTAAT	20	G	G	A	C	T	A	C	C	C	G	G	G	T	A	T	C
GGACTACGCGGGTATCTAAT	20	G	G	A	C	T	A	C	G	C	G	G	G	T	A	T	C
GGACTACTCGGGTATCTAAT	20	G	G	A	C	T	A	C	T	C	G	G	G	T	A	T	C
GGACTACAGGGGTATCTAAT	20	G	G	A	C	T	A	C	A	G	G	G	G	T	A	T	C
GGACTACCGGGGTATCTAAT	20	G	G	A	C	T	A	C	C	G	G	G	G	T	A	T	C
GGACTACGGGGGTATCTAAT	20	G	G	A	C	T	A	C	G	G	G	G	G	T	A	T	C
GGACTACTGGGGTATCTAAT	20	G	G	A	C	T	A	C	T	G	G	G	G	T	A	T	C
GGACTACAAGGGTTTCTAAT	20	G	G	A	C	T	A	C	A	A	G	G	G	T	T	T	C
GGACTACCAGGGTTTCTAAT	20	G	G	A	C	T	A	C	C	A	G	G	G	T	T	T	C
GGACTACGAGGGTTTCTAAT	20	G	G	A	C	T	A	C	G	A	G	G	G	T	T	T	C
GGACTACTAGGGTTTCTAAT	20	G	G	A	C	T	A	C	T	A	G	G	G	T	T	T	C
GGACTACACGGGTTTCTAAT	20	G	G	A	C	T	A	C	A	C	G	G	G	T	T	T	C
GGACTACCCGGGTTTCTAAT	20	G	G	A	C	T	A	C	C	C	G	G	G	T	T	T	C
GGACTACGCGGGTTTCTAAT	20	G	G	A	C	T	A	C	G	C	G	G	G	T	T	T	C
GGACTACTCGGGTTTCTAAT	20	G	G	A	C	T	A	C	T	C	G	G	G	T	T	T	C
GGACTACAGGGGTTTCTAAT	20	G	G	A	C	T	A	C	A	G	G	G	G	T	T	T	C
GGACTACCGGGGTTTCTAAT	20	G	G	A	C	T	A	C	C	G	G	G	G	T	T	T	C
GGACTACGGGGGTTTCTAAT	20	G	G	A	C	T	A	C	G	G	G	G	G	T	T	T	C
GGACTACTGGGGTTTCTAAT	20	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C
A		-	-	24	-	-	24	-	6	8	-	-	-	-	12	-	-
C		-	-	-	24	-	-	24	6	8	-	-	-	-	-	-	24
T		-	-	-	-	24	-	-	6	-	-	-	-	24	12	24	-
G		24	24	-	-	-	-	-	6	8	24	24	24	-	-	-	-
Total		1	1	1	1	1	1	1	4	3	1	1	1	1	2	1	1
Shannon Index		0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.39	1.10	0.00	0.00	0.00	0.00	0.69	0.00	0.00

Starting Nucleotide Diversity	**1.375**
Average Shannon Index	**0.20**

Highest Possible Starting Nucleotide Diversity	**4.00**
Highest Possible Shannon Index	**1.39**

ShortEMP 806R	Length/Nuleotide Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
GGACTACAAGGGTATCTAA	19	G	G	A	C	T	A	C	A	A	G	G	G	T	A	T	C
GACTACCAGGGTATCTAAT	19	G	A	C	T	A	C	C	A	G	G	G	T	A	T	C	T
GGACTACGAGGGTATCTA	18	G	G	A	C	T	A	C	G	A	G	G	G	T	A	T	C
GGACTACTAGGGTATCTAAT	20	G	G	A	C	T	A	C	T	A	G	G	G	T	A	T	C
GACTACACGGGTATCTAAT	19	G	A	C	T	A	C	A	C	G	G	G	T	A	T	C	T
CTACCCGGGTATCTAAT	17	C	T	A	C	C	C	G	G	G	T	A	T	C	T	A	A
CTACGCGGGTATCTAAT	17	C	T	A	C	G	C	G	G	G	T	A	T	C	T	A	A
GGACTACTCGGGTATCTA	18	G	G	A	C	T	A	C	T	C	G	G	G	T	A	T	C
GACTACAGGGGTATCTAAT	19	G	A	C	T	A	C	A	G	G	G	G	T	A	T	C	T
GGACTACCGGGGTATC	16	G	G	A	C	T	A	C	C	G	G	G	G	T	A	T	C
CTACGGGGGTATCTAAT	17	C	T	A	C	G	G	G	G	G	T	A	T	C	T	A	A
GACTACTGGGGTATCTAAT	19	G	A	C	T	A	C	T	G	G	G	G	T	A	T	C	T
GGACTACAAGGGTTTCTA	18	G	G	A	C	T	A	C	A	A	G	G	G	T	T	T	C
ACTACCAGGGTTTCTAAT	18	A	C	T	A	C	C	A	G	G	G	T	T	T	C	T	A
ACTACGAGGGTTTCTAAT	18	A	C	T	A	C	G	A	G	G	G	T	T	T	C	T	A
GGACTACTAGGGTTTCTAA	19	G	G	A		T	A	C	T	A	G	G	G	T	T	T	C
GACTACACGGGTTTCTAA	18	G	A	C	T	A	C	A	C	G	G	G	T	T	T	C	T
ACTACCCGGGTTTCTAA	17	A	C	T	A	C	C	C	G	G	G	T	T	T	C	T	A
ACTACGCGGGTTTCTA	16	A	C	T	A	C	G	C	G	G	G	T	T	T	C	T	A
ACTACTCGGGTTTCTAAT	18	A	C	T	A	C	T	C	G	G	G	T	T	T	C	T	A
ACTACAGGGGTTTCTAAT	18	A	C	T	A	C	A	G	G	G	G	T	T	T	C	T	A
CTACCGGGGTTTCTAAT	17	C	T	A	C	C	G	G	G	G	T	T	T	C	T	A	A
CTACGGGGGTTTCTAAT	17	C	T	A	C	G	G	G	G	G	T	T	T	C	T	A	A
ACTACTGGGGTTTCTAAT	18	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A
A		7	5	12	7	5	8	5	3	5	-	3	-	4	5	5	12
C		5	7	5	12	9	9	11	3	1	-	-	-	5	7	5	7
T		-	5	7	5	7	2	1	3	-	5	9	17	15	12	14	5
G		12	7	-	-	3	5	7	15	18	19	12	7	-	-	-	-
Total		3	4	3	3	4	4	4	4	3	2	3	2	3	3	3	3
Shannon Index		1.03	1.37	1.03	1.03	1.31	1.27	1.18	1.07	0.67	0.51	0.97	0.60	0.92	1.03	0.97	1.03

Starting Nucleotide Diversity	**3.1875**
Average Shannon Index	**1.00**

Chapter V: Quantitating primer-template interactions using a deconstructed PCR methodology

Abstract

When the polymerase chain reaction (PCR) is used to amplify simultaneously multiple templates of unknown and unequal abundance, preferential amplification of certain templates (PCR bias) leads to a distorted representation of the templates in final amplicon pool. PCR selection, a type of PCR bias, is influenced by mismatches between primers and templates, the locations of mismatches, and the nucleotide pairing of mismatches. Direct measurement of primer-template interactions has not been possible, leading to substantial uncertainty when attempting to optimize PCR reactions and primer pools. In this study, we developed an experimental system to systematically study primer-template interactions. We synthesized 10 double-stranded DNA templates with unique priming sites, as well as 64 primers with 0, 1, 2 or 3 mismatches with each of the 10 templates. By using a deconstructed PCR (DePCR) methodology [18, 96], we generated empirical data showing individual primer interactions with templates in complex template-primer amplification reactions. Both standard PCR and DePCR amplification protocols were used to amplify templates in a series of 16 experiments in which templates, primers, and annealing temperature were varied. We observed that although perfect match primer-template interactions are important, the dominant type of interactions are mismatch amplifications, and mismatch annealing and polymerase copying starts immediately during the first two cycle of PCR. In reactions with degenerate primer pools, multiple mismatches are tolerated, and these do not have a strong effect on observed template ratios after amplification when employing the DePCR methodology. We establish here a quantitative experimental system for interrogating primer-template interactions.

Introduction

The polymerase chain reaction (PCR) is a well-established tool for amplification of regions of DNA [8, 28] and is now routinely used in a broad range of biological studies. When PCRs are performed to amplify multiple different templates of unknown and generally unequal abundance, the final pool of PCR amplicons may have an altered ratio of templates relative to the original sample. Such a result is labeled 'PCR bias' and is a well-studied phenomenon, particularly in the context of microbial ecology [14-16]. Wagner et al. [14] defined two broad classes of distortion of underlying template ratios – including PCR selection and PCR drift. In the first category – PCR selection, PCR conditions favor certain templates, and bias generated from selection has been attributed to a broad number of factors, including (but not limited to): annealing temperature [17, 18], mismatches between template and primer [19, 20], location of mismatches between template and primer [21], interference from flanking regions during initial stages of PCR [22], too many PCR cycles [23], input DNA concentration [24-26], preferential amplification of low GC templates in a mixture [13], higher GC content in primer region/differences in primer binding energy [15, 27], template saturation at plateau phase of PCR [28], preferential formation of primer dimers from some primer variants when working with degenerate pools of primers [15], preferential amplification of unmethylated DNA [29], re-annealing of copies to templates leading to reduced amplification efficiency [30, 31], temperature ramp during thermocycling allowing for formation of homoduplexes [32], and combinatorial effects of linear copying of gDNA and exponential amplification of PCR products occurring simultaneously and at different efficiencies [18].

The second category – PCR drift – is caused by stochastic effects during the early stages of PCR when primer-genomic DNA template interactions dominate (as opposed to primer-amplicon

interactions) [14, 15]. To reduce PCR drift, multiple reactions are typically combined. However, Suzuki and Giovannoni [28] suggested that PCR selection was the primary driver of PCR bias, though low input gDNA could lead to higher stochastic effects [15]. A third category of bias should also be considered – the generation of PCR artifacts, such as chimeras [32]. The creation of chimeras – hybrid artifact products of PCR – can be enhanced by using polymerases with low processivity, with short elongation times, and with high cycle number [37]. Reducing cycle number is always recommended with regards to decreasing chimera formation [21, 39, 51, 97].

Thus, many possible sources of PCR bias exist, and many solutions to PCR bias have been attempted. These include: addition of various additives to PCR master mixes, including acetamide [13], DMSO and glycerol [98], running fewer cycles of PCR [28, 34, 88], reducing degeneracies in primers whenever possible [15], increasing ramp rates for transitions between temperatures [32], and use of long elongation times and/or use of highly processive polymerases to ensure complete copying during each cycle [51]. In some systems, higher annealing temperatures are recommended to reduce effects of secondary structure [41], while in complex template systems such as microbial DNA, lower annealing temperatures are recommended to improve tolerance for mismatch annealing [40]. We have also introduced the "deconstructed PCR" (DePCR) method [18, 96] to reduce PCR bias by addressing several issues simultaneously. Briefly, DePCR has two related ways of reducing PCR bias. First, locus-specific primers are only employed for two cycles of PCR, and low efficiency interactions between primers and gDNA templates are minimized. Secondly, exponential amplification of amplicons is performed using non-degenerate primers without mismatches with templates. Locus-specific primer-amplicon interactions are eliminated from the reaction entirely.

Despite the substantial amount of effort that has been invested into identifying and correcting PCR bias, PCR-based studies continue to generate data that distort underlying template ratios. Furthermore, fundamental questions relating to primer-template interactions have not been thoroughly investigated, and these interactions are at the heart of PCR bias. Improvements in fundamental understanding of primer-template interactions can be of benefit by providing guidance for design of primer sets and for selection of optimal PCR conditions. Several recent advances offer a new opportunity to examine fundamental primer-template interactions. First, low cost next-generation sequencing allows for direct interrogation of complex templates without using data reduction strategies such as terminal restriction fragment length polymorphism (TRFLP; [101])or denaturing gradient gel electrophoresis (DGGE; [10]). The second development is the ability to easily and inexpensively synthesize double-stranded DNA templates. The third is the DePCR method, which, in addition to reducing PCR bias by limiting primer-gDNA template interactions to the first two cycles of linear amplification, also provides a mechanism, described below, to identify which primers in a degenerate primer pool interact with each template.

Thus, our studied employed high-throughput amplicon sequencing on an Illumina MiniSeq sequencer, enabling us to generate thousands of sequences per sample for rigorous quantitation of amplicons. Secondly, we synthesized 10 double-stranded DNA templates with unique priming sites, as well as 64 primers, 20 bases in length, with 0, 1, 2 or 3 mismatches with each of the 10 templates. For primers and templates with mismatches, mismatches were located close to the 3' end of the primer (-2 position, counting from the 3' end), the middle of the primer (-8), or closer to the 5' end of the primer (-14). Finally, both standard PCR amplification protocols and DePCR amplification protocols were used to amplify templates in a series of experiments in which

templates, primers, and annealing temperature were varied. Our study also avoided other potential sources of bias by: (1) interrogating only one primer site; (2) using identical DNA concentrations in all experiments; (3) employing synthetic DNA, not genomic DNA; (4) generating short amplicons only; and (5) locus-specific primers used only for 2 cycles – therefore no primer limitation or strong inhibition due to high copy number.

Materials and Methods

Nucleic acids

Artificial double-stranded DNAs (gBlocks Gene Fragments, here called "synthetic templates" or ST) were synthesized by Integrated DNA Technologies, Inc. (IDT; Coralville, Iowa). Prior to pooling, each ST was quantitated using fluorimetry with a Qubit 4.0 fluorometer with the dsDNA BR Assay (Thermo Fisher Scientific, San Jose, CA). DNA concentrations were equalized among all STs prior to pooling. A series of template mixtures were created (see **Table 10, Table S11** for full description), including "A" (single template, ST1), "B" (equimolar pooling of all 10 templates), "C" (equimolar pooling of all templates except ST1, and inclusion of template ST1 at $1/10^{th}$ concentration), "D" (graduated pooling of template ST1, and ST6, ST7, and ST8 templates with differences at the 3' variable position), and "E" (graduated pooling of template ST1, and ST4, ST11 and ST15 templates with differences at the middle variable position). A total of 64 different oligonucleotide primers were synthesized as LabReady primers, normalized to 100 μM concentration (IDT) (**Table S10**). The 64 primers ("806F" primers) were grouped into four categories: (i) primer with no mismatches with template ST1 (1 primer), (ii) primers with one mismatch with template ST1 (9 primers), (iii) primers with two mismatches

with template ST1 (27 primers), and (iv) primers with three mismatches with template ST2 (27 primers) (**Table S11**). When used with other templates than the ST1 template (*i.e.*, ST4, ST6, ST7, ST8, ST11, ST15, ST23, ST39, and ST55), each template had 1 perfect match primer, 9 single mismatch primers, 27 double mismatch primers and 27 triple mismatch primers. For each template mixture, a separate experiment was conducted using one of five primer pools (**Table 10**). Primer pool 1 contained only a single primer, perfectly matching the ST1 template. Primer pool 2 contained ten primers, each perfectly matching one of the ten templates. Primer pool 3 contained nine primers, each perfectly matching one template except for the ST1 template. Primer pool 4 contained 27 primers, each with two mismatches relative to template ST1 and 1-3 mismatches relative to all other templates. Primer pool 5 contained all 64 primers. In total, 640 possible primer-template interactions were considered (10 templates x 64 primers), with a maximum of 3 mismatches between any template and primer. Primer theoretical melting temperatures were calculated using the OligoAnalyzer3.1 calculator [74], assuming 250 nM primer concentration, 2 mM Mg^{2+}, and 0.2 mM dNTPs. All primers contained 5' linker sequences known as common sequence 1 and 2 (CS1: ACACTGACGACATGGTTCTACA and CS2: TACGGTAGCAGAGACTTGGTCT) as described previously [61]. Illumina P5 (AATGATACGGCGACCACCGA) and P7 (CAAGCAGAAGACGGCATACGA) primers, for use in the DePCR protocol, were also synthesized as LabReady primers and normalized to 100 μM concentration (IDT).

Targeted-amplicon sequencing (TAS) Protocol

A standard two-stage PCR amplification method was used to generate amplicons for next-generation sequencing [38]. First stage PCR amplifications were performed in 10 μL reactions in 96-well plates, using MyTaq HS 2X master mix (Bioline, Taunton, MA). 2.5 ng of synthetic ST

template mixtures (A-E, described above) was used for each 10 μL reaction. Primer pools were added at a final concentration of 200 nM. All reactions were performed with eight technical replicates. Thermocycling conditions were 95^0C for 5 minutes, 28 cycles of 95°C for 30 seconds, annealing temperatures of 45°C or 55°C for 45 seconds, and 72°C for 30 seconds, and a final elongation at 72°C for 7 minutes. Subsequently, a second PCR amplification was performed in 10 microliter reactions in 96-well plates. A master mix for the entire plate was made using the MyTaq HS 2X master mix, and each well received a separate primer pair with a unique 10-base barcode, obtained from the Access Array Barcode Library for Illumina (Fluidigm, South San Francisco, CA). These Access Array primers contained the CS1 and CS2 linkers at the 3' ends of the oligonucleotides, and the final concentration was 400 nM. 1 μL of the first stage PCR reaction, without purification, was added to the second stage reaction. Cycling conditions were as follows: 95°C for 5 minutes, followed by 8 cycles of 95°C for 30", 60°C for 30" and 72°C for 30". A final, 7-minute elongation step was performed at 72°C. Second stage PCR amplicons were pooled together in equal concentrations, and the pooled library was purified using an AMPure XP cleanup protocol (0.7X, vol/vol; Agencourt, Beckmann-Coulter) to remove short fragments. Pool, cleaned amplicons were sequenced on an Illumina MiniSeq mid-output flow cell with 2x153 base reads, and with an approximate 30% phiX spike-in.

Deconstructed PCR (DePCR) Protocol

A two-stage deconstructed PCR (DePCR) method [18, 96] was also used to generate amplicons for next-generation sequencing (**Figure 18)**. In this protocol, four primers are added to the first stage reaction, including locus-specific primer pools containing 5' CS1 and CS2 linkers (pools i, ii, iii and iv as described; each pool was added at 200 nM concentration), as well as Fluidigm Access Array Barcode Library primers, containing Illumina sequencing adapters, a sample-

specific 10 nucleotide barcode, and CS1 and CS2 linkers at the 3' ends (added at 400 nM concentration). 2.5 ng of synthetic ST mixtures (A-E, described above) was used for each 10 µL reaction. All reactions were performed using 2× MyTaq HS Mix and reactions were conducted in 96-well plates. First stage thermocycling conditions were: initial denaturation at 95°C for 5 minutes, followed by two cycles of 95°C for 30 seconds and either 45°C and 55°C for 20 minutes, followed by two cycles of 95°C for 30 seconds and 60°C for 2 minutes. Subsequently, technical replicates from each experiment (*e.g.*, A1, A2, A3) were pooled together from both annealing temperatures (16 reactions per pool).

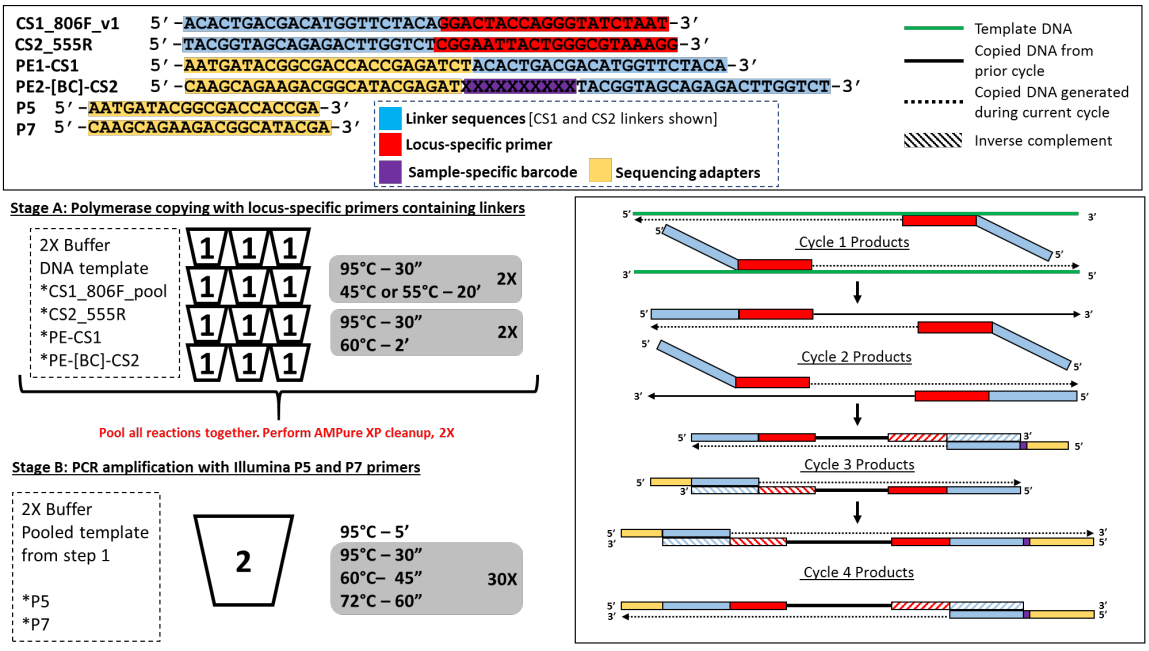


Figure 18: Schematic of Deconstructed PCR (DePCR) workflow

CS1 = common sequence 1 linker sequence. CS2 = common sequence 2 linker sequence. BC = barcode. F = Forward primer. R = Reverse primer, P5/P7 = Illumina primers, PE1/PE2 = Fluidigm Access Array Barcode Library Illumina adapters. In stage A, individual samples are

copied for 4 cycles with locus-specific primers and Fluidigm barcoded primers. Subsequently, all reaction volumes are pooled and purified together, and then amplified with Illumina P5 and P7 primers in stage B. During stage A, linear copying of templates leads to cycle 4 products which contain Illumina sequencing adapters, sample-specific barcodes, and locus-specific region of interest. Only fragments with Illumina adapters and barcodes are amplified in stage B.

Pooled replicates were purified twice purified using an AMPure XP cleanup protocol (0.7X, vol/vol) and eluted in 50 μL. Of this eluate, 20 μL were used as template for amplification in the second stage reaction with P5 and P7 primers. Final volume for each amplification reaction was 50 μL. Thermocycling conditions were 95°C for 5 minutes and 30 cycles of 95°C for 30 seconds, 60°C for 45 seconds, and 72°C for 90 seconds. Amplicons generated from second stage reactions were again purified using an AMPure XP cleanup protocol (0.7X, vol/vol). Pooled and purified amplicons from each experiment were quantified using Qubit fluorimetry (Qubit 4.0, Thermo Fisher Scientific), and further pooled together to generate a final library. Pooled, cleaned amplicons were sequenced on an Illumina MiniSeq mid-output flow cell with 2x153 base reads, and with an approximate 30% phiX spike-in. Library preparation and sequencing were performed at the UIC Sequencing Core (UICSQC).

Sequence Data Analysis

Raw FASTQ files were merged using the software package PEAR [75] using default parameters. Merged reads were then converted from FASTQ to FASTA format using the function convert_fastaqual_fastq.py within the software package QIIME [1]. Sequence data were

analyzed to identify recognition sequences (*i.e.*, identifying which of 10 templates was amplified), and to identify the sequence of the primer used to amplify the template (*i.e.*, identifying which of 64 possible 'forward' primers was used for amplification). In total, 640 possible primer-template pairs were considered, though each experiment individually had fewer possible combinations. A list of template sequences is provided in **Supplemental Materials 1**, and a list of all primer sequences is shown in **Table S10**. All possible primer-template interactions are shown in **Tables S11 and S12**. To calculate utilization profiles for all the samples, a mapping file, containing all possible unique combinations of 806F primers and recognition sequences were generated. To identify the 640 unique primer-recognition sequence combinations that could occur, a custom bash UNIX shell script (**Supplemental Material 2**) was written to search for each combination. Only sequences that matched perfectly with a primer variant sequence and a recognition sequence were counted. In the end, all counts were collated to generate a biological observation matrix (BIOM) [77]. The BIOM was rarefied to a depth of 7,000 counts per sample in the R programming environment [79] for all downstream analyses. The BIOMs were further split into template BIOMs (10 features) and primer BIOMs (64 features). Heatmaps for both template and primer BIOMs were generated using the package pheatmap in R. The *vegan* package [80] was used to generate alpha diversity indices and to calculate pairwise Bray-Curtis dissimilarity scores. Metric multi-dimensional scaling (mMDS) plots were created using the cmdscale and ggplot2 [81] functions within R. Ellipses, representing 95% confidence intervals around group centroids, were created assuming a multivariate t-distribution. Analysis of similarity (ANOSIM) calculations were performed in the software package Primer7 [78]
(Primer-E, Plymouth, UK). Ideal score (IS) analysis was performed using the *vegan* R package slightly modified from the formula described previously [18] to account for uneven distribution

of templates. The IS is a summation of the absolute difference between the expected relative abundance and the observed relative abundance for each feature in a multi-feature dataset [99]. The general IS has a range from 0 (perfect representation of the input template distribution) to 200.

Data Archive

Raw sequence data files were submitted in the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI). The BioProject identifier of the samples is PRJNA513137. Full metadata for each sample are provided in **Table S12**.

Results

Experimental design

As part of this study, 16 different experiments were conducted comparing the effects of PCR amplification method (TAS or DePCR) and annealing temperature (45°C or 55°C). Each experiment was a PCR amplification of synthetic DNA templates, ranging from a single template to a combination of up to 10 different templates. In some experiments, synthetic DNA templates were added to the PCR reaction mixture at equimolar concentration, while in others, each template was added at a different concentration. In addition to varying input templates, 64 primers were used in different combinations to amplify the synthetic templates (STs). In some reactions, only a single primer was used, while in others, various combinations of primers were

used. A full list of experimental conditions is shown in **Table 10 and Figure 18**. 7-8 technical replicates were performed for each experimental condition.

The primary template was designed in a similar manner to synthetic templates described previously [18]. Briefly, the synthetic DNA sequences were based on the 16S rRNA gene sequence from a Gammaproteobacterium, *Rhodanobacter denitrificans* [57]. The prior design was modified by reducing the amplicon size so that the amplification product could be sequenced on an Illumina MiniSeq sequencer that generates paired-end 2x153 nucleotide reads. Furthermore, to reduce complexity of the overall study, primer manipulation was examined only for a single primer site. Synthetic template sequences at the second primer site were identical for all reference templates and targeted by the 555R primer (**Table S10**). The ten synthetic templates were 451 bp in length, and identical except for the forward ('806F') primer region and a so-called "recognition" sequence in the middle of the amplicon (**Supplemental Materials 1; Figure 19**). Each template, when compared to other templates, has variants in up to 3 positions, located at -2, -8, and -14 from the 3' end of the 806F primer annealing site. The -2, -8, and -14 positions represent 3', middle, and 5' mismatches, respectively (**Figure 19**, **Table S11**). In each synthetic template, the recognition sequences are linked to a specific primer site variant, thus allowing identification of the source template primer site, regardless of which primer anneals to the template and initiates template copying (**Figure 19**). Using DePCR, the sequence of the primer annealing to templates is retained during exponential amplification [18, 96], and in this experimental system is linked to a recognition sequence.

Table 10: Description of templates and primers used in experiments conducted as part of this study

Exp Name	Number of Templates Used	Templates Used	Pooling	Ratio for unequal pooling	Number of Primers used	Primer Name (806F_v1 to 806F_v64)	Experimental Aim
A1	1	ST1	Equimolar		1	1	Evaluate the amplification viability of the primer-template system.
A2	1	ST1	Equimolar		10	1, 4, 6, 7, 8, 11, 15, 23, 39, 55	Assess competition between perfect matching and 1 mismatch primers with single template. Assess effect of mismatch position on priming efficiency.
A3	1	ST1	Equimolar		9	4, 6, 7, 8, 11, 15, 23, 39, 55	Assess competition between 1 mismatch primers with single template
A4	1	ST1	Equimolar		27	2, 3, 5, 9, 10, 12-16, 19, 21, 22, 24, 27, 31, 35, 37, 38, 40, 43, 47, 51, 53, 54, 56, 59, 62	Assess competition between 2 mismatch primers with single template when no perfect or 1 mm match primers are available.
A6	1	ST1	Equimolar		64	1-64	Assess competition between 0, 1, 2 and 3 mismatch primers with single template. Assess effect of mismatch position on priming efficiency.
B1	10	ST1, ST4, ST6, ST7, ST8, ST11, ST15, ST23, ST39, ST55	Equimolar		1	1	Assess ability of single primer to amplify 10 templates, including a template perfectly matching, as well as 9 templates with 3', middle, or 5' mismatches. Assess effect of mismatch position on priming efficiency.
B2	10	ST1, ST4, ST6, ST7, ST8, ST11, ST15, ST23, ST39, ST55	Equimolar		10	1, 4, 6, 7, 8, 11, 15, 23, 39, 55	Assess ability of 10 primers to amplify 10 templates where each primer perfectly matches one of the templates. Determine whether perfect match amplification dominates, and whether annealing temperature plays a role. Assess effect of mismatch position on priming efficiency.
B3	10	ST1, ST4, ST6, ST7, ST8, ST11, ST15, ST23, ST39, ST55	Equimolar		9	4, 6, 7, 8, 11, 15, 23, 39, 55	Assess effect of removing one primer from amplification of a pool of 10 templates. 9 templates have perfectly matching primers, 1 template has no perfectly matching primers. Assess effect of mismatch position on priming efficiency.
B4	10	ST1, ST4, ST6, ST7, ST8, ST11, ST15, ST23, ST39, ST55	Equimolar		27	2, 3, 5, 9, 10, 12-16, 19, 21, 22, 24, 27, 31, 35, 37, 38, 40, 43, 47, 51, 53, 54, 56, 59, 62	Assess effect of removing perfect matching primers on amplification of 10 templates. Assess effect of mismatch position on priming efficiency.

Exp Name	Number of Templates Used	Templates Used	Pooling	Ratio for unequal pooling	Number of Primers used	Primer Name (806F_v1 to 806F_v64)	Experimental Aim
C1	10	ST1, ST4, ST6, ST7, ST8, ST11, ST15, ST23, ST39, ST55	Unequal	**0.1**/1/1/1/1/ 1/1/1/1/1	1	1	Assess effect of template concentration on ability of single primer to amplify 10 templates. The primer perfectly matches the low abundant template. Assess effect of mismatch position on priming efficiency.
C2	10	ST1, ST4, ST6, ST7, ST8, ST11, ST15, ST23, ST39, ST55	Unequal	**0.1**/1/1/1/1/ 1/1/1/1/1	10	1, 4, 6, 7, 8, 11, 15, 23, 39, 55	Assess effect of template concentration on ability of 10 primer pool to amplify 10 templates. Assess effect of mismatch position on priming efficiency.
C3	10	ST1, ST4, ST6, ST7, ST8, ST11, ST15, ST23, ST39, ST55	Unequal	**0.1**/1/1/1/1/ 1/1/1/1/1	9	4, 6, 7, 8, 11, 15, 23, 39, 55	Assess effect of template concentration on ability of 9 primer pool to amplify 10 templates. Missing perfect match primer targets the low abundance template. Assess effect of mismatch position on priming efficiency.
D1	4	ST1, ST6, ST7, ST8	Unequal	1/2/4/8	1	1	Assess effect of more dynamic distribution of template abundance, and amplification with single primer. Single primer perfectly matches lowest abundance template and has 3' mismatches with the other three templates; Assess effect of mismatch sequence on priming efficiency.
D2	4	ST1, ST6, ST7, ST8	Unequal	1/2/4/8	10	1, 4, 6, 7, 8, 11, 15, 23, 39, 55	Assess effect of more dynamic distribution of template abundance, and amplification with 10 primer pool. Three templates chosen have a single mismatch at the 3' location relative to the lowest abundance template. Assess effect of mismatch sequence on priming efficiency.
E1	4	ST1, ST4, ST11, ST15	Unequal	1/2/4/8	1	1	Assess effect of more dynamic distribution of template abundance, and amplification with single primer. Single primer perfectly matches lowest abundance template and has middle mismatches with the other three templates; Assess effect of mismatch sequence on priming efficiency.
E2	4	ST1, ST4, ST11, ST15	Unequal	1/2/4/8	10	1, 4, 6, 7, 8, 11, 15, 23, 39, 55	Assess effect of more dynamic distribution of template abundance, and amplification with 10 primer pool. Three templates chosen have a single mismatch at the middle mismatch location relative to the lowest abundance template. Assess effect of mismatch sequence on priming efficiency.

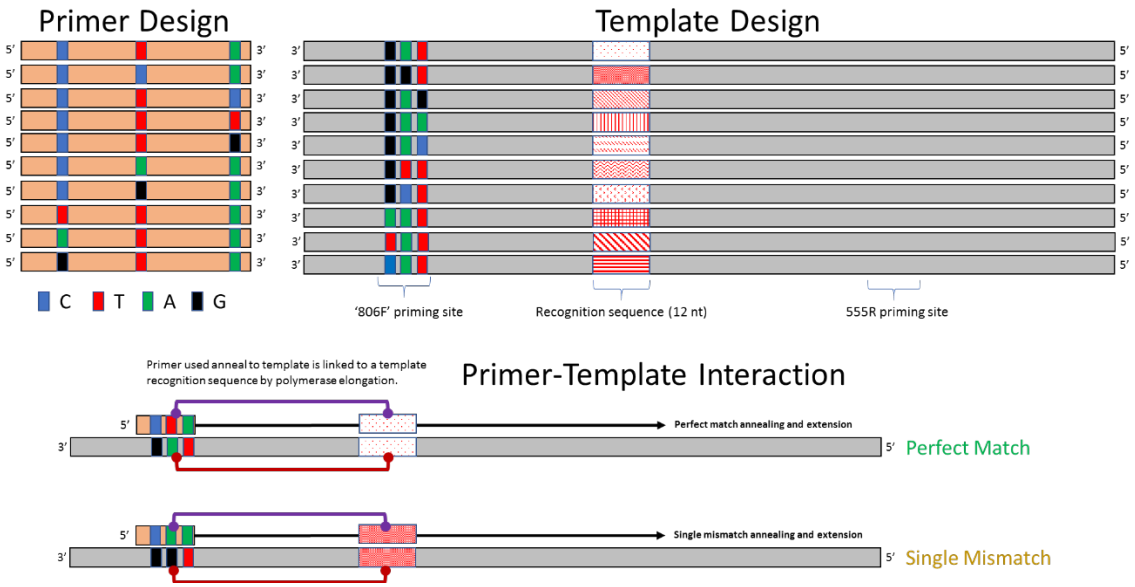


Figure 19: Primer, template and experimental design

(**A**) 64 unique oligonucleotide primers were synthesized of which 10 are shown here. Primers were identical except for 3 positions at -2, -8 and -14 positions relative to the 3' ends. Variant bases have been indicated by color ("C" = Blue, "T" = Red, "A" = Green, and "G" = Black). (**B**) Schematic of 10 synthetic DNA templates used in this study. Each template was identical except for the 806F priming site and the 12-base recognition sequence. Each unique priming site sequence is linked with a unique recognition sequence. (**C**) 640 potential primer-template interactions can occur in this system, of which two are shown here. Shown are primer-template interactions indicating the annealing of a perfectly matched primer and a primer with a single mismatch. Perfect match and mismatch annealing are determined by comparing the recognition sequence to the observed primer sequence in each sequencing reaction. Only reactions conducted using the DePCR methodology retain the sequence of the primer annealing to the source DNA templates. Although not shown, all primers contain common sequence linkers at the 5' ends (**Figure 18**).

In this manner, NGS amplicon sequence data were used to identify which templates were amplified and which primer annealed to each template. These data were used to measure the percentage of sequence reads derived from perfect match and mismatch interactions between primers and templates. Results from each experiment are presented together on single figures (**Figures S11-S26**). Each figure contains results from primer BIOM analysis, including a clustered heatmap, showing the relative abundance of 64 primer variants in the sequence data for that experiment, along with a metric multidimensional scaling plot for the primer variant utilization. In addition, analysis of similarity (ANOSIM; 9999 permutations) calculations were performed to determine if primer profiles were significantly different between TAS and DePCR amplification regardless of annealing temperature, and between annealing temperatures within each amplification method (TAS or DePCR). Based on the known primer site sequence of the template (derived from the recognition sequence), we identified whether the primers annealing to templates represented perfect match, single mismatch, double mismatch or triple mismatch interactions, when templates were amplified using the DePCR method. In addition, location of mismatches and mismatch type (*e.g.*, A-G, G-G, etc.) were identified and quantified. For each experiment, the percentage of reads derived from 0, 1, 2 or 3 mismatch primer-template interactions were counted and differences between experiments conducted at 45°C and 55°C annealing temperatures were examined. For templates amplified with primers containing only single mismatches, the percentage of reads derived from 5' (-14), middle (-8) and 3' (-2) mismatches were measured. The average theoretical melting temperature of primers used in amplifying the templates in each experiment was calculated, in addition to a Shannon Index (loge) based on the relative abundance of primer utilization for each sample. Here, the Shannon index represents evenness, as a fixed number of features are present in each experiment. One-

way analysis of variance [58] was used to determine if values were significantly different by annealing temperature (7-8 replicates per group).

In addition to primer utilization, relative template distribution was also analyzed. Metric MDS (mMDS) plots were generated based on BIOM files with 10 features (*i.e.*, 10 unique templates). In addition, the expected distribution (*i.e.*, input distribution) for each experiment was added to the MDS plots. ANOSIM was performed (9999 permutations) to determine if template distributions differed between amplification method (DePCR or TAS) or by temperature (45°C or 55°C) within each amplification method. A clustered heatmap was generated for the average template profiles for each experimental condition, along with the distribution of the input templates. An Ideal Score (IS) was calculated for each replicate, and ANOVA was performed to determine which method (DePCR or TAS) was able to generate a template distribution profile most similar to that of the input template, as well as which annealing temperature within each method was able to generate a template distribution profile most similar to that of the input template distribution.

Interrogation of single templates with primer pools of varying degeneracy

In the 'A' series of experiments (A1, A2, A3, A4, and A6; **Table 10; Figures 20, S11-S15**), amplification reactions were performed using a single synthetic DNA template (ST1), and from 1 to 64 primers, using both standard (TAS) and DePCR methodologies. In each experiment, template profiling was performed through counting of recognition sequences in datasets, followed by rarefaction (7,000 sequences/sample, 7-8 replicates per condition). All recognition sequences had a minimum Hamming distance of 4 (ranging from 4 to 11 in a recognition

sequence of 12 nucleotides), enabling robust detection of the relative abundance of each template in the dataset. For all studies, we performed analysis of similarity (ANOSIM) tests to determine if the template composition differed between TAS and DePCR methods, and between 45 and 55°C annealing temperatures within TAS and within DePCR. 'Ideal' score analyses were performed to assess how similar observed profiles were to the expected profiles (*i.e.*, input DNA distribution) for each condition. For all "A" experiments, Ideal scores were extremely low (<0.5 on a scale of 0 to 200), regardless of amplification method; this was expected, as only one template was added to each experiment (**Figures S11-S15**).

Primer sequences (variants 1-64) were identified in each sequence, and data were rarefied to 7,000 sequences per sample. The relative abundance profiles of each primer variant in a primer pool is called a 'primer utilization profile' or PUP, and these data can be analyzed in the same manner as any other biological feature. In standard TAS, the PUPs tend to have high diversity and broadly even utilization, leading to a high Shannon index. In systems such as this, with a fixed number of features, the Shannon Index represents feature evenness. The reason for the high diversity is that in standard TAS amplification, primers anneal to both genomic DNA templates, and then later to DNA copies [18]. Due to tolerance to mismatches and possible depletion of specific primer variants during exponential amplification over 25-35 cycles of standard PCR, the signal of specific primers annealing to the source templates is lost. This is observed in all experiments with greater than one forward primer variant (**Figures 22, 23, and S11-S26**). Conversely, the DePCR method allows only two linear cycles of DNA copying with locus-specific primers. Subsequently, exponential amplification is performed using primers targeting linker sequences that are common to all templates; thus, the signal of primers annealing to the source DNA template is preserved (**Figure 18**).

Several patterns were observed when amplifying the single ST1 template with various primer pools (**Figure 20**). First, Shannon indices (*i.e.*, evenness) of primer utilization were generally higher with TAS amplification relative to DePCR amplification for "A" experiments, due to signal scrambling in the TAS method. However, in experiment A1 with only a single primer, the Shannon index was higher in DePCR reactions due to PCR errors derived from polymerase copying through the primer region. In the A1 experiment, 95.9% of reads were annotated as containing the ST1 primer (the only primer added to the reaction), while 98.8% of reads were annotated as containing the ST1 primer in the TAS samples; ANOVA P<0.0001). In experiment A3 with 9 primers, the Shannon index of DePCR at 45°C was lower than for the TAS samples, regardless of annealing temperature, indicating a very even utilization of primers under this condition.

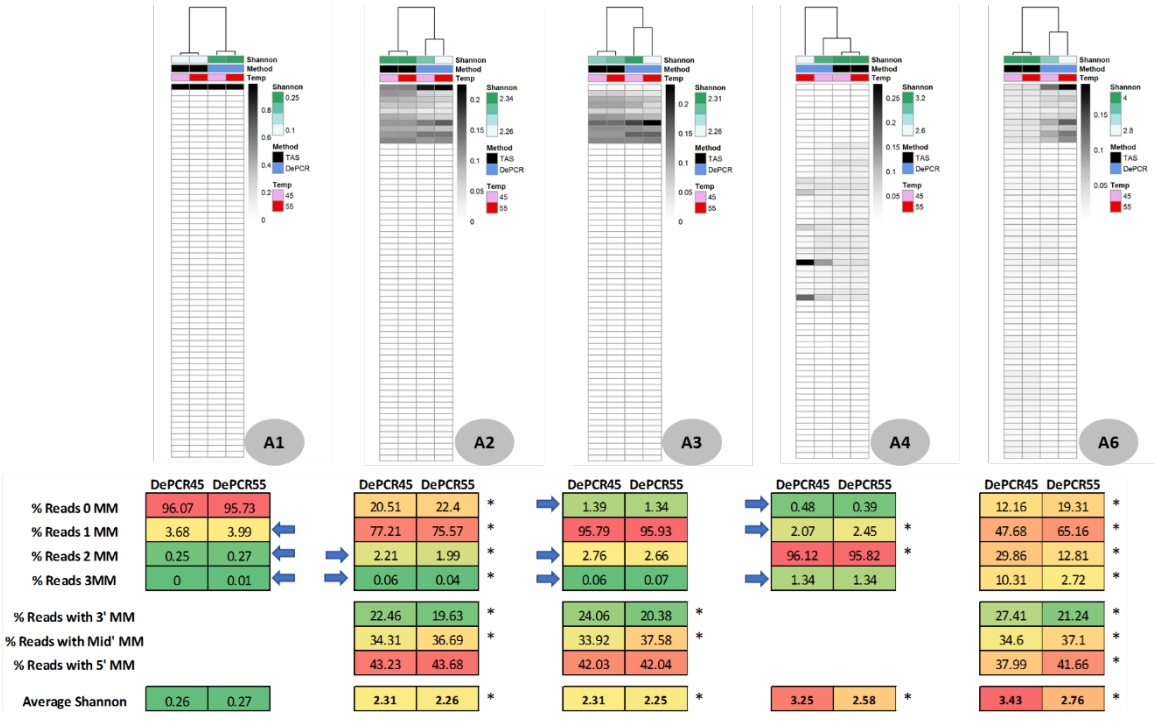


Figure 20: Effect of PCR methodology and annealing temperature on PUPs in reactions with a single template

In experiments A1-6, only template ST1 was added to amplification reactions, while primer pools varied (**Table 10**). One-way clustered heatmaps of untransformed primer variant utilization during amplification of varying primer pools ("A1" = 1 primer, "A2" = 10 primers, "A3" = 9 primers, "A4" = 27 primers, and "A6" = 64 primers). Samples (columns) are color-coded by amplification method (TAS or DePCR), amplification annealing temperature (45°C or 55°C), and average Shannon index of primer utilization. Each column represents the average of 7-8 technical replicates per condition and rarefaction to 7,000 sequences/replicate. Primers (rows) represent all 64 primer variants (806F_v1 – 806F_v64). Percentage of reads with mismatches (0, 1, 2 and 3 mismatches) in amplifications using DePCR are shown in tables below each heatmap. Distribution of position of mismatches (3', middle and 5' mismatch positions) for all reads with one mismatch are also shown. Asterisks indicate significant differences in measured values by annealing temperature (ANOVA, $P < 0.01$). Intensity scales vary between experiments. Certain values represent PCR errors generated during polymerase copying of primer regions, and these are indicated by blue arrows. These known errors are identified by primer-template combinations unavailable in each experiment. Single mismatch positional analysis is not shown for experiments A1 and A4 due to the absence of single mismatch interactions between primers and ST1 template.

Very small effects of annealing temperature on PUPs were observed for TAS amplifications, while significant effects of annealing temperature were observed on PUPs generated using DePCR. An increase in annealing temperature from 45°C to 55°C in DePCR amplifications (except experiment A1) led to reduced Shannon indices for PUPs, with one or several primers becoming increasingly dominant at higher annealing temperatures (**Figure 20**). In experiment A4,

in which a pool of 27 primers each with two mismatches to the ST1 template was used, two primer variants were dominant, particularly at 55°C. These two dominant primers (806F_v47 and 806F_v63) had only 5' and middle mismatches with template ST1, and the mismatch types were primarily A/G mismatches.

We next examined the utilization of primers perfectly matching templates and those with 1, 2, or 3 mismatches to templates in DePCR-amplified reactions. When present, perfect match primers had the highest utilization rate of any single primer (**Figure 20, A2 and A6**). However, the rate of utilization of the primer perfectly matching the ST1 template (*i.e.*, 806F_V1) ranged from approximately 12.2% to 22.4%, depending on annealing temperature and primer pool composition. As show in **Figure 20**, even with perfect match primers available, amplification of the ST1 template was predominantly performed by primers with mismatches. When a heavily degenerate primer pool was employed (64 primers; experiment A6), triple mismatch primers contributed to greater than 10% of reads at 45°C annealing temperature.

We further examined primer-template annealing with regard to position of mismatch. In DePCR amplifications where primers had a single mismatch with the ST1 template, we calculated the percentage of mismatches at the -2 (3'), -8 (middle), and -14 (5') positions. We observed a general trend towards greater utilization of primers with 5' mismatches relative to middle and 3' mismatches, and lowest utilization of 3' mismatched primers. However, 3' mismatched primers amplified a substantial percentage of ST1 template, representing 19-27% of single-mismatch reads, depending on annealing temperature and primer pool. With increasing annealing temperature, the utilization of single mismatch primers with the mismatch at the 3' position decreased significantly but was never below 19% (**Figure 20**).

In the A1 experiment, only a single primer perfectly matching the ST1 template was included. However, we observed that approximately 4% of reads that contained 1, 2, or 3 mismatches. These reads with mismatches represent polymerase error. Specifically, DePCR has a higher observed error rate in the primer site, because the primer sites are copied during amplification, allowing polymerase mistakes to become incorporated. Conversely, in TAS, the primer site sequences are derived directly from the synthesized oligonucleotide primers, and only experience polymerase copying during bridge amplification on the Illumina sequencer. Similar overall rates of known error in primer site attribution of approximately 2-4% were observed in experiments A2 (only perfect match and single mismatch primers added to the reactions), A3 (only single mismatch primers added to the reactions) and A4 (only double mismatch primers added to the reactions). No direct measurement could be made for experiment A6, as all primers, with 0-3 mismatches with the ST1 template, were added to the reactions.

Interrogation of multi-template pools with a non-degenerate primer set

We interrogated multiple template pools (A, B, C, D and E; **Table 10**) with a single primer (806F_v1) which perfectly matched template ST1 and had single mismatches with all other templates (*i.e.*, ST4, ST6, ST7, ST8, ST11, ST15, ST23, ST39, and ST55) (**Figure 21**). DePCR was superior to the TAS for reproducing the expected template distribution in all experiments except for A1 (**Figure 21**).

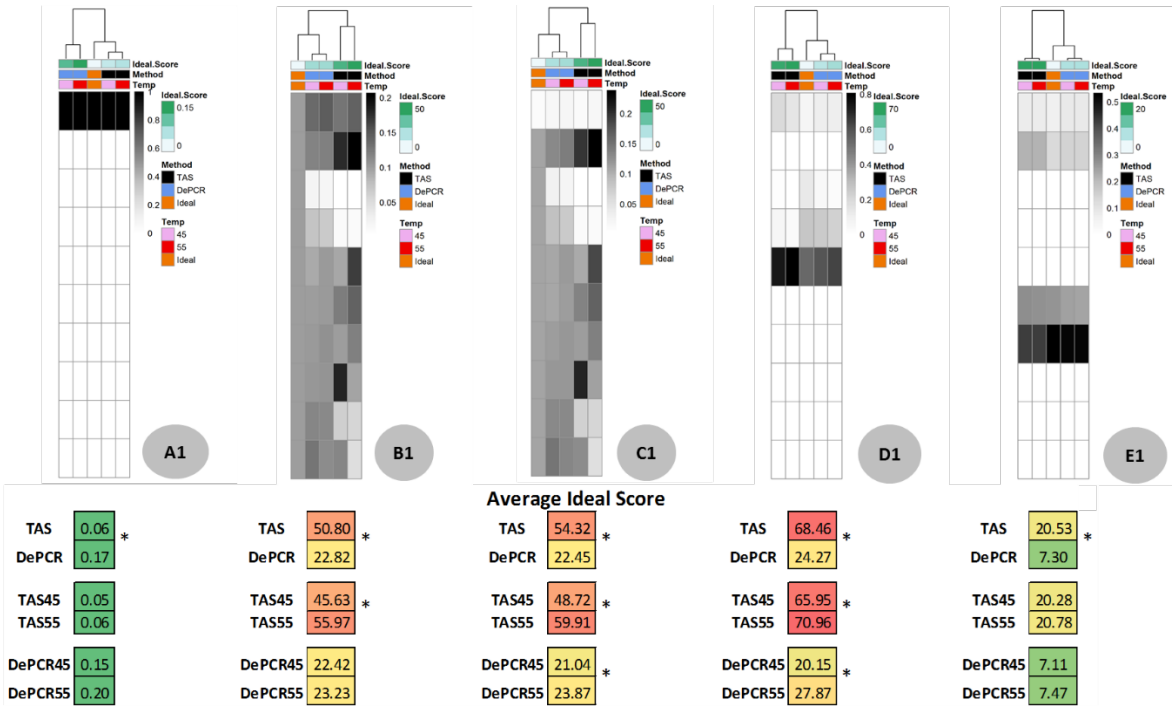


Figure 21: Effect of PCR methodology and annealing temperature on template profiles in amplification reactions utilizing a single primer

One-way clustered heatmaps of untransformed template utilization profiling during amplification with a single primer (806F_v1) with a varying range of templates ("A1" = 1 template, "B1" = 10 templates, "C1" = 10 templates, "D1" = 4 templates, and "E1" = 4 templates) as described in text. Samples (columns) are color-coded by amplification method (TAS or DePCR), amplification annealing temperature (45°C or 55°C), and average Ideal score. Each column represents the average of 7-8 technical replicates per condition and rarefaction to 7,000 sequences/replicate. Templates (rows) represent all 10 templates (ordered from top to bottom; ST1, ST4, ST6, ST7, ST8, ST11, ST15, ST23, ST39, and ST55). Ideal score comparisons between TAS and DePCR (across both annealing temperatures), within TAS (45°C or 55°C), and within DePCR (45°C or 55°C) are shown in tables. Asterisks indicate significant differences in

measured values by annealing temperature (ANOVA, $P < 0.01$). Intensity scales vary between experiments.

This was determined by calculation of the Ideal Score, which represents a summation of difference in relative abundance for each feature from the expected relative abundance, and mMDS profiles and template heatmaps where the expected template structure clustered with DePCR profiles (**Figures S11-S26**). Higher values represent a greater distortion of the expected structure. Lower Ideal scores were observed at the lower annealing temperature of 45°C relative to annealing temperatures of 55°C, for both TAS and DePCR (**Figure 21**).

Of the ten templates, templates ST6 and ST7 proved difficult to amplify using either TAS or DePCR methods at either annealing temperatures, and regardless of which template pool was used (**Figure 21**). The ST6 and ST7 templates each have a single 3' mismatch with the 806F_v1 primer (primer A annealing to template G or template A). Conversely, template ST8, with a 3' mismatch (primer A annealing to template C) could be amplified with both TAS and DePCR (**Figure 21**; **Table S11**). Although poorly amplified, template ST6 could be amplified with primer 806F_v1 using DePCR at an average rate of approximately 2.1% of all reads in comparison to 0.3% for TAS (experiment B1, annealing temperature 45°C; ANOVA $P<0.001$). Similarly, template ST7 could be amplified with primer 806F_v1 using DePCR at an average rate of approximately 6.7% of all reads in comparison to 1.2% for TAS (ANOVA $P<0.001$).

Interrogation of complex template pools with degenerate primer pools

We interrogated multi- template pools (B, C, D and E; **Table 10**) with degenerate primer pools to determine if such pools could improve recovery of expected template distribution relative to non-degenerate primers as shown above. Results from the "B" experiment, with 10 unique templates, are shown in **Figures 22, 23 and S16-S19**. Results from the "C" experiment, with 10 unique templates but with ST1 at $1/10^{th}$ concentration are shown in **Figures S20-S22**. Results from the "D" experiment, with four unique templates (including ST1 and three 3' single mismatch templates) at graduated concentrations are shown in **Figures S23-S24**. Results from the "E" experiment, with four unique templates (including ST1 and three middle position single mismatch templates) at graduated concentrations are shown in **Figures S25-S26**.

Amplification method (DePCR or TAS) yielded significantly different PUPs in "B" experiments with 10 templates and varying number of primers (**Figure 22).** As above, TAS amplification 'scrambles' the PUP signature, leading to highly even primer utilization with high Shannon index. When using the DePCR methodology at 45°C and employing 10 primers, each matching a single template perfectly (experiment B2), the observed Shannon Index approached that observed in the TAS reactions (Shannon index ranging from 2.31 to 2.34 between DePCR and TAS; **Figure 22**). In experiments B2 and B3 which utilized 10 or 9 primers, perfect match amplification was particularly favored at the higher annealing temperature of 55°C and this correlated with lower Shannon Index. Although perfect match amplification was higher than for "A" experiments in which only a single primer was utilized, perfect match annealing never contributed more than 50% of all observed sequencing reads, across all temperature and primer pools (**Figure 22**). In experiment B1, where only a single primer matching the ST1 template was used, perfect match annealing represented approximately 14-17% of all reads. With 10 primers,

each perfectly matching one of the 10 templates, perfect match annealing represented approximately 29-48%, with the higher value occurring at the 55°C annealing temperature (**Figure 22**). Two mismatch annealing interactions contributed substantially at 45°C, but not nearly as much at 55°C. As observed previously, 5' mismatch annealing interactions were generally favored relative to middle and 3' mismatches.

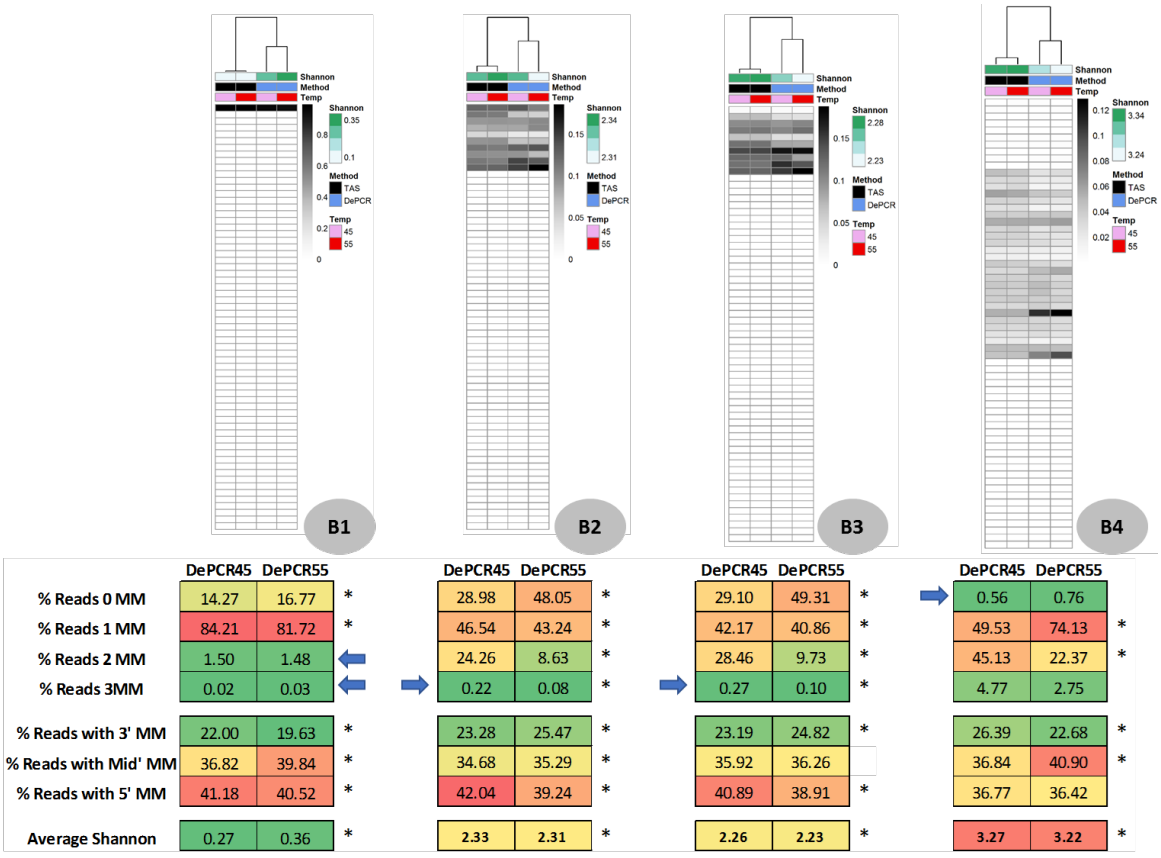


Figure 22: Effect of PCR methodology, annealing temperature, and primer pool on PUPs of ten templates

In experiments B1-4, all ten synthetic DNA templates were added to amplification reactions at equimolar concentrations, while primer pools varied (**Table 10**). One-way clustered heatmaps of

untransformed primer variant utilization during amplification of varying primer pools ("B1" = 1 primer, "B2" = 10 primers, "B3" = 9 primers, and "B4" = 27 primers). Samples (columns) are color-coded by amplification method (TAS or DePCR), annealing temperature (45°C or 55°C), and average Shannon index of primer utilization. Each column represents the average of 8 technical replicates per condition and rarefaction to 7,000 sequences/replicate. Primers (rows) represent all 64 primer variants (806F_v1 – 806F_v64). Percentage of reads with mismatches (0, 1, 2 and 3 mismatches) in amplifications using DePCR are shown in tables below each heatmap. Distribution of position of mismatches (3', middle and 5' mismatch positions) for all reads with one mismatch are also shown. Asterisks indicate significant differences in measured values by annealing temperature (ANOVA, $P < 0.01$). Intensity scales vary between experiments. Certain values represent PCR errors generated during polymerase copying of primer regions, and these are indicated by blue arrows. These known errors are identified by primer-template combinations unavailable in each experiment.

We next examined template profiles generated with these complex template and primer pools. As before, we observed that the DePCR method generated profiles significantly closer to the expected template distribution, relative to amplification using TAS, as assessed by Ideal scores (**Figure 23**). Using a single primer with the DePCR method generated a relatively high Ideal Score (approximately 23) but increasing primer pool complexity led to improved accuracy of profile **(Figures 23, S27, S28**). Unlike experiments with a single primer, we observed that increasing annealing temperature generated significantly better template profiling (*i.e.*, Ideal scores) when 10 templates and 9 or 10 perfect match primers were used (Experiments B2 and B3; **Figure 23**). When a broad range of mismatch primers (pool of 27 primers with 2 mismatches

to ST1 and 1-3 mismatches to all other templates) was used with the DePCR method, the lowest Ideal scores (highest accuracy) were generated, and no significant effect of annealing temperature was observed **(Figure 23**). The ST6 and ST7 templates continued to be difficult to amplify with TAS even with greater numbers of primers or low annealing temperature (*e.g.*, Experiments B2 and B3, **Figure 23**). When amplified using DePCR with pools of 9, 10 or 27 primers, templates ST6 and ST7 were robustly amplified relative to DePCR with only a single primer (*i.e.*, Experiment B1, **Figures 22 and 23**). The use of greater number of primers, therefore, directly contributed to the significantly lower Ideal scores observed in Experiment B2, B3 and B4 relative to B1. The lowest Ideal scores were generated using DePCR without any perfect match primers (*i.e.*, Experiment B4, **Figure 23**).

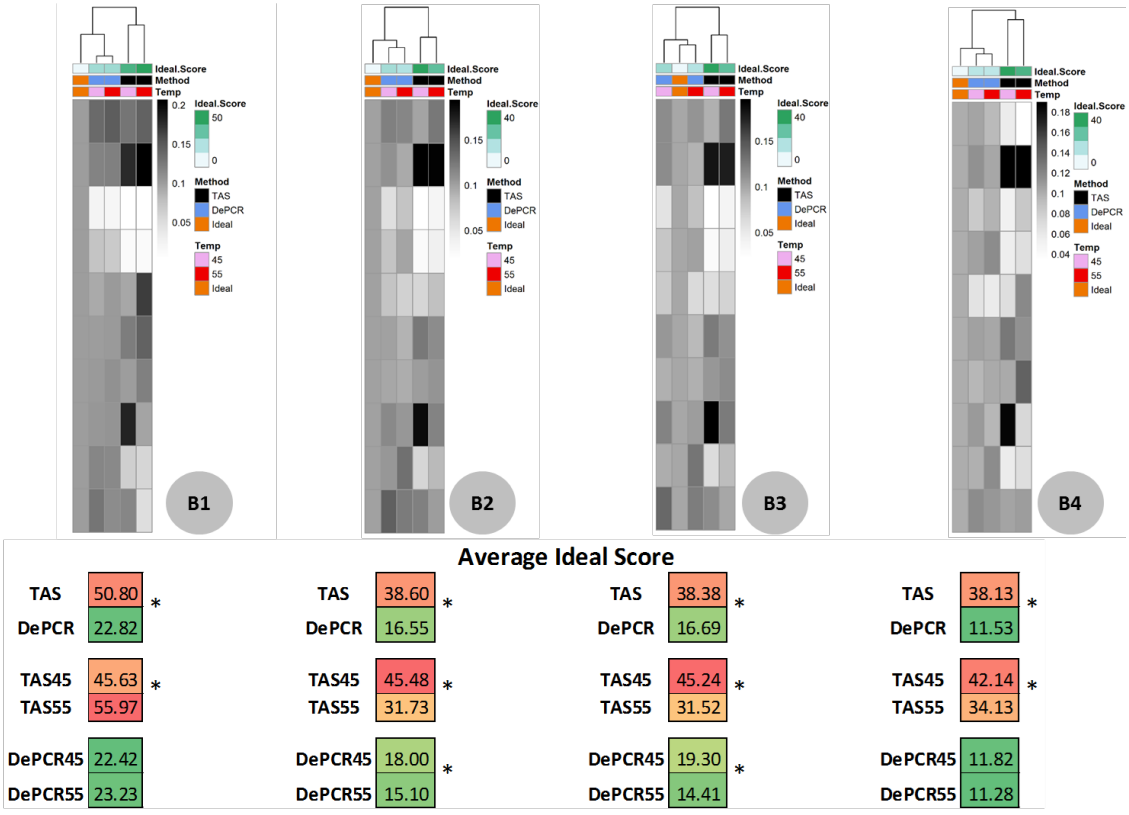


Figure 23: Effect of PCR methodology and annealing temperature on template profiles in amplification reactions utilizing varying primer pools

One-way clustered heatmaps of untransformed template utilization profiling during amplification of an equimolar pooling of all ten synthetic DNA templates and varying primer pools ("B1" = 1 primer, "B2" = 10 primers, "B3" = 9 primers, and "B4" = 27 primers) as described in text and **Table 10**. Samples (columns) are color-coded by amplification method (TAS or DePCR), amplification annealing temperature (45°C or 55°C), and average Ideal score. Each column represents the average of 7-8 technical replicates per condition and rarefaction to 7,000 sequences/replicate. Templates (rows) represent all 10 templates (ordered from top to bottom; ST1, ST4, ST6, ST7, ST8, ST11, ST15, ST23, ST39, and ST55). Ideal score comparisons between TAS and DePCR (across both annealing temperatures), within TAS (45°C or 55°C), and within DePCR (45°C or 55°C) are shown in tables. Asterisks indicate significant differences in measured values by annealing temperature (ANOVA, $P < 0.01$). Intensity scales vary between experiments.

Primer utilization profiles for each template within a complex template pool

Using the DePCR methodology and experimental setup described here, we were able to recover PUPs for each template independently. For example, in experiment B2, a total of 10 templates were pooled and 10 primers used for amplification. PUPs presented in **Figure 22** represent average primer utilization across all templates. PUPs presented in **Figure 24** present primer utilization for each of the 10 templates in experiment B2 at 45° and 55°C annealing temperatures.

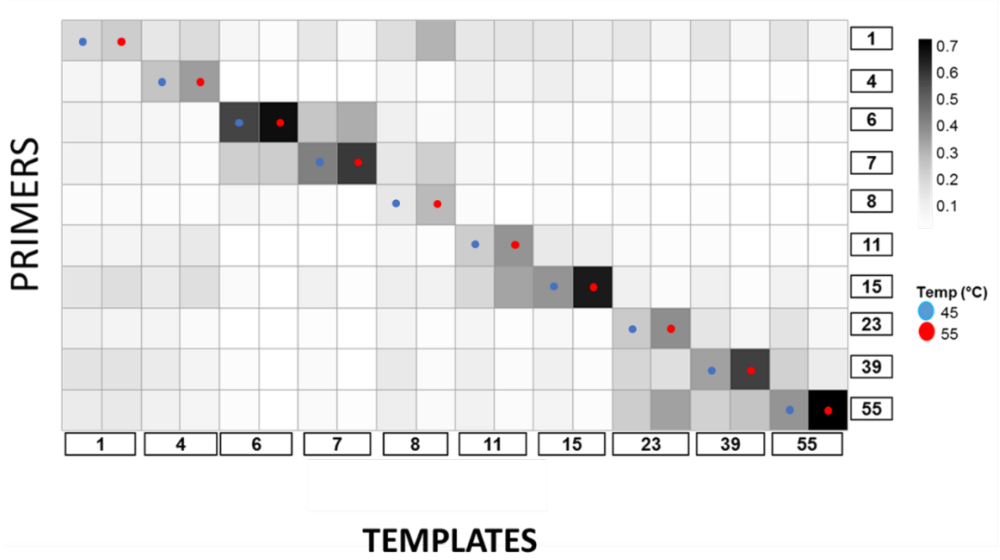


Figure 24: Template-specific primer utilization profiling

In experiment B2, all 10 DNA templates were amplified with a pool of 10 primers, each perfectly matching a single template, and with 1-3 mismatches with the remaining 9 templates. PUPs for each template were separated from the averaged PUPs shown in **Figure 23.** Primer utilization is shown for annealing temperature of 45°C and 55°C. Blue dots indicate perfect match annealing at an annealing temperature of 45°C, and red dots indicate perfect match annealing at 55°C. For each primer-template combination, the gray-scale intensity is proportional to the relative abundance of reads with that combination.

In experiment B2, two patterns were observed in template-specific PUPs: (a) dominant annealing to templates with perfect match primers and one or two other primers (*i.e.*, templates ST4, ST6, ST7, ST11, ST15, ST23, ST39, and ST55), and (b) broad annealing to templates with multiple primers (*i.e.*, templates ST1 and ST8). In templates that favored amplification by perfect match primers, a strong effect of annealing temperature was observed, increasing perfect match

annealing at higher annealing temperatures (**Figure 24**). The even utilization pattern observed for the ST1 template is likely a result of the large number of single mismatch primers available to anneal to the template (*i.e.*, of the 10 available primers in experiment B2, 1 primer matched the ST1 template perfectly, while the remaining 9 primers each had a single mismatch with ST1). Conversely, for all other templates, there was a mixture of 1 perfect match, 3 single mismatch, and 6 double mismatch primers. The ST8 template was unique – with a broad PUP at 45°C and a much lower diversity profile at 55°C. This template was the only one with a nucleotide of G at the -2 position on the 5'-3' strand (**Table S11**).

Discussion

PCR bias has been thoroughly studied, and a wide range of factors contributing to bias are known. In particular, PCR selection – wherein factors within PCR preferentially amplify some templates [15] – can strongly distort underlying biological structure. We focus in this study on primer-template interactions, as mismatches are known to lead to selective amplification, and poor representation of source template structure [13, 55]. As has been shown previously, templates with mismatches to primers can be difficult to detect, and mismatches close to 3' ends are particularly damaging [67, 68]. We previously developed a novel method for reducing PCR bias [18, 96], and one of the features of this method is the ability to measure primer-template annealing and elongation events empirically. Thus, we sought to use this method ('DePCR') to explore primer-template interactions in a systematic manner under controlled experimental conditions. The fundamental questions of this study included: (a) Is the DePCR method an improvement over standard amplification methods for maintaining the underlying community structure after amplification in systems with complex primer pools and template pools?, (b) Do

perfect match primer-template interactions dominate in PCRs?, (c) Can we quantify the effect of mismatch position on template amplification?, (d) How does annealing temperature alter primer-template interactions?, and (e) How effective are non-degenerate primers for amplification of complex templates?

We previously developed the DePCR methodology to reduce bias associated with PCR amplification of complex DNA templates [18, 96]. In the original study [18], we identified a novel source of PCR bias – namely, the combined action of linear copying of genomic DNA templates and exponential amplification of DNA copies generated during PCR. Furthermore, in standard PCR, lower efficiency primer-DNA template interactions are compounded over many cycles of amplification. To alleviate this, DePCR limits primer-template interactions to the first two cycles of linear copy, and additional PCR bias is avoided by performing exponential amplification using primers targeting only non-degenerate adapter sequences. A second benefit of this approach is that the locus-specific primers that anneal to DNA templates and are used to initiate polymerase copying are preserved. After linear copying, exponential amplification is performed with primers that do not contain any locus-specific information, and therefore do not continuously interact with locus-specific primer sites, as is common in standard PCR amplification reactions. As such, DePCR provides an unprecedented view into primer-template interactions; so-called primer utilization profiles (PUPs) represent data that cannot be generated in any other manner. Conversely, standard PCR (TAS) is definitively shown to 'scramble' primer utilization profiles, as locus-specific primers are used to copying DNA templates and DNA copies throughout the exponential cycles of PCR.

We previously showed that the DePCR method improved the representation of a mock community of known composition when compared to standard TAS amplification [18]. In that

study, however, the mock community was comprised of only 4 templates, with relatively low complexity. In a second manuscript, describing the development of an improved DePCR workflow, we examined effects of annealing temperature, template concentration and primer degeneracy on the observed microbial community structure in gDNA derived from mammalian feces [96]. Thus, in this manuscript, we sought to systematically explore primer-template interactions with the novel workflow but with a more complex mock community. By utilizing a suite of experiments with different template and primer complexity, we demonstrate here that the DePCR methodology consistently improves sequence-based representation of complex communities. This is shown through the calculation of a univariate metric – the Ideal score – which is a summation of divergence from the expected underlying distribution and the observed distribution of reads from each template in a known pool of templates. Ideals scores, except for the series of "A" experiments which contained only a single template, were substantially and significantly lower for all experiments run with DePCR relative to TAS. The improved accuracy of the DePCR method is derived from several basic mechanisms. First, Suzuki and Giovanonni [28] demonstrated that the evenness of amplifications products is dependent on the efficiency of polymerase copying during each amplification cycle. Thus, bias can be modeled by a formula including molarity of starting template, amplification efficiency of each template, and number of cycles (*i.e.* formula 3, Suzuki and Giovanonni [28]). In DePCR, only two cycles of amplification with locus-specific primers are used, thus, bias derived from differing amplification efficiency is greatly limited. A second mechanism is the difference between amplification efficiency associated with primer-template interactions and efficiency associated with primer-amplicon interactions [18]. For example, in microbial DNA samples, when primers anneal to gDNA templates, the potential positions and numbers of mismatches is very large due to high sequence

diversity of ribosomal RNA genes, even in highly conserved primer regions [42, 43]. However, when primers interact with PCR copies, the primer region represents the synthetic oligonucleotide primers rather than the original gDNA sequence, thereby limiting the number of possible primer-template interactions. The combination of both linear copying of gDNA templates and copying of PCR copies during exponential amplification cycle leads to complex rates of amplification efficiency [18]. Using DePCR, this second form of bias is removed, as locus-specific primer-PCR copy interactions are removed completely. Finally, we previously demonstrated that DePCR lowers detectable chimera rates significantly, and this too can contribute to lower overall distortion of underlying community structures [96].

We observed that an additional feature of DePCR was a greater tolerance for mismatches relative to TAS. Detrimental effects of primer-template mismatches have been previously studied, including a system in which base alterations were introduced into 21 primers and 19 DNA templates [68]. Among other findings, Bru et al. [68] observed that mismatches closest to the 3' end of primers were the most detrimental to PCR efficiency, leading to as great as a one log underestimation of gene copy number in quantitative PCR assays. However, other studies have shown small or no effects of 3' mismatches [95]. In our study, we observed that both number of mismatches and inclusion of 3' mismatches lowered amplification efficiency. For example, certain synthetic templates (*e.g.*, ST6, with a 3' mismatch) were poorly amplified under many PCR conditions, including conditions in which a perfect match primer was available (*i.e.*, experiment B2). However, as primer diversity increased, ST6 amplification did not greatly improve with TAS PCR. Using the DePCR method, however, template ST6 could be routinely amplified provided that degenerate primer pools were employed. The improved amplification of such templates with DePCR is in part due to the fact that low efficiency primer annealing and

elongation is limited to 2 cycles only. Across all datasets with more than a single template and primer, primer-template interactions containing single mismatches had efficiency profiles with 5' mismatches > middle mismatches > 3' mismatches. However, 3' mismatches were still tolerated. Wu et al. [67] observed that mismatches within the last 3-4 bases of primers led to almost complete lack of amplification; however, this is likely a result of low amplification efficiency compounded over 30 cycles of PCR. Such low efficiency can lead to distorted microbial community structures, and even loss of phylum-level detection in environmental samples [55].

We demonstrate here that primer-template interactions favor perfect matches, but not overwhelmingly so. In fact, most annealing and copying in the DePCR experiments was conducted with primers that did not perfectly match templates, even during the very first cycles of PCR when no primers are limiting. Although efficiency of amplification using primer-template interactions with more than one mismatch is lower than perfect matching amplification, reasonable amplification was possible even with one, two or three mismatches using DePCR. Interestingly, in experiment B3, the removal of a primer perfectly matching one of the ten templates (806F_v1, matching template ST1) did not substantially decrease the ability of the primer pool to profile the mock template community, in part due to the presence of nine primers, each with a single mismatch to the ST1 template. The tolerance of mismatches occurs during the first two cycles of PCR, when all primer variants are present at equal concentrations and perfect match primers are available at high concentration. We observed that in the B2 experiment (10 templates and 10 primers, with each primer perfectly matching one template), perfect match interactions were most heavily favored, but still only represented 29% (45°C annealing temperature) or 48% (55°C annealing temperature) of amplicons. This was further shown to be template and primer-pool dependent. Based on these results, it appears that when there are a

matched number of templates and perfect matching primers, higher annealing temperatures are favored to profile complex template mixtures. However, this condition is extremely unlikely in natural environments, where numerous and potentially unpredictable mismatches are possible. When using the DePCR method, the PCR amplification system can amplify mock community DNA templates even with primers that have a minimum of 1 or 2 mismatches with all templates (*i.e.*, experiment B4). The use of 10 perfectly matching primers was less successful at evenly amplifying the 10 templates than were 27 primers, each with 1-3 mismatches with each of the templates (Ideal score of 15.1-18.0 for experiment B2 relative to 11.3 to 11.8 for experiment B4). However, this phenomenon was not observed for standard (TAS) amplification. We suggest that for environmental systems, tolerance of mismatches is better for recovering the underlying structure, particularly when employing DePCR.

Within DePCR experiments, annealing temperature played a strong role in determining PUPs, and in some experiments also significantly altered Ideal scores. In experiments without degenerate primer pools (*i.e.*, with only a single 806F primer variant – experiments A1, B1, C1, D1 and E1), lower annealing temperatures led to significantly improved representation of the mock communities. In systems where the number of templates were matched or nearly matched with perfect matching primers (*i.e.*, B2 and B3), PCR conditions favoring perfect match interaction (*i.e.*, elevated annealing temperature) led to improved representation. This was not the case for experiment B4, in which all primer-template interactions were mismatch interactions; here, no significant effect of annealing temperature was observed. Analysis of the PUPs indicate that lower annealing temperature is more tolerant of 3' mismatches, and this leads to greater evenness (high Shannon index) of primer utilization. We previously observed a quadratic relationship between annealing temperature in DePCR and Shannon index of PUPs within a

complex microbial sample [96]. This temperature relationship with primer utilization is confirmed here, and we also demonstrate that the shift towards lower evenness of primer utilization is a shift towards higher rate of perfect match annealing. This observation is consistent with very early studies of primer-template interactions showing that increased annealing temperature reduced mis-extension of incorrect nucleotides at the 3' ends of primers [33]. As we demonstrated previously, the shift in primer utilization associated with annealing temperature in DePCR leads to a shift in the observed complex template structure.

Conclusions

We provide a novel strategy for exploring primer-template interactions, providing a mechanism for acquiring previously inaccessible information. Some phenomena are confirmed – 3' mismatches are destabilizing, and perfect matches favored. Other phenomena are novel: perfect matches may be favored, but mismatch primer-template annealing is the dominant type of interaction, and non-perfect match copying starts immediately during the first cycles of PCR, not in later cycles. Primer-template interactions can tolerate multiple mismatches without dramatic effect on observed community structure when employing the DePCR methodology. We establish here an experimental system for interrogating primer-template interactions, by providing a mechanism for identifying perfect match and mismatch primer-template interactions. Such an experimental system has broad applicability and will provide empirical evidence for future studies of primer design. Ultimately, we sought to better understand the relationship between primers and templates, particularly with regard to mismatch tolerance, to help improve the design of primer pools for amplification of complex environmental samples. Caveats of this study include: (a) study was performed with synthetic DNA templates, and not more complex

environmental samples; and (b) the standard polymerase used in this study introduced sequence errors creating limited uncertainty regarding exact primer utilization profiles. In future studies, proof-reading enzymes can be used to reduce such error.

Acknowledgements

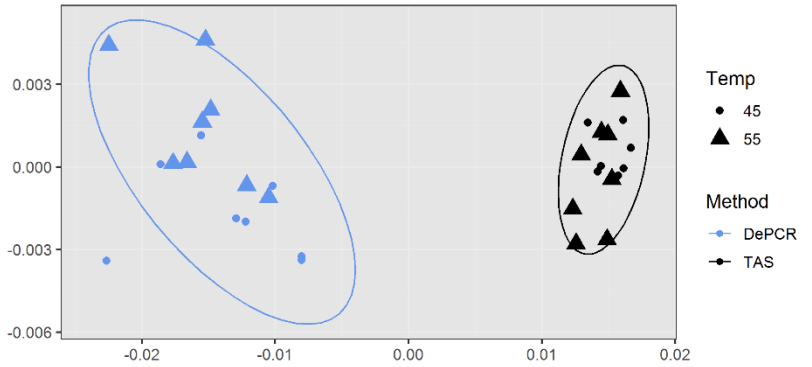
We greatly acknowledge the support of the members of the University of Illinois at Chicago Sequencing Core for assistance in the study.

Supporting Information

Figure S11

Primer Utilizing Profiles

Experiment: A1
Primers: 1
Templates: 1

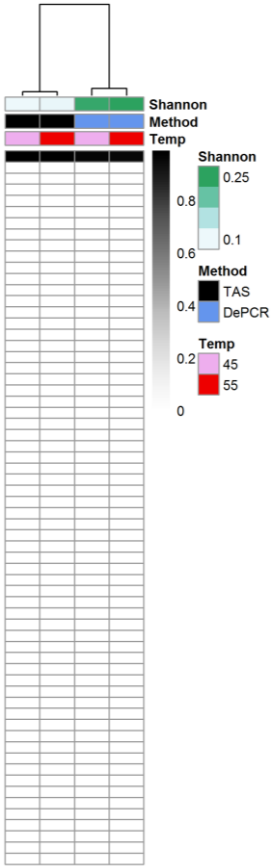


ANOSIM

DePCR vs TAS:
R=1, P=0.0001

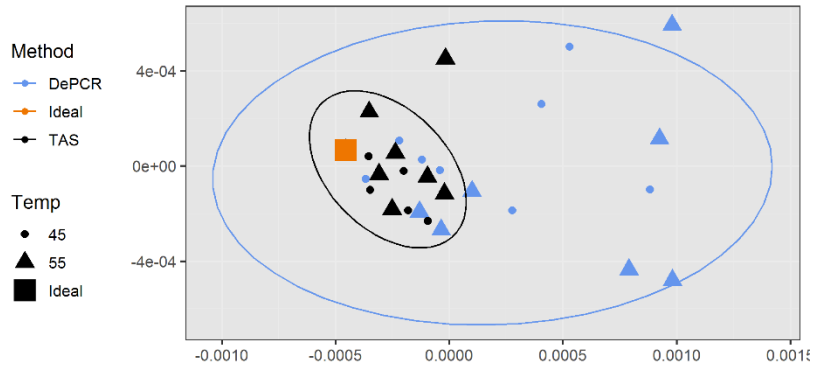
TAS45 vs TAS55:
R=0.12; p=0.0012

DePCR45 vs DePCR55:
R=0.099, P=0.109



Comparison	DePCR45	DePCR55	ANOVA
% Reads 0 MM	96.07	95.73	0.07
% Reads 1 MM	3.68	3.99	0.06
% Reads 2 MM	0.25	0.27	0.62
% Reads 3MM	0.00	0.01	0.28
% Reads with 3' MM	26.81	25.29	0.23
% Reads with Mid' MM	30.87	34.19	0.01
% Reads with 5' MM	42.32	40.51	0.17
Average Primer Tm	57.42	57.43	0.02
Average Shannon	0.26	0.27	0.41

Template Profiles

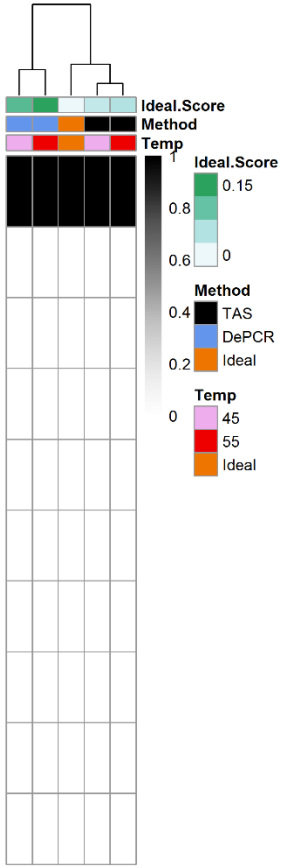


ANOSIM

DePCR vs TAS:
R=0.201, P=0.0003

TAS45 vs TAS55:
R=-0.049; p=0.762

DePCR45 vs DePCR55:
R=-0.015, P=0.478

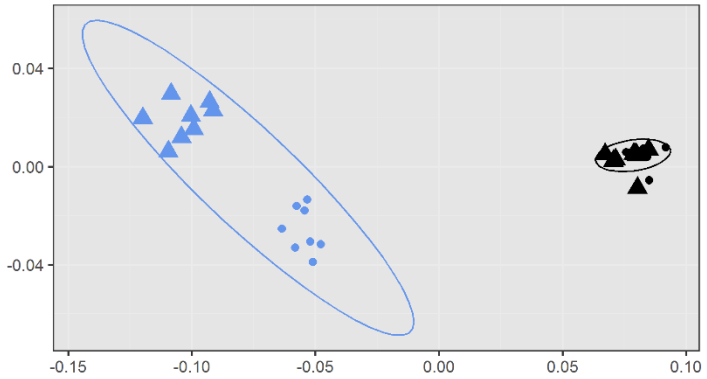


	Comparison	Average Ideal Score	ANOVA
1	TAS	0.06	0.00
	DePCR	0.17	
2	TAS45	0.05	0.39
	TAS55	0.06	
3	DePCR45	0.15	0.38
	DePCR55	0.20	

Figure S12

Primer Utilizing Profiles

Experiment: A2
Primers: 10
Templates: 1

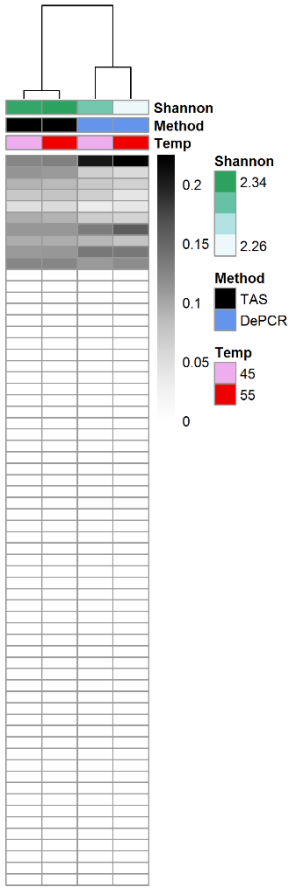


ANOSIM

DePCR vs TAS:
R=1, P=0.0001

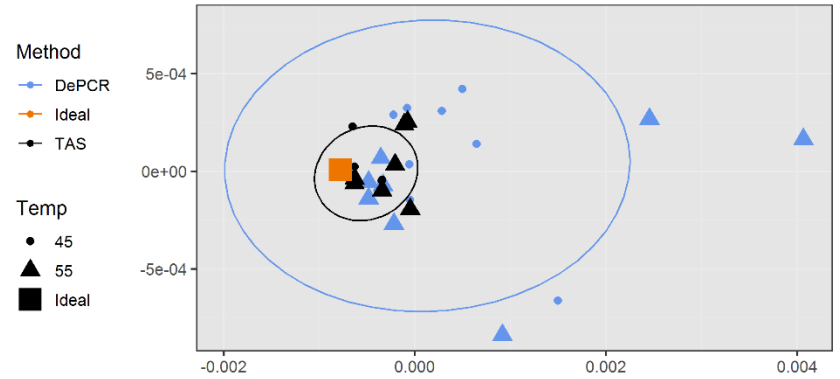
TAS45 vs TAS55:
R=0.425; p=0.0009

DePCR45 vs DePCR55:
R= 1, P=0.0002



Comparison	DePCR45	DePCR55	ANOVA
% Reads 0 MM	20.51	22.40	0.00
% Reads 1 MM	77.21	75.57	0.00
% Reads 2 MM	2.21	1.99	0.03
% Reads 3MM	0.06	0.04	0.05
% Reads with 3' MM	22.46	19.63	0.00
% Reads with Mid' MM	34.31	36.69	0.00
% Reads with 5' MM	43.23	43.68	0.28
Average Primer Tm	57.79	57.84	0.00
Average Shannon	2.31	2.26	0.00

Template Profiles

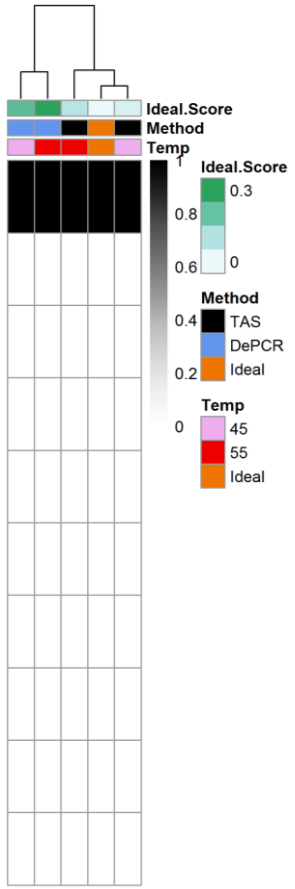


ANOSIM

DePCR vs TAS:
R=0.169, P=0.0003

TAS45 vs TAS55:
R=0.183; p=.031

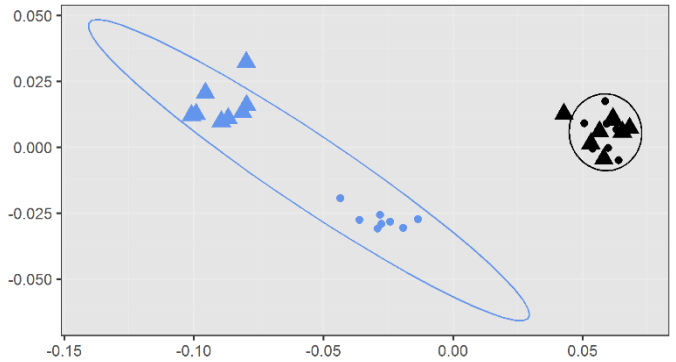
DePCR45 vs DePCR55:
R=0.063, P=0.00063



	Comparison	Average Ideal Score	ANOVA
1	TAS	0.06	0.00
	DePCR	0.26	
2	TAS45	0.04	0.04
	TAS55	0.09	
3	DePCR45	0.22	0.53
	DePCR55	0.30	

Figure S13

Primer Utilizing Profiles

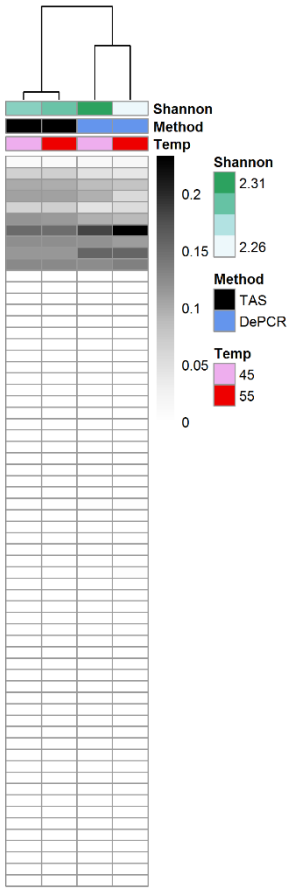


ANOSIM

DePCR vs TAS:
R=0.968, P=0.0001

TAS45 vs TAS55:
R=0.207; p=0.02

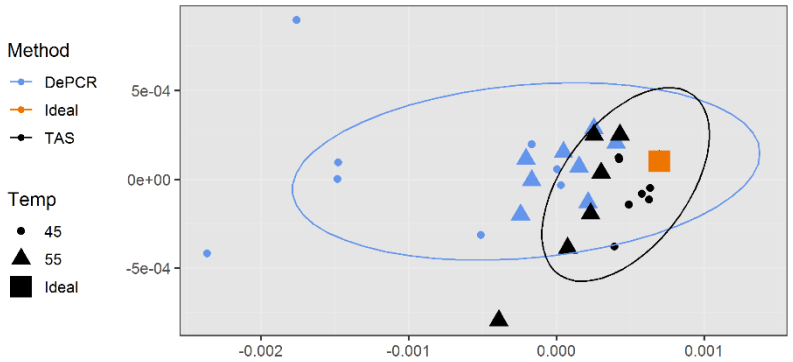
DePCR45 vs DePCR55:
R=1, P=0.0002



Comparison	DePCR45	DePCR55	ANOVA
% Reads 0 MM	1.39	1.34	0.50
% Reads 1 MM	95.79	95.93	0.28
% Reads 2 MM	2.76	2.66	0.43
% Reads 3MM	0.06	0.07	0.59
% Reads with 3' MM	24.06	20.38	0.00
% Reads with Mid' MM	33.92	37.58	0.00
% Reads with 5' MM	42.03	42.04	0.98
Average Primer Tm	57.81	57.95	0.00
Average Shannon	2.31	2.25	0.00

Experiment: A3
Primers: 9
Templates: 1

Template Profiles

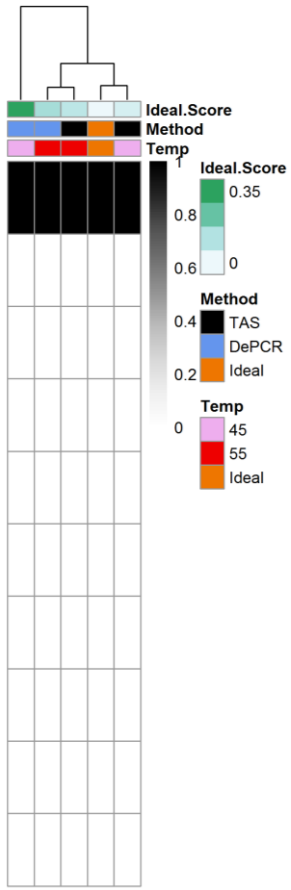


ANOSIM

DePCR vs TAS:
R=0.25, P=0.0001

TAS45 vs TAS55:
R=0.064; p=0.137

DePCR45 vs DePCR55:
R=0.26, P=0.006

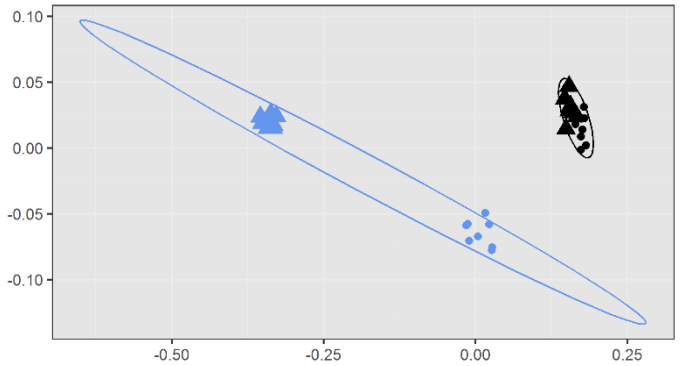


	Comparison	Average Ideal Score	ANOVA
1	TAS	0.08	0.00
	DePCR	0.25	
2	TAS45	0.05	0.17
	TAS55	0.10	
3	DePCR45	0.36	0.00
	DePCR55	0.14	

Figure S14

Primer Utilizing Profiles

Experiment: A4
Primers: 27
Templates: 1

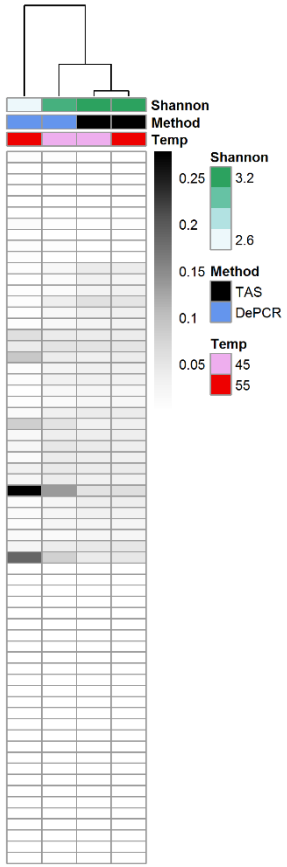


ANOSIM

DePCR vs TAS:
R=0.733, P=000.01

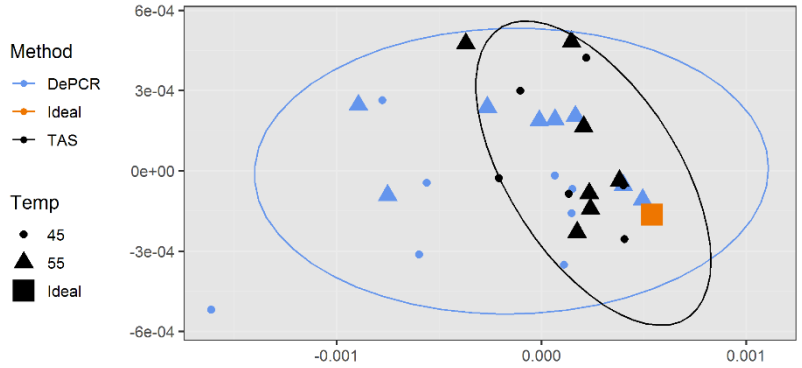
TAS45 vs TAS55:
R=0.894; p=0.0002

DePCR45 vs DePCR55:
R=1, P=0.0002



Comparison	DePCR45	DePCR55	ANOVA
% Reads 0 MM	0.48	0.39	0.12
% Reads 1 MM	2.07	2.45	0.01
% Reads 2 MM	96.12	95.82	0.04
% Reads 3MM	1.34	1.34	0.95
% Reads with 3' MM	25.84	17.10	0.00
% Reads with Mid' MM	37.01	44.30	0.00
% Reads with 5' MM	37.15	38.59	0.00
Average Primer Tm	58.56	58.97	0.00
Average Shannon	3.25	2.58	0.00

Template Profiles

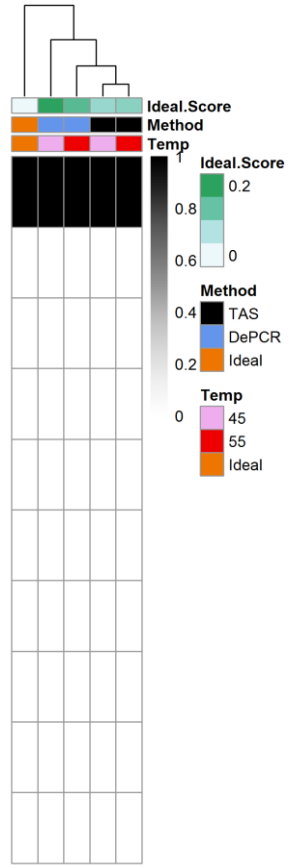


ANOSIM

DePCR vs TAS:
R=0.108, P=0.013

TAS45 vs TAS55:
R=-0.048; p=0.71

DePCR45 vs DePCR55:
R=-0.01, P=0.469

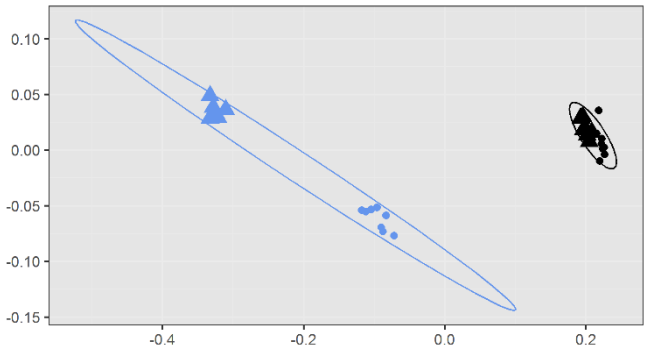


	Comparison	Average Ideal Score	ANOVA
1	TAS	0.10	0.02
	DePCR	0.18	
2	TAS45	0.09	0.71
	TAS55	0.11	
3	DePCR45	0.21	0.37
	DePCR55	0.16	

Figure S15

Primer Utilizing Profiles

Experiment: A6
Primers: 64
Templates: 1

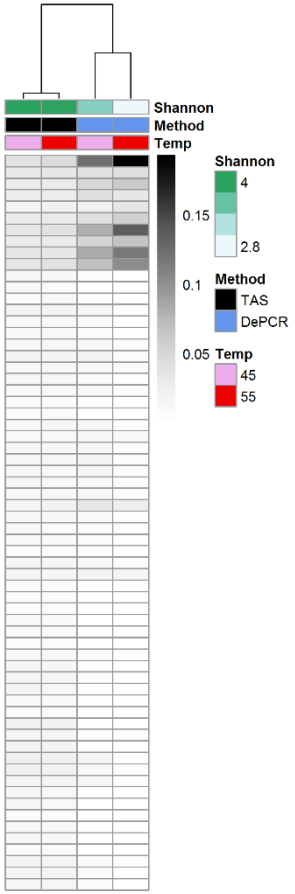


ANOSIM

DePCR vs TAS:
R=1, P=0.0001

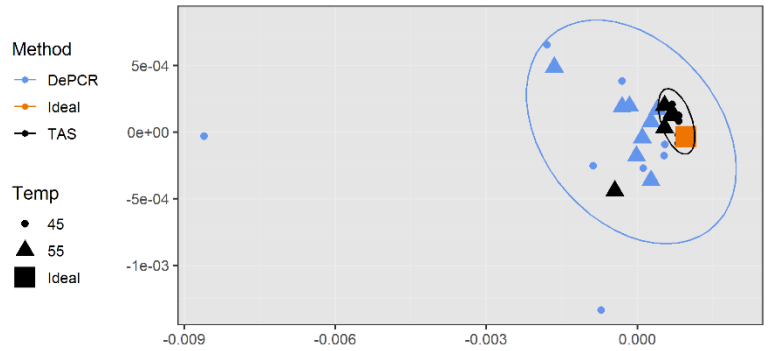
TAS45 vs TAS55:
R=1; p=0.0002

DePCR45 vs DePCR55:
R=1, P=0.0002



Comparison	DePCR45	DePCR55	ANOVA
% Reads 0 MM	12.16	19.31	0.00
% Reads 1 MM	47.68	65.16	0.00
% Reads 2 MM	29.86	12.81	0.00
% Reads 3MM	10.31	2.72	0.00
% Reads with 3' MM	27.41	21.24	0.00
% Reads with Mid' MM	34.60	37.10	0.00
% Reads with 5' MM	37.99	41.66	0.00
Average Primer Tm	58.18	57.99	0.00
Average Shannon	3.43	2.76	0.00

Template Profiles

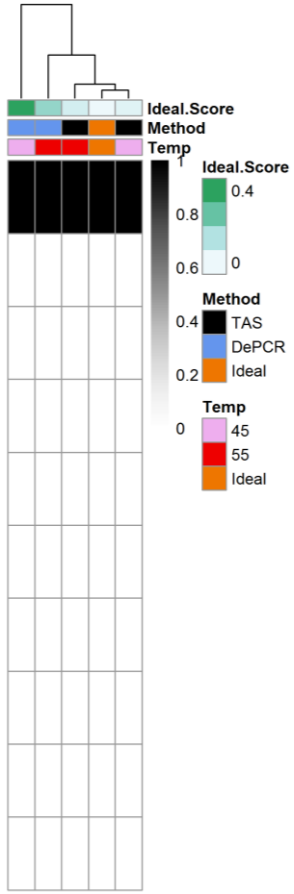


ANOSIM

DePCR vs TAS:
R=0.168, P=0.0002

TAS45 vs TAS55:
R=0.074; p=0.109

DePCR45 vs DePCR55:
R=-0.007, P=0.469

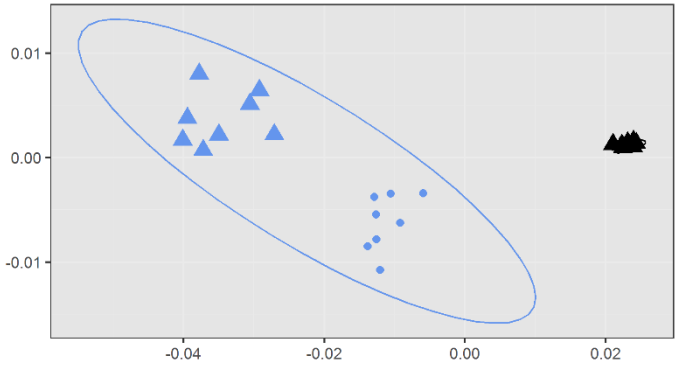


	Comparison	Average Ideal Score	ANOVA
1	TAS	0.05	0.01
	DePCR	0.36	
2	TAS45	0.04	0.31
	TAS55	0.07	
3	DePCR45	0.49	0.25
	DePCR55	0.22	

Figure S16

Primer Utilizing Profiles

Experiment: B1
Primers: 1
Templates: 10

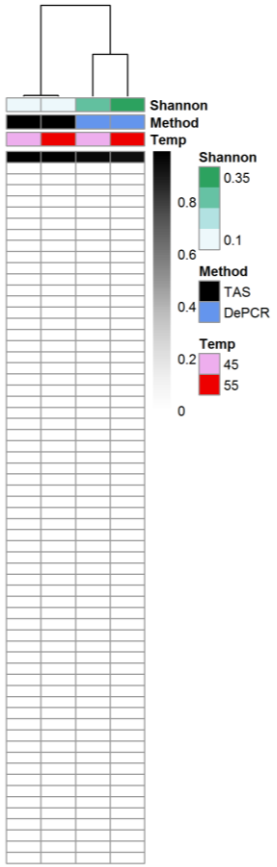


ANOSIM

DePCR vs TAS:
R=0.987, P=0.0001

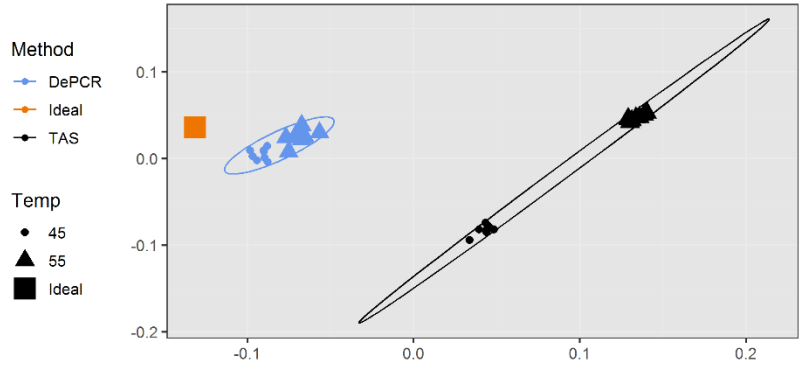
TAS45 vs TAS55:
R=-0.108; p=0.934

DePCR45 vs DePCR55:
R=1, P=0.0002



Comparison	DePCR45	DePCR55	ANOVA
% Reads 0 MM	14.27	16.77	0.00
% Reads 1 MM	84.21	81.72	0.00
% Reads 2 MM	1.50	1.48	0.79
% Reads 3MM	0.02	0.03	0.36
% Reads with 3' MM	22.00	19.63	0.00
% Reads with Mid' MM	36.82	39.84	0.00
% Reads with 5' MM	41.18	40.52	0.02
Average Primer Tm	57.42	57.43	0.01
Average Shannon	0.27	0.36	0.00

Template Profiles

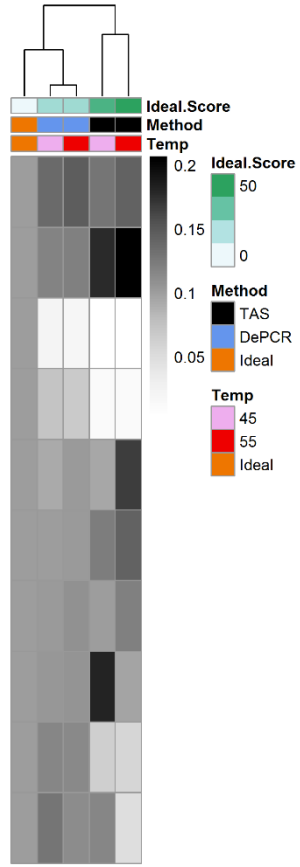


ANOSIM

DePCR vs TAS:
R=0.878, P=0.0001

TAS45 vs TAS55:
R=1; p=0.0002

DePCR45 vs DePCR55:
R=0.909, P=0.0002

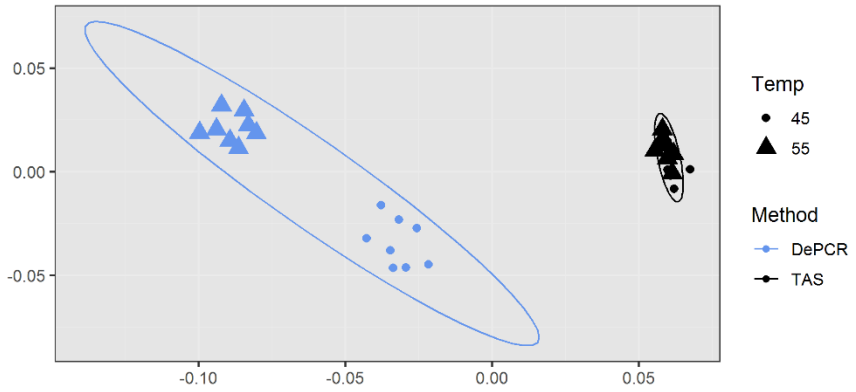


	Comparison	Average Ideal Score	ANOVA
1	TAS	50.80	0.00
	DePCR	22.82	
2	TAS45	45.63	0.00
	TAS55	55.97	
3	DePCR45	22.42	0.12
	DePCR55	23.23	

Figure S17

Primer Utilizing Profiles

Experiment: B2
Primers: 10
Templates: 10

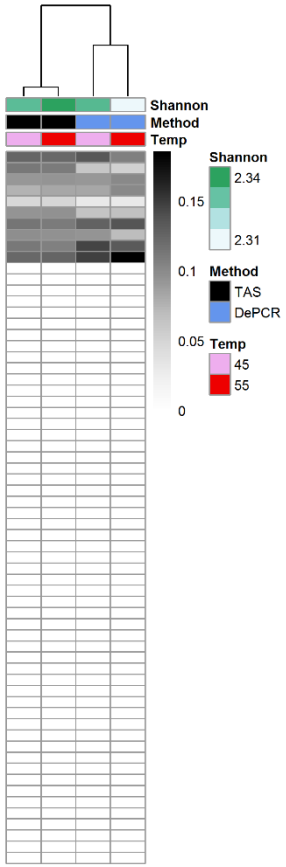


ANOSIM

DePCR vs TAS:
R=0.986, P=0.0001

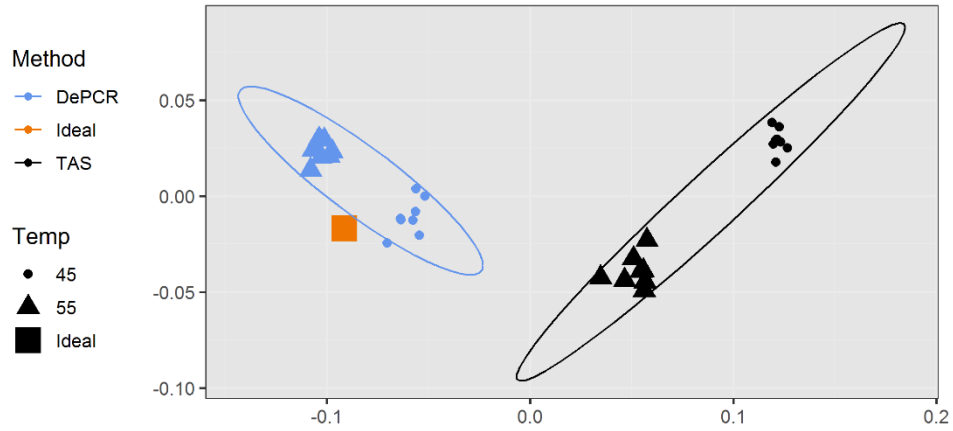
TAS45 vs TAS55:
R=0.422; p=0.0002

DePCR45 vs DePCR55:
R=1, P=0.0002



Comparison	DePCR45	DePCR55	ANOVA
% Reads 0 MM	28.98	48.05	0.00
% Reads 1 MM	46.54	43.24	0.00
% Reads 2 MM	24.26	8.63	0.00
% Reads 3MM	0.22	0.08	0.00
% Reads with 3' MM	23.28	25.47	0.00
% Reads with Mid' MM	34.68	35.29	0.10
% Reads with 5' MM	42.04	39.24	0.00
Average Primer Tm	57.78	57.87	0.00
Average Shannon	2.33	2.31	0.00

Template Profiles

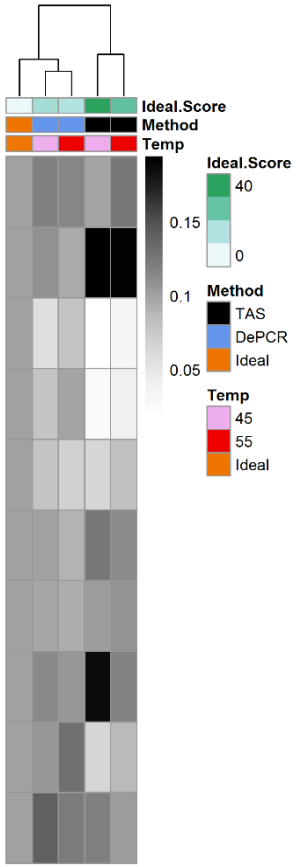


ANOSIM

DePCR vs TAS:
R=0.993, P=0.0001

TAS45 vs TAS55:
R=1; p=0.0002

DePCR45 vs DePCR55:
R=1, P=0.0002

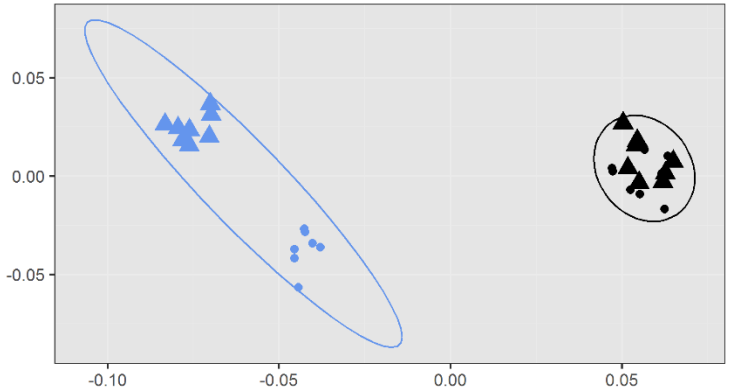


	Comparison	Average Ideal Score	ANOVA
1	TAS	38.60	0.00
	DePCR	16.55	
2	TAS45	45.48	0.00
	TAS55	31.73	
3	DePCR45	18.00	0.00
	DePCR55	15.10	

Figure S18

Primer Utilizing Profiles

Experiment: B3
Primers: 9
Templates: 10

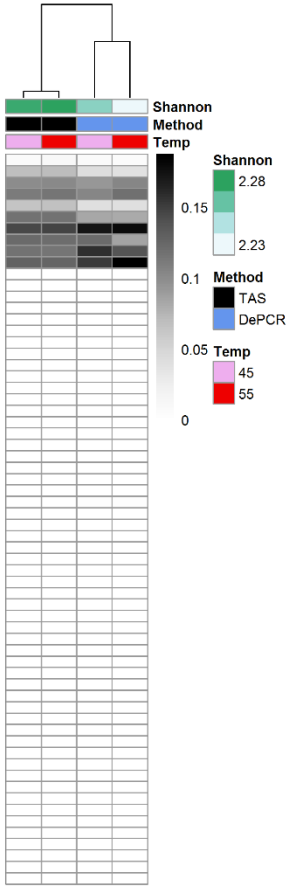


ANOSIM

DePCR vs TAS:
R=1, P=0.0001

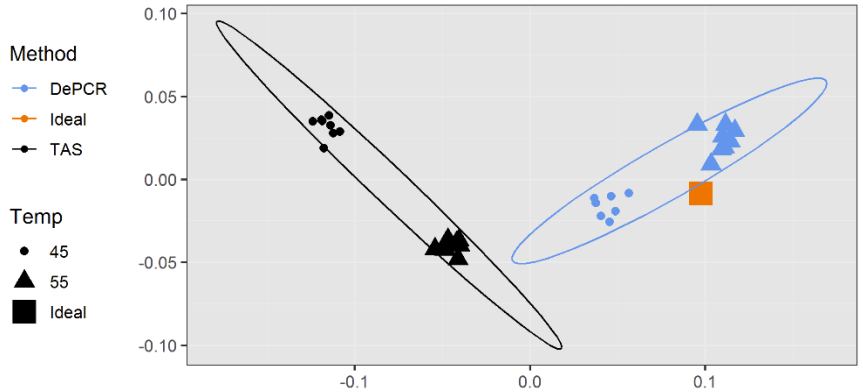
TAS45 vs TAS55:
R=-0.051; p=0.671

DePCR45 vs DePCR55:
R=1, P=0.0002



Comparison	DePCR45	DePCR55	ANOVA
% Reads 0 MM	29.10	49.31	0.00
% Reads 1 MM	42.17	40.86	0.00
% Reads 2 MM	28.46	9.73	0.00
% Reads 3MM	0.27	0.10	0.00
% Reads with 3' MM	23.19	24.82	0.00
% Reads with Mid' MM	35.92	36.26	0.40
% Reads with 5' MM	40.89	38.91	0.00
Average Primer Tm	57.77	57.91	0.00
Average Shannon	2.26	2.23	0.00

Template Profiles

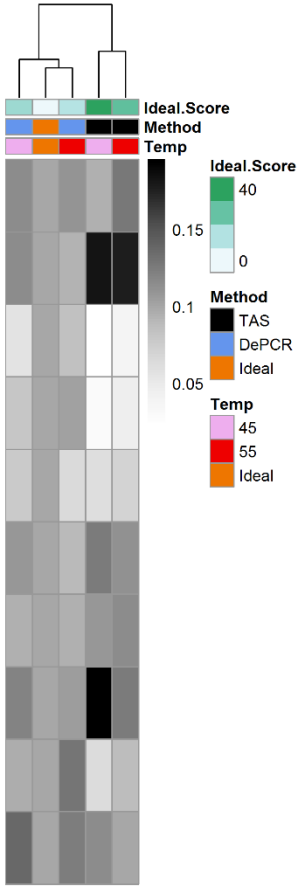


ANOSIM

DePCR vs TAS:
R=0.916, P=0.0001

TAS45 vs TAS55:
R=1; p=0.0002

DePCR45 vs DePCR55:
R=1, P=0.0002

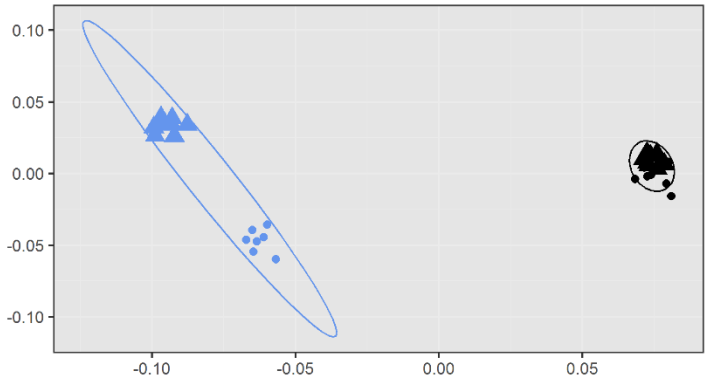


	Comparison	Average Ideal Score	ANOVA
1	TAS	38.38	0.00
	DePCR	16.69	
2	TAS45	45.24	0.00
	TAS55	31.52	
3	DePCR45	19.30	0.00
	DePCR55	14.41	

Figure S19

Primer Utilizing Profiles

Experiment: B4
Primers: 27
Templates: 10

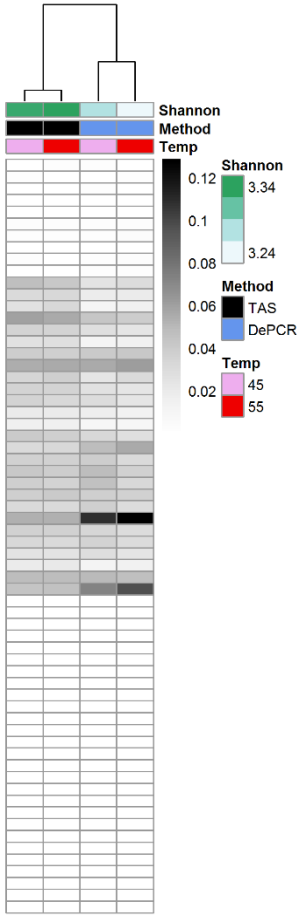


ANOSIM

DePCR vs TAS:
R=1, P=0.0001

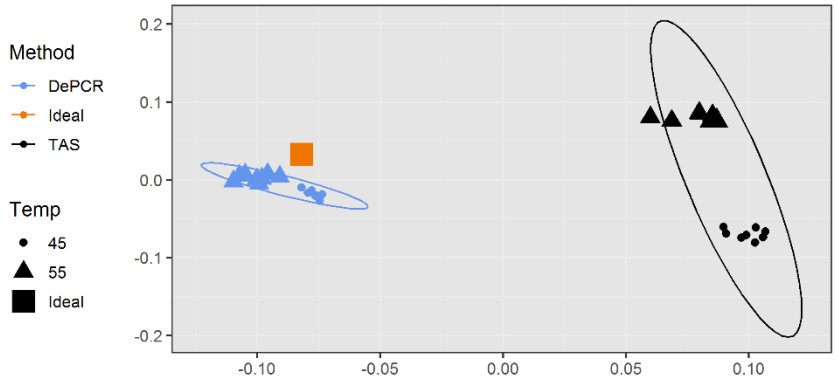
TAS45 vs TAS55:
R=0.47; p=0.0002

DePCR45 vs DePCR55:
R=1, P=0.0002



Comparison	DePCR45	DePCR55	ANOVA
% Reads 0 MM	0.56	0.76	0.00
% Reads 1 MM	49.53	74.13	0.00
% Reads 2 MM	45.13	22.37	0.00
% Reads 3MM	4.77	2.75	0.00
% Reads with 3' MM	26.39	22.68	0.00
% Reads with Mid' MM	36.84	40.90	0.00
% Reads with 5' MM	36.77	36.42	0.07
Average Primer Tm	58.59	58.76	0.00
Average Shannon	3.27	3.22	0.00

Template Profiles

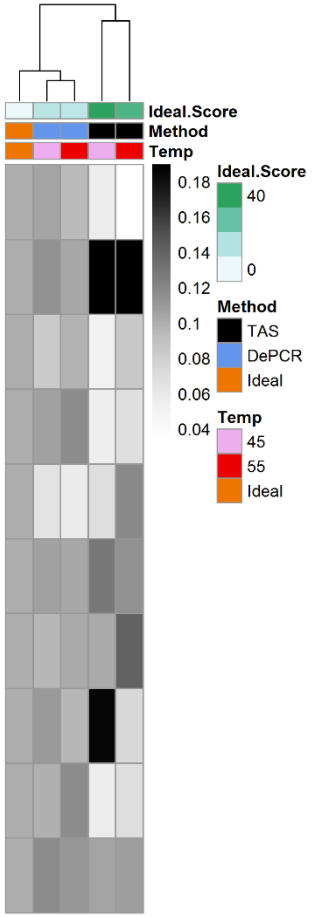


ANOSIM

DePCR vs TAS:
R=1, P=0.0001

TAS45 vs TAS55:
R=1; p=0.0002

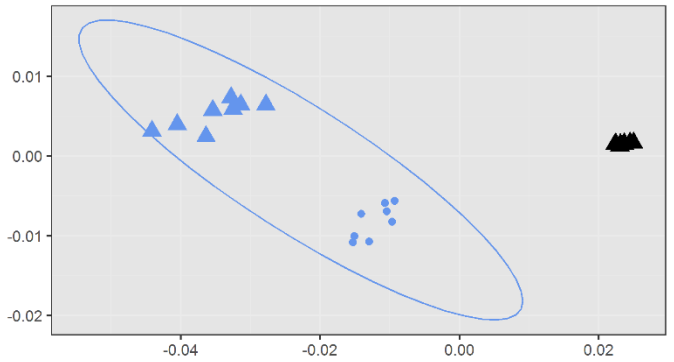
DePCR45 vs DePCR55:
R=1, P=0.0002



	Comparison	Average Ideal Score	ANOVA
1	TAS	38.13	0.00
	DePCR	11.53	
2	TAS45	42.14	0.00
	TAS55	34.13	
3	DePCR45	11.82	0.41
	DePCR55	11.28	

Figure S20

Primer Utilizing Profiles

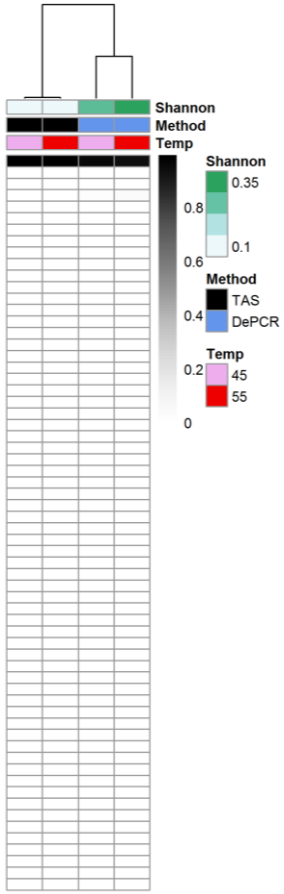


ANOSIM

DePCR vs TAS:
R=0.992, P=0.0001

TAS45 vs TAS55:
R=0.104; p=0.117

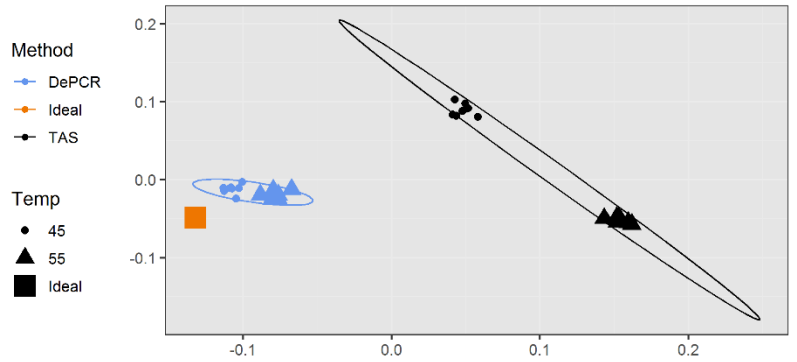
DePCR45 vs DePCR55:
R=1, P=0.0002



Comparison	DePCR45	DePCR55	ANOVA
% Reads 0 MM	3.42	5.77	0.00
% Reads 1 MM	94.72	92.70	0.00
% Reads 2 MM	1.83	1.51	0.00
% Reads 3MM	0.02	0.02	0.35
% Reads with 3' MM	22.75	20.24	0.00
% Reads with Mid' MM	36.06	39.65	0.00
% Reads with 5' MM	41.19	40.11	0.02
Average Primer Tm	57.42	57.43	0.16
Average Shannon	0.28	0.36	0.00

Experiment: C1
Primers: 1
Templates: 10

Template Profiles

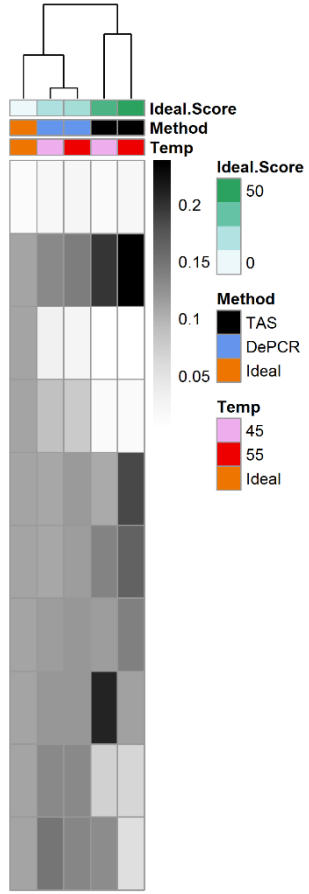


ANOSIM

DePCR vs TAS:
R=0.875, P=0.0001

TAS45 vs TAS55:
R=1; p=0.0002

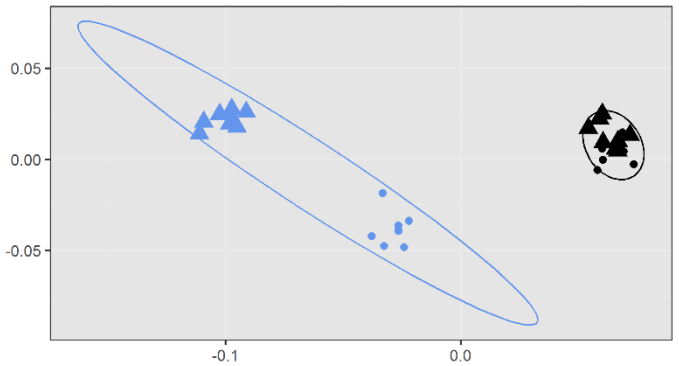
DePCR45 vs DePCR55:
R=0.993, P=0.0002



	Comparison	Average Ideal Score	ANOVA
1	TAS	54.32	0.00
	DePCR	22.45	
2	TAS45	48.72	0.00
	TAS55	59.91	
3	DePCR45	21.04	0.00
	DePCR55	23.87	

Figure S21

Primer Utilizing Profiles

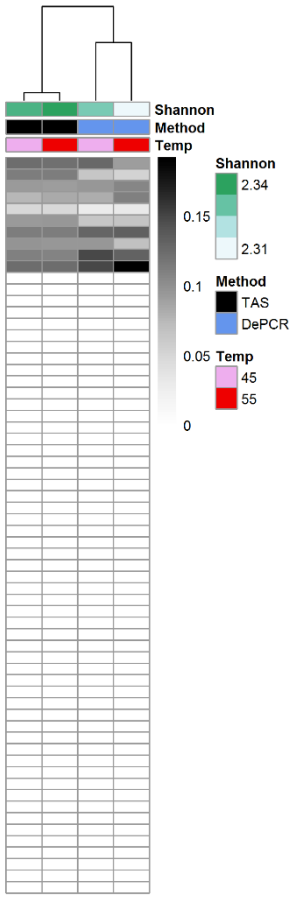


ANOSIM

DePCR vs TAS:
R=0.956, P=0.0001

TAS45 vs TAS55:
R=0.335; p=0.002

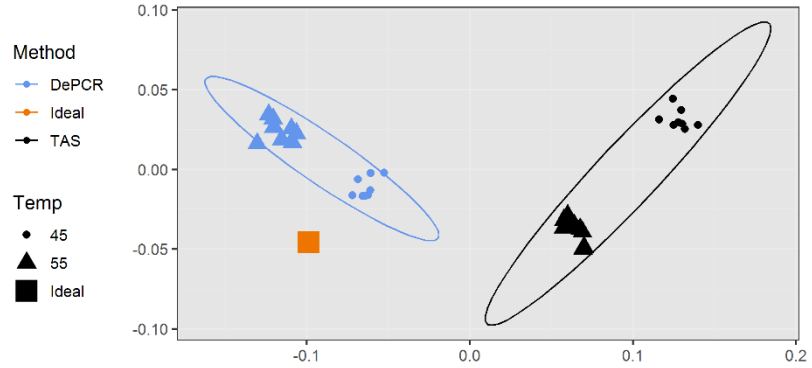
DePCR45 vs DePCR55:
R=1, P=0.0002



Comparison	DePCR45	DePCR55	ANOVA
% Reads 0 MM	29.58	51.32	0.00
% Reads 1 MM	42.98	38.78	0.00
% Reads 2 MM	27.22	9.83	0.00
% Reads 3MM	0.23	0.07	0.00
% Reads with 3' MM	23.43	26.64	0.00
% Reads with Mid' MM	34.80	34.94	0.65
% Reads with 5' MM	41.77	38.42	0.00
Average Primer Tm	57.79	57.89	0.00
Average Shannon	2.33	2.30	0.00

Experiment: C2
Primers: 10
Templates: 10

Template Profiles

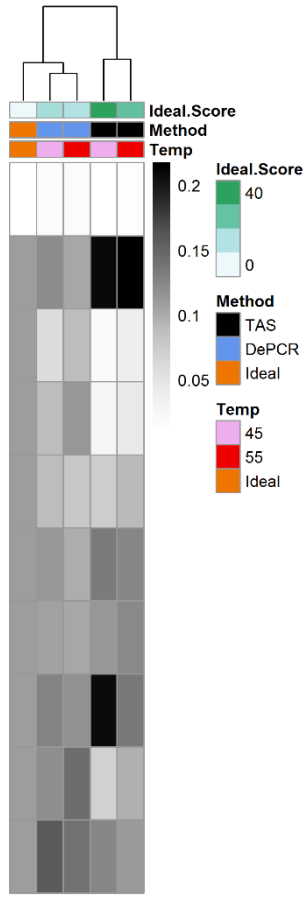


ANOSIM

DePCR vs TAS:
R=1, P=0.0001

TAS45 vs TAS55:
R=1; p=0.0002

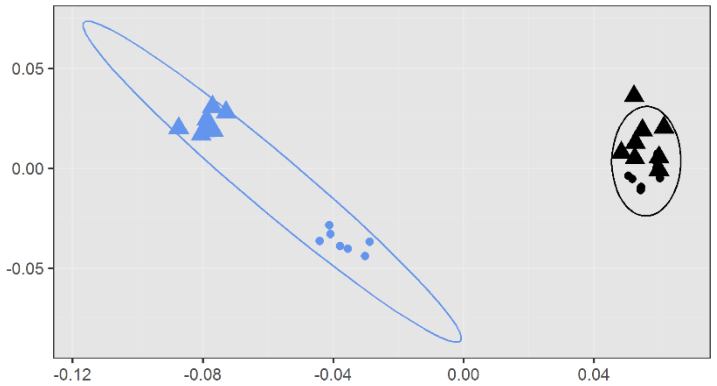
DePCR45 vs DePCR55:
R=1, P=0.0002



	Comparison	Average Ideal Score	ANOVA
1	TAS	41.48	0.00
	DePCR	17.56	
2	TAS45	49.17	0.00
	TAS55	33.79	
3	DePCR45	19.07	0.00
	DePCR55	16.04	

Figure S22

Primer Utilizing Profiles

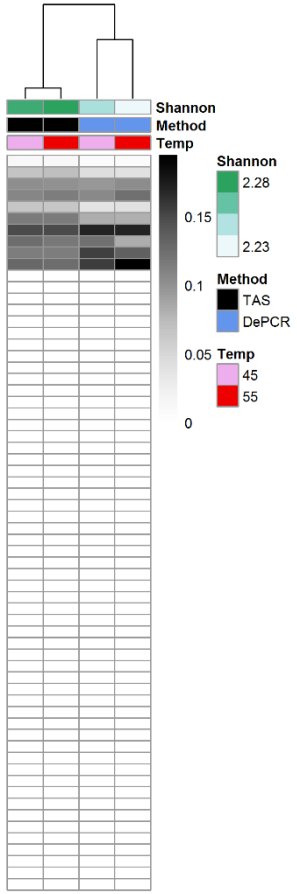


ANOSIM

DePCR vs TAS:
R=1, P=0.0001

TAS45 vs TAS55:
R=0.328; p=0.002

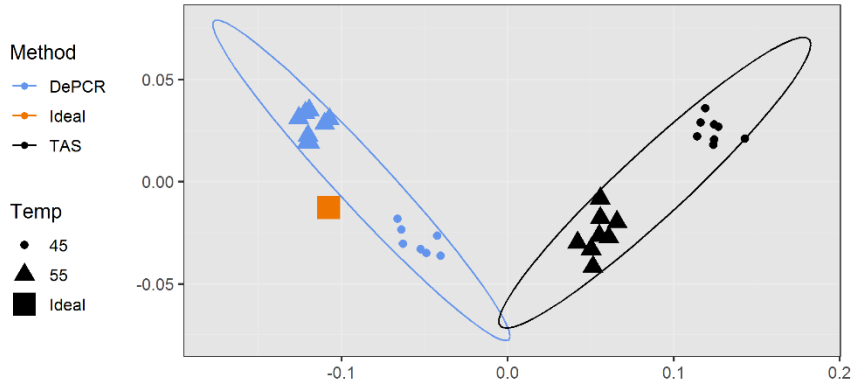
DePCR45 vs DePCR55:
R=1, P=0.0002



Comparison	DePCR45	DePCR55	ANOVA
% Reads 0 MM	31.68	55.21	0.00
% Reads 1 MM	35.80	34.32	0.00
% Reads 2 MM	32.24	10.39	0.00
% Reads 3MM	0.28	0.09	0.00
% Reads with 3' MM	23.20	25.56	0.00
% Reads with Mid' MM	35.69	36.68	0.03
% Reads with 5' MM	41.11	37.77	0.00
Average Primer Tm	57.75	57.93	0.00
Average Shannon	2.25	2.22	0.00

Experiment: C3
Primers: 9
Templates: 10

Template Profiles

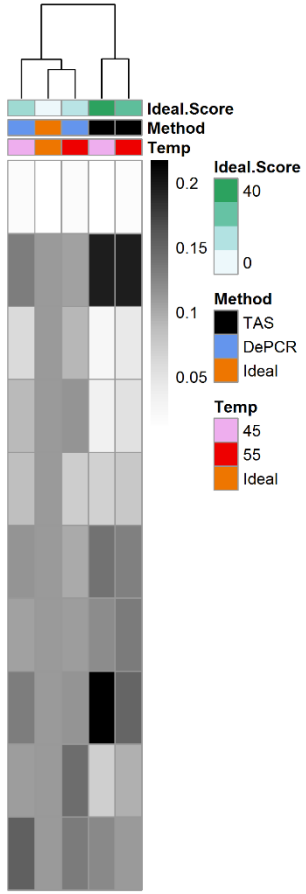


ANOSIM

DePCR vs TAS:
R=0.991, P=0.0001

TAS45 vs TAS55:
R=1; p=0.0002

DePCR45 vs DePCR55:
R=1, P=0.0002

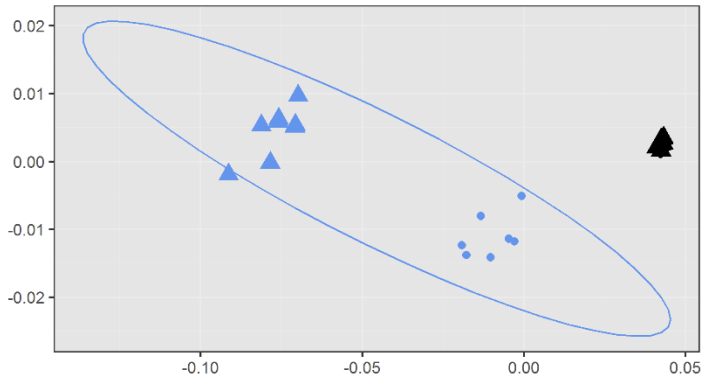


	Comparison	Average Ideal Score	ANOVA
1	TAS	41.53	0.00
	DePCR	17.05	
2	TAS45	48.81	0.00
	TAS55	34.24	
3	DePCR45	20.19	0.00
	DePCR55	14.31	

Figure S23

Primer Utilizing Profiles

Experiment: D1
Primers: 1
Templates: 4

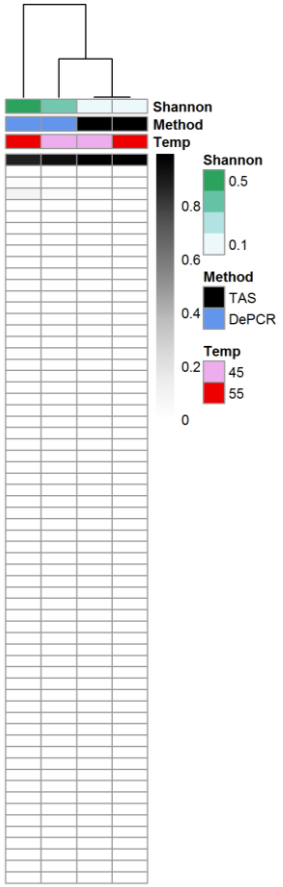


ANOSIM

DePCR vs TAS:
R=0.791, P=0.0001

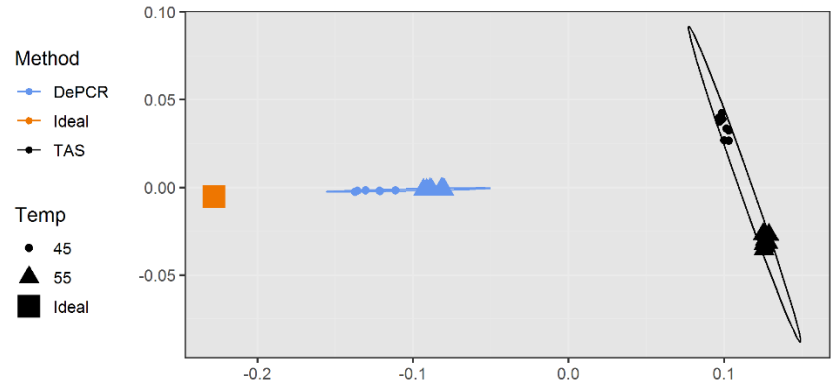
TAS45 vs TAS55:
R=0.009; p=0.399

DePCR45 vs DePCR55:
R=1, P=0.002



Comparison	DePCR45	DePCR55	ANOVA
% Reads 0 MM	13.18	19.78	0.00
% Reads 1 MM	85.31	78.86	0.00
% Reads 2 MM	1.47	1.34	0.08
% Reads 3MM	0.03	0.02	0.13
% Reads with 3' MM	97.79	97.65	0.30
% Reads with Mid' MM	0.94	0.98	0.66
% Reads with 5' MM	1.27	1.37	0.21
Average Primer Tm	57.43	57.43	0.00
Average Shannon	0.34	0.53	0.00

Template Profiles

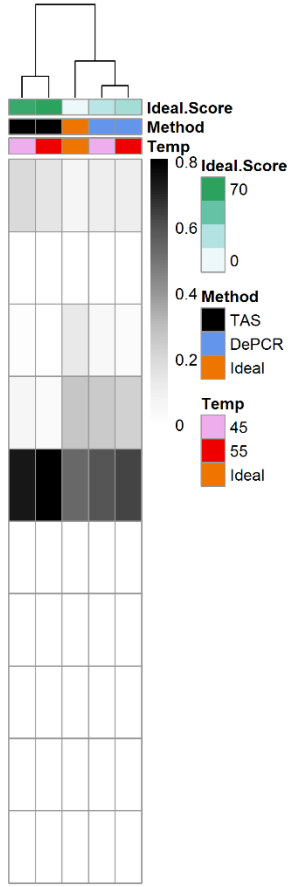


ANOSIM

DePCR vs TAS:
R=1, P=0.0001

TAS45 vs TAS55:
R=1; p=0.0002

DePCR45 vs DePCR55:
R=0.996, P=0.0002

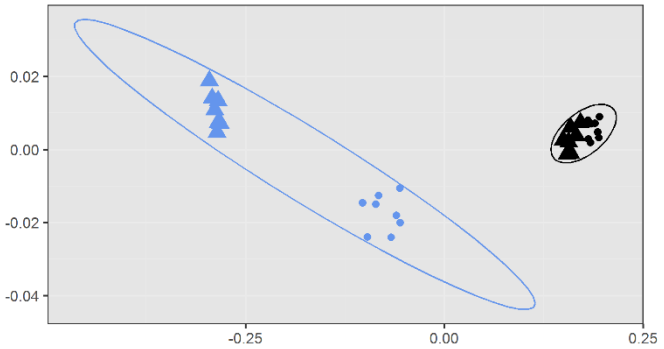


	Comparison	Average Ideal Score	ANOVA
1	TAS	68.46	0.00
	DePCR	24.27	
2	TAS45	65.95	0.00
	TAS55	70.96	
3	DePCR45	20.15	0.00
	DePCR55	27.87	

Figure S24

Primer Utilizing Profiles

Experiment: D2
Primers: 10
Templates: 4

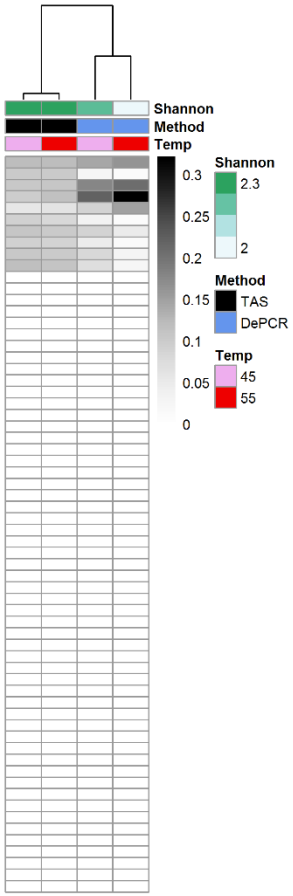


ANOSIM

DePCR vs TAS:
R=0.97, P=0.0001

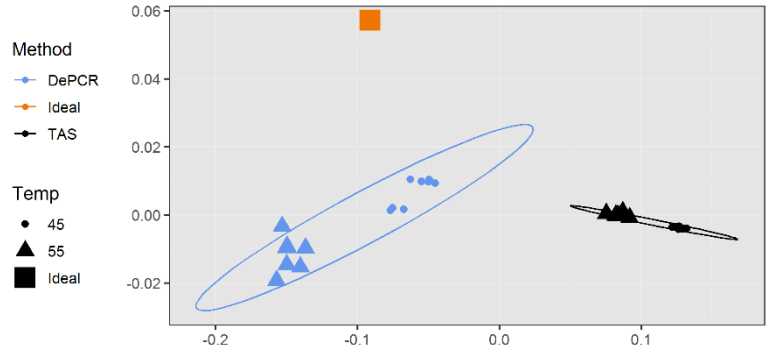
TAS45 vs TAS55:
R=0.866; p=0.0002

DePCR45 vs DePCR55:
R=1, P=0.0002



Comparison	DePCR45	DePCR55	ANOVA
% Reads 0 MM	26.59	43.80	0.00
% Reads 1 MM	43.50	46.84	0.00
% Reads 2 MM	29.69	9.28	0.00
% Reads 3MM	0.22	0.08	0.00
% Reads with 3' MM	64.27	75.44	0.00
% Reads with Mid' MM	15.38	12.56	0.00
% Reads with 5' MM	20.35	12.00	0.00
Average Primer Tm	57.98	58.11	0.00
Average Shannon	2.23	1.90	0.00

Template Profiles

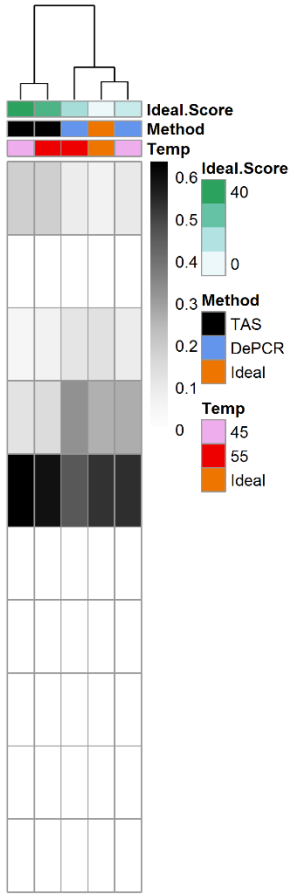


ANOSIM

DePCR vs TAS:
R=1, P=0.0001

TAS45 vs TAS55:
R=1; p=0.0002

DePCR45 vs DePCR55:
R=1 P=0.0002

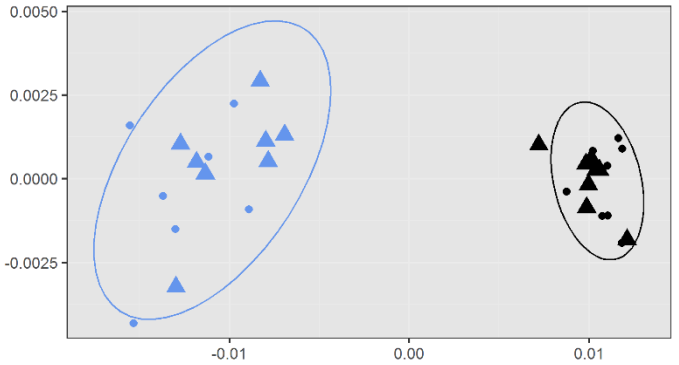


	Comparison	Average Ideal Score	ANOVA
1	TAS	41.16	0.00
	DePCR	13.23	
2	TAS45	45.17	0.00
	TAS55	36.58	
3	DePCR45	9.53	0.00
	DePCR55	17.45	

Figure S25

Primer Utilizing Profiles

Experiment: E1
Primers: 1
Templates: 4

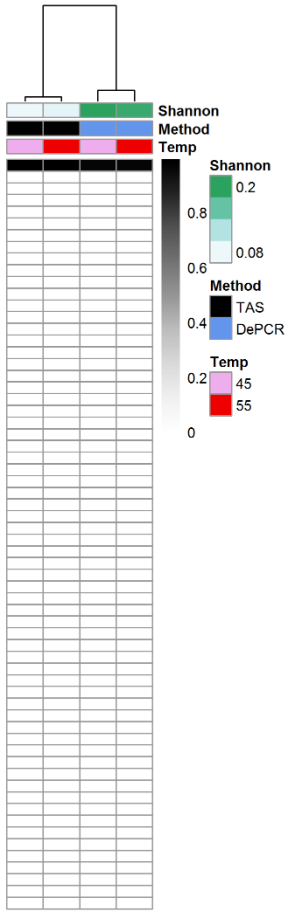


ANOSIM

DePCR vs TAS:
R=1, P=0.0001

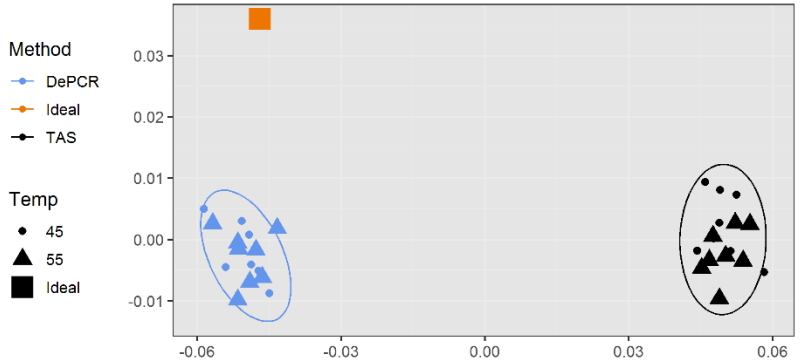
TAS45 vs TAS55:
R=0.126; p=0.054

DePCR45 vs DePCR55:
R=0.078, P=0.174



Comparison	DePCR45	DePCR55	ANOVA (p-value)
% Reads 0 MM	9.30	9.15	0.48
% Reads 1 MM	89.14	89.39	0.29
% Reads 2 MM	1.54	1.45	0.31
% Reads 3MM	0.02	0.01	0.23
% Reads with 3' MM	0.81	0.77	0.34
% Reads with Mid' MM	97.93	98.03	0.37
% Reads with 5' MM	1.26	1.20	0.54
Average Primer Tm	57.42	57.42	0.76
Average Shannon	0.20	0.19	0.10

Template Profiles

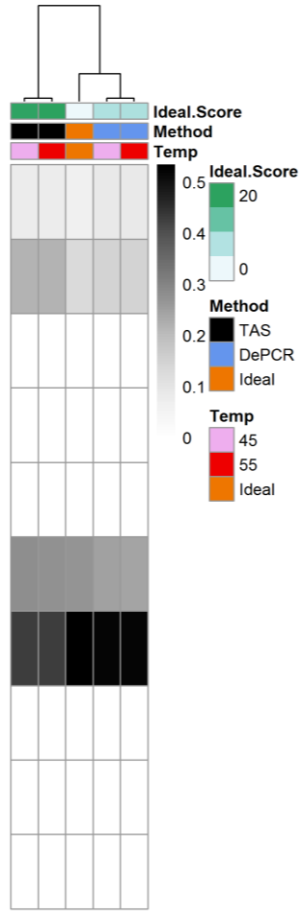


ANOSIM

DePCR vs TAS:
R=1, P=0.0001

TAS45 vs TAS55:
R=0.1; p=0.116

DePCR45 vs DePCR55:
R=-0.103, P=0.963

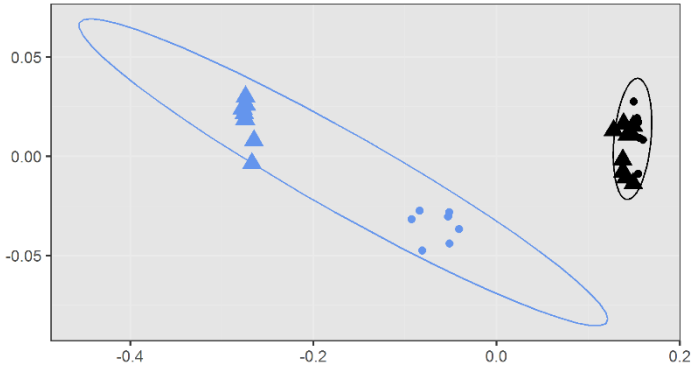


	Comparison	Average Ideal Score	ANOVA
1	TAS	20.53	0.00
	DePCR	7.30	
2	TAS45	20.28	0.26
	TAS55	20.78	
3	DePCR45	7.11	0.56
	DePCR55	7.47	

Figure S26

Primer Utilizing Profiles

Experiment: E2
Primers: 10
Templates: 4

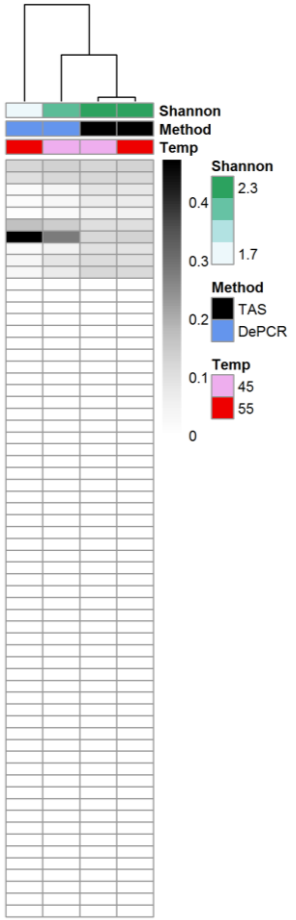


ANOSIM

DePCR vs TAS:
R=0.893, P=0.0001

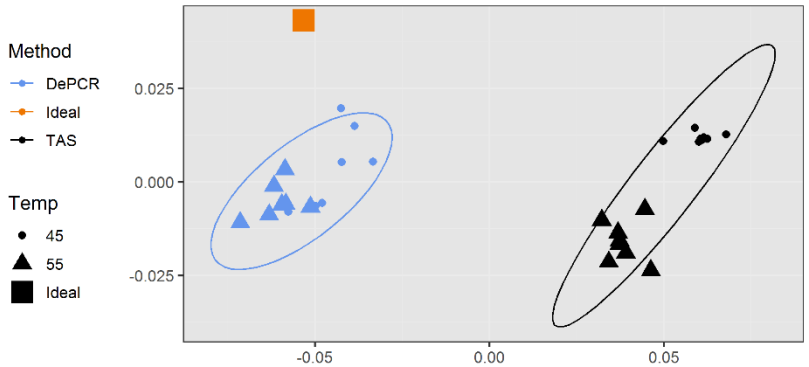
TAS45 vs TAS55:
R=0.743; p=0.0002

DePCR45 vs DePCR55:
R=1, P=0.0006



Comparison	DePCR45	DePCR55	ANOVA
% Reads 0 MM	30.88	51.68	0.00
% Reads 1 MM	41.43	38.94	0.00
% Reads 2 MM	27.43	9.31	0.00
% Reads 3MM	0.25	0.07	0.00
% Reads with 3' MM	10.65	7.69	0.00
% Reads with Mid' MM	66.09	76.16	0.00
% Reads with 5' MM	23.26	16.15	0.00
Average Primer Tm	58.44	58.93	0.00
Average Shannon	2.16	1.70	0.00

Template Profiles

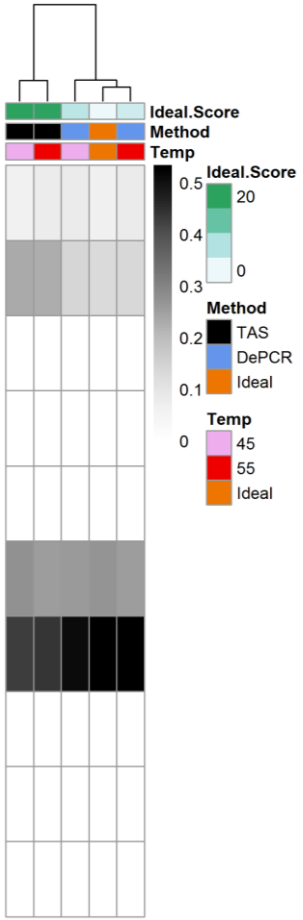


ANOSIM

DePCR vs TAS:
R=1, P=0.0001

TAS45 vs TAS55:
R=0.984; p=0.0002

DePCR45 vs DePCR55:
R=0.633, P=0.0001



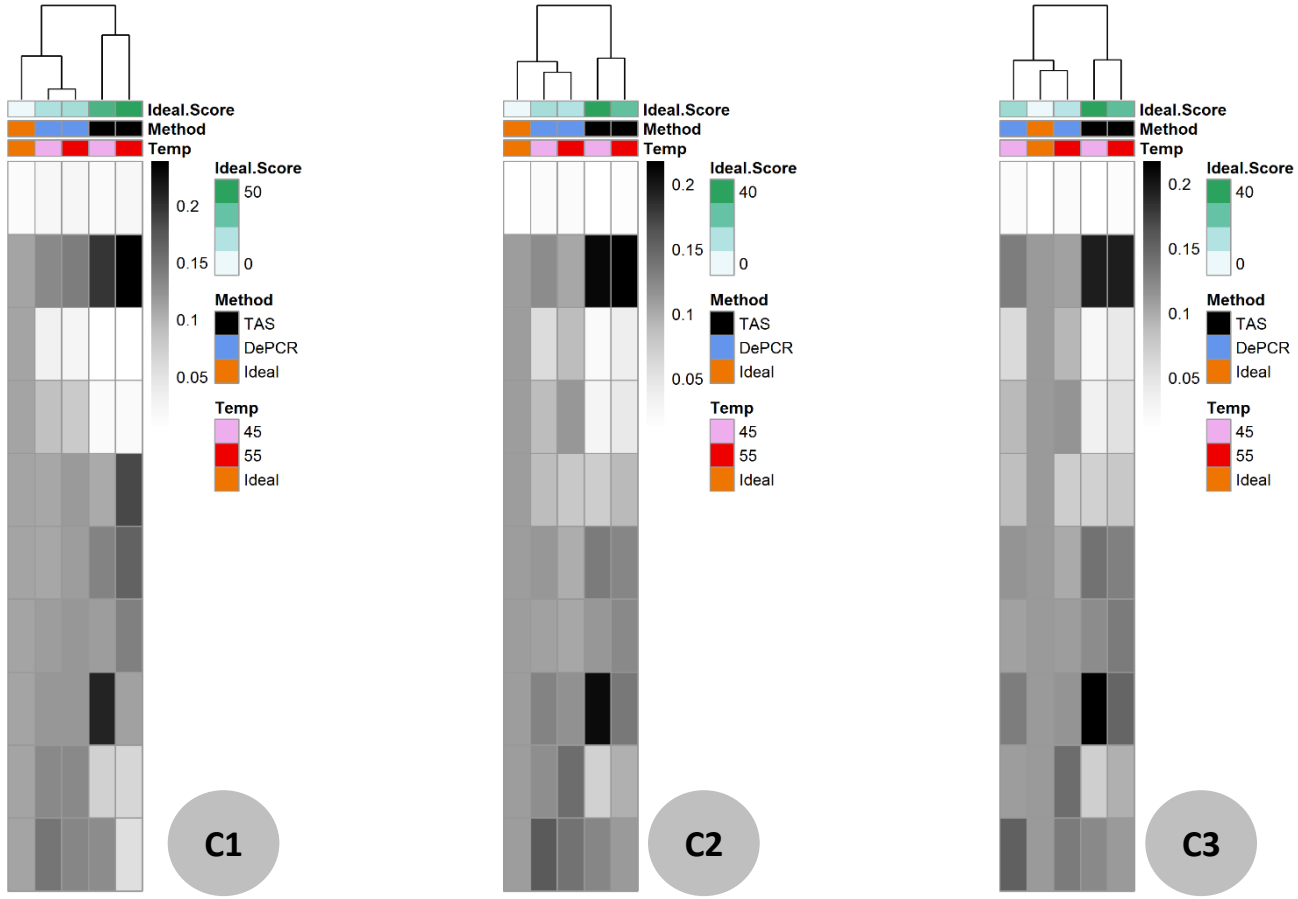
	Comparison	Average Ideal Score	ANOVA
1	TAS	21.50	0.00
	DePCR	4.91	
2	TAS45	21.66	0.59
	TAS55	21.35	
3	DePCR45	5.95	0.00
	DePCR55	3.87	

Figures S11-S26. Template and primer utilization profiles for 16 individual experiments conducted in this study.

For each study, varying number of primers and templates were used, as described in **Table 10**. For mMDS plots, samples were color coded by amplification method and different annealing temperatures indicated by shape. Ellipses represent a 95% confidence interval around the centroid. ANOVA was performed to measure differences in measured values by annealing temperature. Intensity scales vary between experiments. All samples were rarefied to 7,000 sequences. Heatmaps are the average of 7-8 technical replicates per condition; all replicates are shown in mMDS plots. (A) For each experiment, primer utilization profiles (PUPs) were generated (left side), and data are presented as mMDS plots (top) and as clustered heatmaps (bottom). Analysis of similarity (ANOSIM) was performed to determine if PUPs were significantly different between TAS and DePCR, regardless of annealing temperature, and within method across annealing temperature. Each slide contains a table showing the percentage of reads with 0, 1, 2 and 3 mismatches between primers and templates, as indicated in experiments with DePCR amplifications. For primer-template interactions with only a single mismatch, percentage of reads with 3' (-2), middle (-8) and 5' (-14) mismatches are shown. The average theoretical melting temperature of primers used in each study are shown. (B) Template profiling analyses were performed (right side), and data are presented as mMDS plots (top) and as clustered heatmaps (bottom). In addition to analysis of sequence data, the expected distribution of reads is shown in orange, both in the mMDS plots and in the heatmap. ANOSIM was performed to determine if template profiles were significantly different between TAS and DePCR, regardless of annealing temperature, and within method across annealing temperature.

Ideal scores, as described in text, were calculated to determine which method and annealing temperature generated the closest approximation of the expected template distribution.

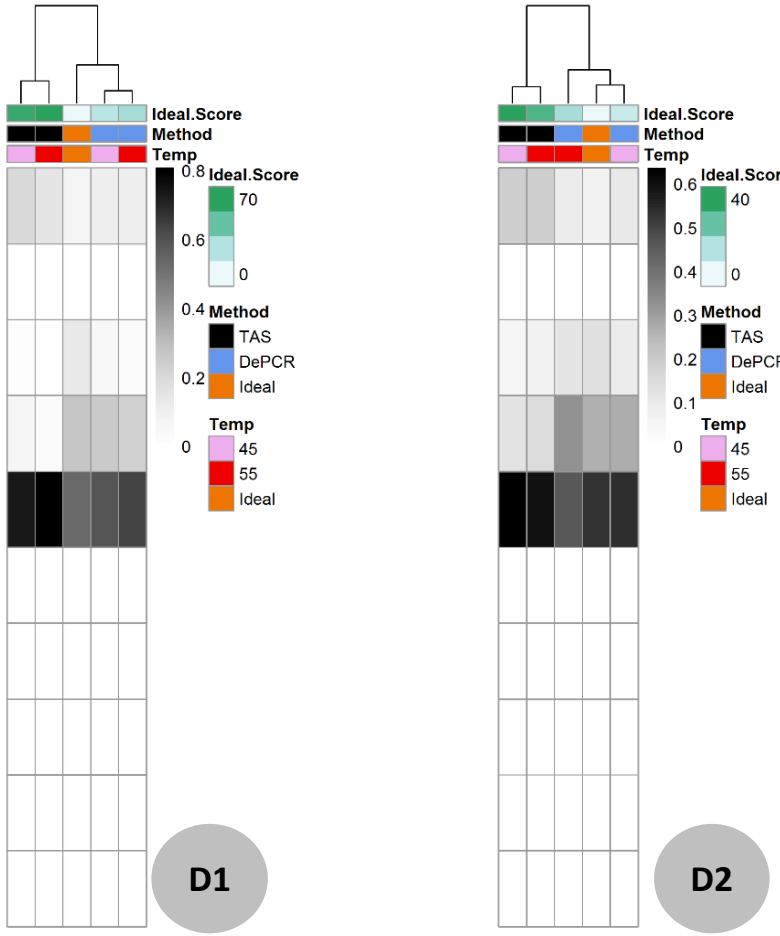
Figure S27

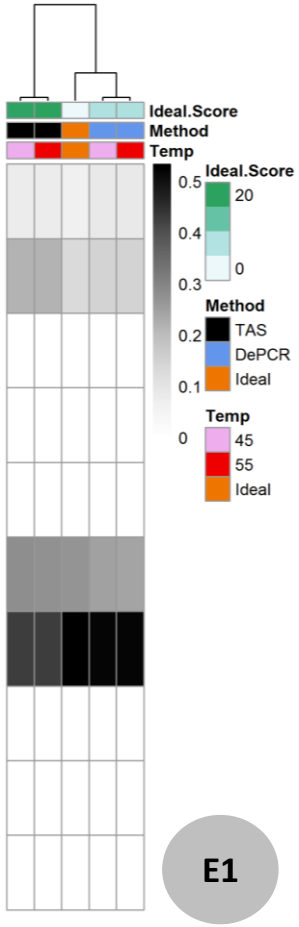


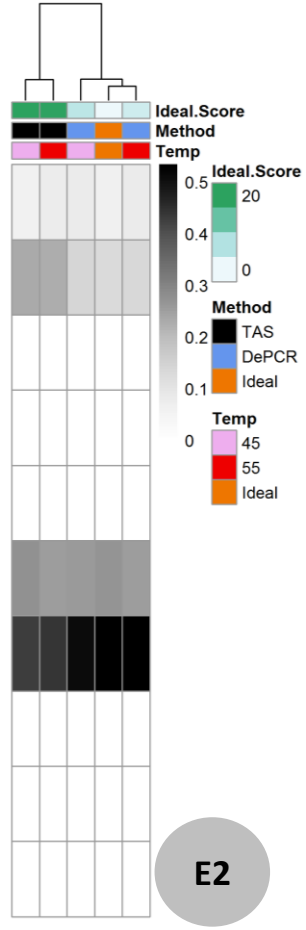
Average Ideal Score

	C1		C2		C3	
TAS	54.32	*	41.48	*	41.53	*
DePCR	22.45		17.56		17.05	
TAS45	48.72	*	49.17	*	48.81	*
TAS55	59.91		33.79		34.24	
DePCR45	21.04	*	19.07	*	20.19	*
DePCR55	23.87		16.04		14.31	

Figure S28







Average Ideal Score

TAS	68.46	*
DePCR	24.27	
TAS45	65.95	*
TAS55	70.96	
DePCR45	20.15	*
DePCR55	27.87	

TAS	41.16	*
DePCR	13.23	
TAS45	45.17	*
TAS55	36.58	
DePCR45	9.53	*
DePCR55	17.45	

Average Ideal Score

TAS	20.53	*
DePCR	7.30	
TAS45	20.28	
TAS55	20.78	
DePCR45	7.11	
DePCR55	7.47	

TAS	21.50	*
DePCR	4.91	
TAS45	21.66	
TAS55	21.35	
DePCR45	5.95	*
DePCR55	3.87	

Figures S27-28. Effect of PCR methodology and annealing temperature on template profiles in amplification reactions utilizing varying primer pools.

One-way clustered heatmaps of untransformed template utilization profiling during amplification of an uneven pooling of synthetic DNA templates and varying primer pools (**Figure S27** = C1, C2 and C3 experiments with all ten templates present, and template ST1 at $1/10^{th}$ the concentration of the other nine templates; **Figure S28** = D1, D2, E1 and E2 experiments with four templates). For experiments C1, D1 and E1, only a single primer variant was used

(806F_v1), while in experiments C2, D2 and E2, 10 primers were used. In experiment C3, 9 primers were used (806F_v1 was removed). Primer and template details are shown in **Table 10**. Samples (columns) are color-coded by amplification method (TAS or DePCR), amplification annealing temperature (45°C or 55°C), and average Ideal score. Each column represents the average of 7-8 technical replicates per condition and rarefaction to 7,000 sequences/replicate. Templates (rows) represent all 10 templates (ordered from top to bottom; ST1, ST4, ST6, ST7, ST8, ST11, ST15, ST23, ST39, and ST55). Ideal score comparisons between TAS and DePCR (across both annealing temperatures), within TAS (45°C or 55°C), and within DePCR (45°C or 55°C) are shown in tables. Asterisks indicate significant differences in measured values by annealing temperature (ANOVA, $P < 0.01$). Intensity scales vary between experiments.

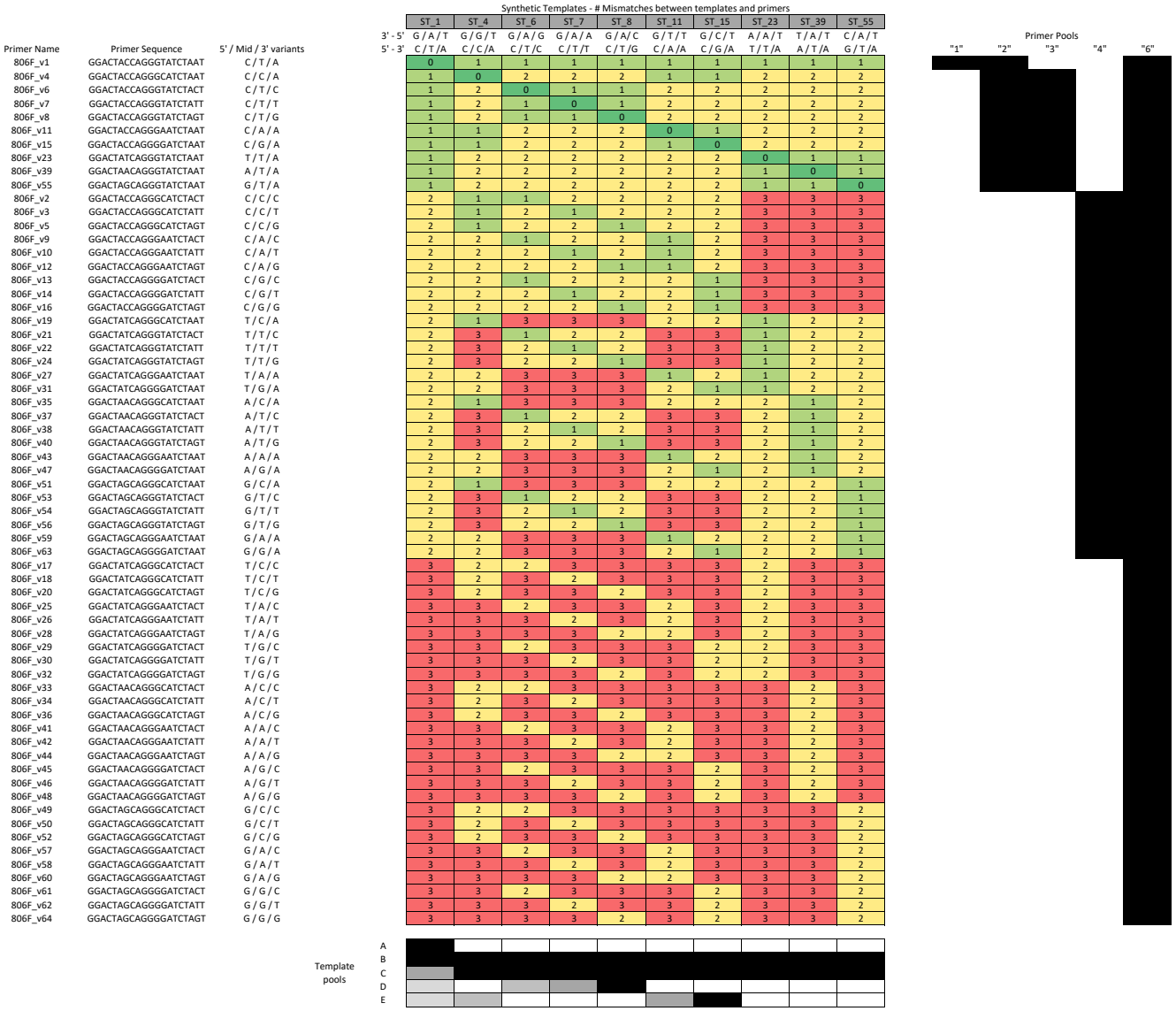
Table S10. Locus-specific primer sequences used in this study

Primer Name	"Reverse" Primer Sequence	Length	Tm (°C)		CS2 Linker Sequence	Final Primer Sequence Used	Length
555R	CGGAATTACTGGGCGTAAAGG	21	64.1		TACGGTAGCAGAGACTTGGTCT	TACGGTAGCAGAGACTTGGTCTCGGAATTACTGGGCGTAAAGG	43

Primer Name	"Forward" Primer Sequences	Length	Tm (°C)	MMs *	CS1 Linker Sequence	Final Primer Sequence Used	Length
806F_v1	GGACTACCAGGGTATCTAAT	20	57.4	0	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTACCAGGGTATCTAAT	42
806F_v2	GGACTACCAGGGCATCTACT	20	62.3	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTACCAGGGCATCTACT	42
806F_v3	GGACTACCAGGGCATCTATT	20	60.6	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTACCAGGGCATCTATT	42
806F_v4	GGACTACCAGGGCATCTAAT	20	60.6	1	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTACCAGGGCATCTAAT	42
806F_v5	GGACTACCAGGGCATCTAGT	20	62.3	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTACCAGGGCATCTAGT	42
806F_v6	GGACTACCAGGGTATCTACT	20	59.1	1	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTACCAGGGTATCTACT	42
806F_v7	GGACTACCAGGGTATCTATT	20	57.4	1	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTACCAGGGTATCTATT	42
806F_v8	GGACTACCAGGGTATCTAGT	20	59.1	1	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTACCAGGGTATCTAGT	42
806F_v9	GGACTACCAGGGAATCTACT	20	59.7	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTACCAGGGAATCTACT	42
806F_v10	GGACTACCAGGGAATCTATT	20	58	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTACCAGGGAATCTATT	42
806F_v11	GGACTACCAGGGAATCTAAT	20	58	1	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTACCAGGGAATCTAAT	42
806F_v12	GGACTACCAGGGAATCTAGT	20	59.7	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTACCAGGGAATCTAGT	42
806F_v13	GGACTACCAGGGGATCTACT	20	61.5	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTACCAGGGGATCTACT	42
806F_v14	GGACTACCAGGGGATCTATT	20	59.8	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTACCAGGGGATCTATT	42
806F_v15	GGACTACCAGGGGATCTAAT	20	59.8	1	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTACCAGGGGATCTAAT	42
806F_v16	GGACTACCAGGGGATCTAGT	20	61.5	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTACCAGGGGATCTAGT	42
806F_v17	GGACTATCAGGGCATCTACT	20	60	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTATCAGGGCATCTACT	42
806F_v18	GGACTATCAGGGCATCTATT	20	58.4	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTATCAGGGCATCTATT	42
806F_v19	GGACTATCAGGGCATCTAAT	20	58.4	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTATCAGGGCATCTAAT	42
806F_v20	GGACTATCAGGGCATCTAGT	20	60	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTATCAGGGCATCTAGT	42
806F_v21	GGACTATCAGGGTATCTACT	20	56.8	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTATCAGGGTATCTACT	42
806F_v22	GGACTATCAGGGTATCTATT	20	55.2	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTATCAGGGTATCTATT	42
806F_v23	GGACTATCAGGGTATCTAAT	20	55.2	1	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTATCAGGGTATCTAAT	42
806F_v24	GGACTATCAGGGTATCTAGT	20	56.8	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTATCAGGGTATCTAGT	42
806F_v25	GGACTATCAGGGAATCTACT	20	57.4	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTATCAGGGAATCTACT	42
806F_v26	GGACTATCAGGGAATCTATT	20	55.8	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTATCAGGGAATCTATT	42
806F_v27	GGACTATCAGGGAATCTAAT	20	55.8	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTATCAGGGAATCTAAT	42
806F_v28	GGACTATCAGGGAATCTAGT	20	57.4	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTATCAGGGAATCTAGT	42
806F_v29	GGACTATCAGGGGATCTACT	20	59.2	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTATCAGGGGATCTACT	42
806F_v30	GGACTATCAGGGGATCTATT	20	57.5	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTATCAGGGGATCTATT	42
806F_v31	GGACTATCAGGGGATCTAAT	20	57.5	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTATCAGGGGATCTAAT	42
806F_v32	GGACTATCAGGGGATCTAGT	20	59.2	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTATCAGGGGATCTAGT	42
806F_v33	GGACTAACAGGGCATCTACT	20	60.5	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAACAGGGCATCTACT	42
806F_v34	GGACTAACAGGGCATCTATT	20	58.9	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAACAGGGCATCTATT	42
806F_v35	GGACTAACAGGGCATCTAAT	20	58.9	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAACAGGGCATCTAAT	42
806F_v36	GGACTAACAGGGCATCTAGT	20	60.5	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAACAGGGCATCTAGT	42
806F_v37	GGACTAACAGGGTATCTACT	20	57.3	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAACAGGGTATCTACT	42
806F_v38	GGACTAACAGGGTATCTATT	20	55.7	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAACAGGGTATCTATT	42
806F_v39	GGACTAACAGGGTATCTAAT	20	55.7	1	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAACAGGGTATCTAAT	42
806F_v40	GGACTAACAGGGTATCTAGT	20	57.3	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAACAGGGTATCTAGT	42
806F_v41	GGACTAACAGGGAATCTACT	20	57.9	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAACAGGGAATCTACT	42
806F_v42	GGACTAACAGGGAATCTATT	20	56.2	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAACAGGGAATCTATT	42
806F_v43	GGACTAACAGGGAATCTAAT	20	56.2	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAACAGGGAATCTAAT	42
806F_v44	GGACTAACAGGGAATCTAGT	20	57.9	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAACAGGGAATCTAGT	42
806F_v45	GGACTAACAGGGGATCTACT	20	59.7	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAACAGGGGATCTACT	42
806F_v46	GGACTAACAGGGGATCTATT	20	58	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAACAGGGGATCTATT	42
806F_v47	GGACTAACAGGGGATCTAAT	20	58	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAACAGGGGATCTAAT	42
806F_v48	GGACTAACAGGGGATCTAGT	20	59.7	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAACAGGGGATCTAGT	42
806F_v49	GGACTAGCAGGGCATCTACT	20	62.6	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAGCAGGGCATCTACT	42
806F_v50	GGACTAGCAGGGCATCTATT	20	60.9	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAGCAGGGCATCTATT	42
806F_v51	GGACTAGCAGGGCATCTAAT	20	60.9	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAGCAGGGCATCTAAT	42
806F_v52	GGACTAGCAGGGCATCTAGT	20	62.6	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAGCAGGGCATCTAGT	42
806F_v53	GGACTAGCAGGGTATCTACT	20	59.4	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAGCAGGGTATCTACT	42
806F_v54	GGACTAGCAGGGTATCTATT	20	57.8	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAGCAGGGTATCTATT	42
806F_v55	GGACTAGCAGGGTATCTAAT	20	57.8	1	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAGCAGGGTATCTAAT	42
806F_v56	GGACTAGCAGGGTATCTAGT	20	59.4	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAGCAGGGTATCTAGT	42
806F_v57	GGACTAGCAGGGAATCTACT	20	60	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAGCAGGGAATCTACT	42
806F_v58	GGACTAGCAGGGAATCTATT	20	58.3	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAGCAGGGAATCTATT	42
806F_v59	GGACTAGCAGGGAATCTAAT	20	58.3	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAGCAGGGAATCTAAT	42
806F_v60	GGACTAGCAGGGAATCTAGT	20	60	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAGCAGGGAATCTAGT	42
806F_v61	GGACTAGCAGGGGATCTACT	20	61.7	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAGCAGGGGATCTACT	42
806F_v62	GGACTAGCAGGGGATCTATT	20	60.1	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAGCAGGGGATCTATT	42
806F_v63	GGACTAGCAGGGGATCTAAT	20	60.1	2	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAGCAGGGGATCTAAT	42
806F_v64	GGACTAGCAGGGGATCTAGT	20	61.7	3	ACACTGACGACATGGTTCTACA	ACACTGACGACATGGTTCTACAGGACTAGCAGGGGATCTAGT	42

* Mismatches relative to the first 806 primer, 806Syn_1

Table S11. Distribution of mismatches between primers and templates used in this study.



Locus-specific primer names and primer sequences (columns A and B) are shown next to variant position sequences (column C). Columns F-O represent each of the 10 synthetic DNA templates used in this study, with nucleotide sequences at each potential mismatch position shown in rows 3 and 4. Number of mismatches between templates and primers are colored in columns F-O and rows 5-68. Columns Q-U indicate which primers are used in which series of experiments (1-6). Rows 70-74 indicate which templates are used in which series of experiments (A-E).

Table S12. Metadata associated with all samples used in this study

#	Exp	SRA Sample Name	SampleID	PCR Method	Annealing Temp	Barcode Name	Barcode Sequence	Input DNA Template	gDNA Conc (ng/ul)	Raw Reads	Merged Reads	Trimmed	Shannon	Ideal Score	Rarefaction	Ref. Figure
1	A1	214644436_Ankur001-A1-45-1_S1	A1-45-1	**DePCR**	45	FLD0001	GTATCGTCGT	A	2.5	28286	27923	22817	0.28	0.09	7000	20, 21, S11
2	A1	214656448_Ankur002-A1-45-2_S2	A1-45-2	**DePCR**	45	FLD0002	GTGTATGCGT	A	2.5	27630	27230	22095	0.24	0.17	7000	20, 21, S11
3	A1	214657444_Ankur003-A1-45-3_S3	A1-45-3	**DePCR**	45	FLD0003	TGCTCGTAGT	A	2.5	21137	20852	16965	0.25	0.23	7000	20, 21, S11
4	A1	214639438_Ankur004-A1-45-4_S4	A1-45-4	**DePCR**	45	FLD0004	GTCGTCGTCT	A	2.5	22604	22264	18145	0.24	0.06	7000	20, 21, S11
5	A1	214639439_Ankur005-A1-45-5_S5	A1-45-5	**DePCR**	45	FLD0005	GTGCGTGTGT	A	2.5	27497	27071	21900	0.31	0.29	7000	20, 21, S11
6	A1	214639440_Ankur006-A1-45-6_S6	A1-45-6	**DePCR**	45	FLD0006	GCGTCGTGTA	A	2.5	30594	30146	24423	0.22	0.03	7000	20, 21, S11
7	A1	214651443_Ankur007-A1-45-7_S7	A1-45-7	**DePCR**	45	FLD0007	GTCGTGTACT	A	2.5	24900	24532	19854	0.23	0.20	7000	20, 21, S11
8	A1	214645438_Ankur008-A1-45-8_S8	A1-45-8	**DePCR**	45	FLD0008	GATGTAGCGT	A	2.5	24892	24537	19878	0.27	0.11	7000	20, 21, S11
9	A1	214653458_Ankur129-A1-55-1_S129	A1-55-1	**DePCR**	55	FLD0193	TTGTTGCTGT	A	2.5	26736	26383	21328	0.26	0.29	7000	20, 21, S11
10	A1	214651458_Ankur130-A1-55-2_S130	A1-55-2	**DePCR**	55	FLD0194	GTGTGGTTGT	A	2.5	28474	28050	22900	0.26	0.29	7000	20, 21, S11
11	A1	214662464_Ankur131-A1-55-3_S131	A1-55-3	**DePCR**	55	FLD0195	TAGGTGGAAT	A	2.5	22295	21994	17833	0.26	0.11	7000	20, 21, S11
12	A1	214644447_Ankur132-A1-55-4_S132	A1-55-4	**DePCR**	55	FLD0196	TGTAGGTGGA	A	2.5	25461	25125	20214	0.27	0.31	7000	20, 21, S11
13	A1	214644448_Ankur133-A1-55-5_S133	A1-55-5	**DePCR**	55	FLD0197	TTAGTGGTGA	A	2.5	22265	21955	17755	0.30	0.14	7000	20, 21, S11
14	A1	214644449_Ankur134-A1-55-6_S134	A1-55-6	**DePCR**	55	FLD0198	GTGAAGGTAA	A	2.5	31800	31385	25576	0.29	0.34	7000	20, 21, S11
15	A1	214661471_Ankur135-A1-55-7_S135	A1-55-7	**DePCR**	55	FLD0199	TGTTGTGGTA	A	2.5	22893	22595	18527	0.26	0.00	7000	20, 21, S11
16	A1	214645450_Ankur136-A1-55-8_S136	A1-55-8	**DePCR**	55	FLD0200	GTTGATGAGT	A	2.5	36120	35654	29575	0.23	0.09	7000	20, 21, S11
17	A2	214646443_Ankur009-A2-45-1_S9	A2-45-1	**DePCR**	45	FLD0009	GAGTGATCGT	A	2.5	26209	25875	21104	2.31	0.14	7000	20, S12
18	A2	214660446_Ankur010-A2-45-2_S10	A2-45-2	**DePCR**	45	FLD0010	CGCTATCAGT	A	2.5	10636	10485	8410	2.33	0.23	7000	20, S12
19	A2	214659446_Ankur011-A2-45-3_S11	A2-45-3	**DePCR**	45	FLD0011	CGCTGTAGTC	A	2.5	13271	13104	10504	2.32	0.14	7000	20, S12
20	A2	214647438_Ankur012-A2-45-4_S12	A2-45-4	**DePCR**	45	FLD0012	GCTAGTGAGT	A	2.5	25609	25282	20599	2.30	0.14	7000	20, S12
21	A2	214661447_Ankur013-A2-45-5_S13	A2-45-5	**DePCR**	45	FLD0013	GAGCTAGTGA	A	2.5	24160	23869	19279	2.31	0.46	7000	20, S12
22	A2	214646444_Ankur014-A2-45-6_S14	A2-45-6	**DePCR**	45	FLD0014	CGTGCTGTCA	A	2.5	17036	16837	13503	2.32	0.11	7000	20, S12
23	A2	214659447_Ankur015-A2-45-7_S15	A2-45-7	**DePCR**	45	FLD0015	GATCGTCTCT	A	2.5	24371	24114	19782	2.31	0.29	7000	20, S12
24	A2	214661448_Ankur016-A2-45-8_S16	A2-45-8	**DePCR**	45	FLD0016	GTGCTGTCGT	A	2.5	27262	26922	21774	2.30	0.26	7000	20, S12
25	A2	214653459_Ankur137-A2-55-1_S137	A2-55-1	**DePCR**	55	FLD0201	GGTCAGTGTA	A	2.5	43420	42934	34962	2.25	0.97	7000	20, S12
26	A2	214659462_Ankur138-A2-55-2_S138	A2-55-2	**DePCR**	55	FLD0202	GTAATGGAGT	A	2.5	38711	38223	31164	2.28	0.09	7000	20, S12
27	A2	214644450_Ankur139-A2-55-3_S139	A2-55-3	**DePCR**	55	FLD0203	CTCGTTATTC	A	2.5	24514	24277	19809	2.26	0.11	7000	20, S12
28	A2	214644451_Ankur140-A2-55-4_S140	A2-55-4	**DePCR**	55	FLD0204	GGAAGTAAGG	A	2.5	40211	39756	31960	2.26	0.06	7000	20, S12
29	A2	214661473_Ankur141-A2-55-5_S141	A2-55-5	**DePCR**	55	FLD0205	CGGTGTGTGT	A	2.5	26713	26398	21108	2.26	0.06	7000	20, S12
30	A2	214647461_Ankur142-A2-55-6_S142	A2-55-6	**DePCR**	55	FLD0206	CGTCTTCTTA	A	2.5	23891	23632	19512	2.24	0.66	7000	20, S12
31	A2	214640453_Ankur143-A2-55-7_S143	A2-55-7	**DePCR**	55	FLD0207	TGTGAATCTC	A	2.5	33895	33582	27555	2.26	0.40	7000	20, S12
32	A2	214651459_Ankur144-A2-55-8_S144	A2-55-8	**DePCR**	55	FLD0208	CTAATCGTGT	A	2.5	26669	26396	21373	2.26	0.09	7000	20, S12
33	A3	214654441_Ankur017-A3-45-1_S17	A3-45-1	**DePCR**	45	FLD0017	TGAGCGTGCT	A	2.5	17976	17790	14232	2.30	0.20	7000	20, S13
34	A3	214640436_Ankur018-A3-45-2_S18	A3-45-2	**DePCR**	45	FLD0018	CATGTCGTCA	A	2.5	12633	12485	10117	2.32	0.14	7000	20, S13
35	A3	214654442_Ankur019-A3-45-3_S19	A3-45-3	**DePCR**	45	FLD0019	TCAGTGTCTC	A	2.5	15576	15423	12408	2.30	0.26	7000	20, S13
36	A3	214656450_Ankur020-A3-45-4_S20	A3-45-4	**DePCR**	45	FLD0020	GTGCTCATGT	A	2.5	17686	17496	14100	2.31	0.46	7000	20, S13
37	A3	214647441_Ankur021-A3-45-5_S21	A3-45-5	**DePCR**	45	FLD0021	CGTATCTCGA	A	2.5	12402	12269	9850	2.33	0.63	7000	20, S13
38	A3	214651445_Ankur022-A3-45-6_S22	A3-45-6	**DePCR**	45	FLD0022	GTCATGCGTC	A	2.5	22081	21815	17380	2.31	0.51	7000	20, S13
39	A3	214645439_Ankur023-A3-45-7_S23	A3-45-7	**DePCR**	45	FLD0023	CTATGCGATC	A	2.5	14770	14622	11698	2.31	0.23	7000	20, S13
40	A3	214656451_Ankur024-A3-45-8_S24	A3-45-8	**DePCR**	45	FLD0024	TGCTATGCTG	A	2.5	20548	20343	16420	2.33	0.46	7000	20, S13
41	A3	214640454_Ankur145-A3-55-1_S145	A3-55-1	**DePCR**	55	FLD0209	CTCTTAGTTC	A	2.5	25170	24914	20118	2.26	0.20	7000	20, S13
42	A3	214663449_Ankur146-A3-55-2_S146	A3-55-2	**DePCR**	55	FLD0210	GGATAGGATC	A	2.5	36725	36354	29256	2.23	0.20	7000	20, S13
43	A3	214646468_Ankur147-A3-55-3_S147	A3-55-3	**DePCR**	55	FLD0211	GGTGTCTTGT	A	2.5	35953	35551	29014	2.25	0.20	7000	20, S13
44	A3	214657465_Ankur148-A3-55-4_S148	A3-55-4	**DePCR**	55	FLD0212	GATGGTTGTA	A	2.5	34298	33952	27604	2.26	0.09	7000	20, S13
45	A3	214640455_Ankur149-A3-55-5_S149	A3-55-5	**DePCR**	55	FLD0213	CCTCGTTGTT	A	2.5	26826	26535	21587	2.24	0.11	7000	20, S13
46	A3	214651460_Ankur150-A3-55-6_S150	A3-55-6	**DePCR**	55	FLD0214	GGTTGGAGTT	A	2.5	48675	48151	39122	2.26	0.06	7000	20, S13
47	A3	214661474_Ankur151-A3-55-7_S151	A3-55-7	**DePCR**	55	FLD0215	TGGTGTCCGT	A	2.5	28399	28109	22346	2.26	0.11	7000	20, S13
48	A3	214646469_Ankur152-A3-55-8_S152	A3-55-8	**DePCR**	55	FLD0216	CGTTAGCGTA	A	2.5	25751	25479	20644	2.27	0.14	7000	20, S13

#	Exp	SRA Sample Name	SampleID	PCR Method	Annealing Temp	Barcode Name	Barcode Sequence	Input DNA Template	gDNA Conc (ng/ul)	Raw Reads	Merged Reads	Trimmed	Shannon	Ideal Score	Rarefaction	Ref. Figure
49	A4	214644437_Ankur025-A4-45-1_S25	A4-45-1	**DePCR**	45	FLD0025	TGTGTGCATG	A	2.5	13399	13262	10357	3.24	0.26	7000	20, S14
50	A4	214639441_Ankur026-A4-45-2_S26	A4-45-2	**DePCR**	45	FLD0026	GAGTGTCACT	A	2.5	21108	20903	16722	3.22	0.23	7000	20, S14
51	A4	214639442_Ankur027-A4-45-3_S27	A4-45-3	**DePCR**	45	FLD0027	CTAGTCTCGT	A	2.5	14160	14026	11235	3.27	0.29	7000	20, S14
52	A4	214646445_Ankur028-A4-45-4_S28	A4-45-4	**DePCR**	45	FLD0028	GAGTGCATCT	A	2.5	16759	16601	13219	3.22	0.09	7000	20, S14
53	A4	214656452_Ankur029-A4-45-5_S29	A4-45-5	**DePCR**	45	FLD0029	TGCGTAGTCG	A	2.5	17584	17399	13985	3.25	0.11	7000	20, S14
54	A4	214651446_Ankur030-A4-45-6_S30	A4-45-6	**DePCR**	45	FLD0030	CTGTGTCGTC	A	2.5	14909	14755	11646	3.27	0.14	7000	20, S14
55	A4	214662451_Ankur031-A4-45-7_S31	A4-45-7	**DePCR**	45	FLD0031	CTGTAGTGCG	A	2.5	12971	12827	9832	3.27	0.11	7000	20, S14
56	A4	214645441_Ankur032-A4-45-8_S32	A4-45-8	**DePCR**	45	FLD0032	GTGCGCTAGT	A	2.5	22953	22725	17909	3.22	0.46	7000	20, S14
57	A4	214644453_Ankur153-A4-55-1_S153	A4-55-1	**DePCR**	55	FLD0217	TACTAGGATC	A	2.5	25046	24801	19613	2.60	0.20	7000	20, S14
58	A4	214648454_Ankur154-A4-55-2_S154	A4-55-2	**DePCR**	55	FLD0218	GTCTCAATGT	A	2.5	26145	25909	20668	2.61	0.11	7000	20, S14
59	A4	214653461_Ankur155-A4-55-3_S155	A4-55-3	**DePCR**	55	FLD0219	GATGAGGTAT	A	2.5	36776	36423	28723	2.59	0.03	7000	20, S14
60	A4	214645451_Ankur156-A4-55-4_S156	A4-55-4	**DePCR**	55	FLD0220	GGTGTTAGTG	A	2.5	34498	34136	27055	2.60	0.14	7000	20, S14
61	A4	214662466_Ankur157-A4-55-5_S157	A4-55-5	**DePCR**	55	FLD0221	CATTCTCTGA	A	2.5	24559	24353	19177	2.56	0.31	7000	20, S14
62	A4	214662467_Ankur158-A4-55-6_S158	A4-55-6	**DePCR**	55	FLD0222	CATCTGGAGT	A	2.5	27127	26876	21309	2.54	0.29	7000	20, S14
63	A4	214657466_Ankur159-A4-55-7_S159	A4-55-7	**DePCR**	55	FLD0223	GAATGGAAGA	A	2.5	39819	39471	31305	2.58	0.14	7000	20, S14
64	A4	214653462_Ankur160-A4-55-8_S160	A4-55-8	**DePCR**	55	FLD0224	GGCTGTGATC	A	2.5	38638	38257	30041	2.59	0.03	7000	20, S14
65	A6	214662453_Ankur033-A6-45-1_S33	A6-45-1	**DePCR**	45	FLD0033	TGTGCTCGCA	A	2.5	15364	15191	11918	3.45	0.57	7000	20, S15
66	A6	214639444_Ankur034-A6-45-2_S34	A6-45-2	**DePCR**	45	FLD0034	GATGCGAGCT	A	2.5	19278	19066	15127	3.39	0.37	7000	20, S15
67	A6	214646447_Ankur035-A6-45-3_S35	A6-45-3	**DePCR**	45	FLD0035	CTGTACGTGA	A	2.5	11167	11040	8455	3.46	1.91	7000	20, S15
68	A6	214640438_Ankur036-A6-45-4_S36	A6-45-4	**DePCR**	45	FLD0036	GCGATGATGA	A	2.5	23662	23405	18613	3.37	0.09	7000	20, S15
69	A6	214644438_Ankur037-A6-45-5_S37	A6-45-5	**DePCR**	45	FLD0037	TGTCGAGTCA	A	2.5	16504	16337	12962	3.44	0.17	7000	20, S15
70	A6	214657446_Ankur038-A6-45-6_S38	A6-45-6	**DePCR**	45	FLD0038	GTCTACTGTC	A	2.5	22237	22017	17905	3.40	0.43	7000	20, S15
71	A6	214653448_Ankur039-A6-45-7_S39	A6-45-7	**DePCR**	45	FLD0039	CAGTCAGAGT	A	2.5	16452	16283	13033	3.49	0.26	7000	20, S15
72	A6	214662454_Ankur040-A6-45-8_S40	A6-45-8	**DePCR**	45	FLD0040	CGCAGTCTAT	A	2.5	21797	21586	17187	3.44	0.09	7000	20, S15
73	A6	214648455_Ankur161-A6-55-1_S161	A6-55-1	**DePCR**	55	FLD0225	TGGTGCTGGA	A	2.5	26257	25975	20454	2.75	0.20	7000	20, S15
74	A6	214659463_Ankur162-A6-55-2_S162	A6-55-2	**DePCR**	55	FLD0226	TATGGTAAGG	A	2.5	31581	31251	25102	2.74	0.14	7000	20, S15
75	A6	214646471_Ankur163-A6-55-3_S163	A6-55-3	**DePCR**	55	FLD0227	GTTCGATTGT	A	2.5	38524	38134	31113	2.79	0.23	7000	20, S15
76	A6	214657467_Ankur164-A6-55-4_S164	A6-55-4	**DePCR**	55	FLD0228	GGTAGAATGA	A	2.5	35956	35599	28708	2.81	0.17	7000	20, S15
77	A6	214660466_Ankur165-A6-55-5_S165	A6-55-5	**DePCR**	55	FLD0229	TTCTCATCGT	A	2.5	28500	28238	22994	2.75	0.14	7000	20, S15
78	A6	214648456_Ankur166-A6-55-6_S166	A6-55-6	**DePCR**	55	FLD0230	CTCAATCGTA	A	2.5	28160	27915	22717	2.76	0.11	7000	20, S15
79	A6	214652466_Ankur167-A6-55-7_S167	A6-55-7	**DePCR**	55	FLD0231	CGCTAATGTA	A	2.5	31353	31075	25164	2.71	0.54	7000	20, S15
80	A6	214664451_Ankur168-A6-55-8_S168	A6-55-8	**DePCR**	55	FLD0232	GCGTCTGAAT	A	2.5	40228	39842	32369	2.77	0.26	7000	20, S15
81	B1	214659450_Ankur041-B1-45-1_S41	B1-45-1	**DePCR**	45	FLD0041	GTATGAGCAC	B	2.5	40846	40394	32793	0.27	22.26	7000	21, 22, 23, S16
82	B1	214660449_Ankur042-B1-45-2_S42	B1-45-2	**DePCR**	45	FLD0042	CGAGTGCTGT	B	2.5	24913	24621	19776	0.27	22.89	7000	21, 22, 23, S16
83	B1	214661450_Ankur043-B1-45-3_S43	B1-45-3	**DePCR**	45	FLD0043	TATAGCACGC	B	2.5	20629	20386	15893	0.28	20.46	7000	21, 22, 23, S16
84	B1	214646449_Ankur044-B1-45-4_S44	B1-45-4	**DePCR**	45	FLD0044	TCATGCGCGA	B	2.5	22712	22467	17661	0.24	22.00	7000	21, 22, 23, S16
85	B1	214654444_Ankur045-B1-45-5_S45	B1-45-5	**DePCR**	45	FLD0045	TATGCGCTGC	B	2.5	18330	18131	13627	0.29	23.63	7000	21, 22, 23, S16
86	B1	214647442_Ankur046-B1-45-6_S46	B1-45-6	**DePCR**	45	FLD0046	TCTCTGTGCA	B	2.5	28169	27863	22536	0.28	21.00	7000	21, 22, 23, S16
87	B1	214647443_Ankur047-B1-45-7_S47	B1-45-7	**DePCR**	45	FLD0047	CTATCGCGTG	B	2.5	18120	17903	14400	0.28	23.71	7000	21, 22, 23, S16
88	B1	214650446_Ankur048-B1-45-8_S48	B1-45-8	**DePCR**	45	FLD0048	TACGCTGCTG	B	2.5	27804	27506	22124	0.26	23.43	7000	21, 22, 23, S16
89	B1	214647462_Ankur169-B1-55-1_S169	B1-55-1	**DePCR**	55	FLD0233	TTCTGTTGCC	B	2.5	29914	29586	24123	0.38	22.97	7000	21, 22, 23, S16
90	B1	214653463_Ankur170-B1-55-2_S170	B1-55-2	**DePCR**	55	FLD0234	TTGTCCTTGC	B	2.5	31551	31203	25098	0.39	23.17	7000	21, 22, 23, S16
91	B1	214645452_Ankur171-B1-55-3_S171	B1-55-3	**DePCR**	55	FLD0235	CCTGTGTAGA	B	2.5	31728	31402	25736	0.35	22.31	7000	21, 22, 23, S16
92	B1	214647464_Ankur172-B1-55-4_S172	B1-55-4	**DePCR**	55	FLD0236	GATAAGAAGG	B	2.5	44053	43556	35776	0.34	22.66	7000	21, 22, 23, S16
93	B1	214653464_Ankur173-B1-55-5_S173	B1-55-5	**DePCR**	55	FLD0237	CAGGTCACAT	B	2.5	25388	25110	20526	0.37	24.03	7000	21, 22, 23, S16
94	B1	214644455_Ankur174-B1-55-6_S174	B1-55-6	**DePCR**	55	FLD0238	GCCATGTCAT	B	2.5	49358	48807	40055	0.33	23.06	7000	21, 22, 23, S16
95	B1	214648458_Ankur175-B1-55-7_S175	B1-55-7	**DePCR**	55	FLD0239	TCTGCCTATA	B	2.5	42913	42467	34819	0.38	23.89	7000	21, 22, 23, S16
96	B1	214662469_Ankur176-B1-55-8_S176	B1-55-8	**DePCR**	55	FLD0240	CTTAGTTCGC	B	2.5	30678	30345	24199	0.37	23.71	7000	21, 22, 23, S16

#	Exp	SRA Sample Name	SampleID	PCR Method	Annealing Temp	Barcode Name	Barcode Sequence	Input DNA Template	gDNA Conc (ng/ul)	Raw Reads	Merged Reads	Trimmed	Shannon	Ideal Score	Rarefaction	Ref. Figure
97	B2	214651449_Ankur049-B2-45-1_S49	B2-45-1	**DePCR**	45	FLD0049	CTGCATGATC	B	2.5	14618	14462	11611	2.33	16.00	7000	22, 23, 24, S17
98	B2	214656455_Ankur050-B2-45-2_S50	B2-45-2	**DePCR**	45	FLD0050	CGCGTATCAT	B	2.5	18443	18247	14675	2.34	17.71	7000	22, 23, 24, S17
99	B2	214652452_Ankur051-B2-45-3_S51	B2-45-3	**DePCR**	45	FLD0051	GTATCTCTCG	B	2.5	18010	17789	14428	2.32	18.77	7000	22, 23, 24, S17
100	B2	214645443_Ankur052-B2-45-4_S52	B2-45-4	**DePCR**	45	FLD0052	GCTCATATGC	B	2.5	26451	26108	20756	2.34	18.37	7000	22, 23, 24, S17
101	B2	214653450_Ankur053-B2-45-5_S53	B2-45-5	**DePCR**	45	FLD0053	CACTATGTCG	B	2.5	21476	21258	17103	2.35	17.86	7000	22, 23, 24, S17
102	B2	214647445_Ankur054-B2-45-6_S54	B2-45-6	**DePCR**	45	FLD0054	TAGCGCGTAG	B	2.5	14709	14535	11547	2.32	19.40	7000	22, 23, 24, S17
103	B2	214650448_Ankur055-B2-45-7_S55	B2-45-7	**DePCR**	45	FLD0055	CGTCACAGTA	B	2.5	18340	18122	14748	2.35	19.23	7000	22, 23, 24, S17
104	B2	214660452_Ankur056-B2-45-8_S56	B2-45-8	**DePCR**	45	FLD0056	TCGCGTGAGA	B	2.5	21240	21008	16693	2.33	16.63	7000	22, 23, 24, S17
105	B2	214650465_Ankur177-B2-55-1_S177	B2-55-1	**DePCR**	55	FLD0241	CGTAATGAGC	B	2.5	36455	36091	28446	2.31	16.00	7000	22, 23, 24, S17
106	B2	214657468_Ankur178-B2-55-2_S178	B2-55-2	**DePCR**	55	FLD0242	TTGCTTAGTC	B	2.5	37685	37337	30418	2.31	15.94	7000	22, 23, 24, S17
107	B2	214644457_Ankur179-B2-55-3_S179	B2-55-3	**DePCR**	55	FLD0243	TCTTGTTCAC	B	2.5	35905	35581	28833	2.32	12.77	7000	22, 23, 24, S17
108	B2	214656468_Ankur180-B2-55-4_S180	B2-55-4	**DePCR**	55	FLD0244	GTGGCTTCGT	B	2.5	45125	44601	35516	2.31	16.66	7000	22, 23, 24, S17
109	B2	214648459_Ankur181-B2-55-5_S181	B2-55-5	**DePCR**	55	FLD0245	TGTTCGATAG	B	2.5	39219	38857	31107	2.29	13.91	7000	22, 23, 24, S17
110	B2	214654456_Ankur182-B2-55-6_S182	B2-55-6	**DePCR**	55	FLD0246	TCATTCAGTG	B	2.5	38241	37871	30768	2.31	14.06	7000	22, 23, 24, S17
111	B2	214660467_Ankur183-B2-55-7_S183	B2-55-7	**DePCR**	55	FLD0247	GTGGAGAGCT	B	2.5	45082	44649	36160	2.29	14.97	7000	22, 23, 24, S17
112	B2	214646474_Ankur184-B2-55-8_S184	B2-55-8	**DePCR**	55	FLD0248	GTAGAAGTGG	B	2.5	48743	48270	38741	2.30	16.46	7000	22, 23, 24, S17
113	B3	214659453_Ankur057-B3-45-1_S57	B3-45-1	**DePCR**	45	FLD0057	TACATCGCTG	B	2.5	15185	15058	12093	2.27	18.23	7000	22, 23, S18
114	B3	214662455_Ankur058-B3-45-2_S58	B3-45-2	**DePCR**	45	FLD0058	GTGAGAGACA	B	2.5	26630	26377	21461	2.24	20.74	7000	22, 23, S18
115	B3	214639447_Ankur059-B3-45-3_S59	B3-45-3	**DePCR**	45	FLD0059	GACTGTACGT	B	2.5	17776	17587	14028	2.27	19.14	7000	22, 23, S18
116	B3	214646454_Ankur060-B3-45-4_S60	B3-45-4	**DePCR**	45	FLD0060	GCACGTAGCT	B	2.5	19283	19057	15016	2.25	19.40	7000	22, 23, S18
117	B3	214644439_Ankur061-B3-45-5_S61	B3-45-5	**DePCR**	45	FLD0061	TCACGCTATG	B	2.5	18120	17949	14515	2.24	19.97	7000	22, 23, S18
118	B3	214648444_Ankur062-B3-45-6_S62	B3-45-6	**DePCR**	45	FLD0062	CGTACTACGT	B	2.5	11212	11106	8879	NA	NA	7000	Removed from analysis
119	B3	214648445_Ankur063-B3-45-7_S63	B3-45-7	**DePCR**	45	FLD0063	CAGCTGAGTA	B	2.5	25407	25164	20346	2.26	20.80	7000	22, 23, S18
120	B3	214644440_Ankur064-B3-45-8_S64	B3-45-8	**DePCR**	45	FLD0064	GAGATCAGTC	B	2.5	28880	28619	23159	2.25	16.83	7000	22, 23, S18
121	B3	214645454_Ankur185-B3-55-1_S185	B3-55-1	**DePCR**	55	FLD0249	TGGAGCATGT	B	2.5	34582	34257	27632	2.21	15.91	7000	22, 23, S18
122	B3	214662470_Ankur186-B3-55-2_S186	B3-55-2	**DePCR**	55	FLD0250	GAAGGAGATA	B	2.5	46421	45968	37436	2.22	11.37	7000	22, 23, S18
123	B3	214653465_Ankur187-B3-55-3_S187	B3-55-3	**DePCR**	55	FLD0251	CGAATGTATG	B	2.5	38789	38428	31005	2.23	14.09	7000	22, 23, S18
124	B3	214663452_Ankur188-B3-55-4_S188	B3-55-4	**DePCR**	55	FLD0252	TCGTGAATGA	B	2.5	39371	38978	31579	2.23	12.80	7000	22, 23, S18
125	B3	214645457_Ankur189-B3-55-5_S189	B3-55-5	**DePCR**	55	FLD0253	GAATAGCTGA	B	2.5	51336	50874	41058	2.22	15.31	7000	22, 23, S18
126	B3	214645458_Ankur190-B3-55-6_S190	B3-55-6	**DePCR**	55	FLD0254	TTGTCACATC	B	2.5	33892	33588	27032	2.24	14.69	7000	22, 23, S18
127	B3	214651462_Ankur191-B3-55-7_S191	B3-55-7	**DePCR**	55	FLD0255	CTGGAGGCTA	B	2.5	33464	33188	25826	2.22	15.74	7000	22, 23, S18
128	B3	214663453_Ankur192-B3-55-8_S192	B3-55-8	**DePCR**	55	FLD0256	TGTCAGCTTA	B	2.5	35977	35680	28959	2.23	15.37	7000	22, 23, S18
129	B4	214656460_Ankur065-B4-45-1_S65	B4-45-1	**DePCR**	45	FLD0065	TACTGAGCTG	B	2.5	16472	16320	12861	3.25	11.54	7000	22, 23, S19
130	B4	214657452_Ankur066-B4-45-2_S66	B4-45-2	**DePCR**	45	FLD0066	TAGTAGCGCG	B	2.5	12691	12539	9489	3.26	11.80	7000	22, 23, S19
131	B4	214662458_Ankur067-B4-45-3_S67	B4-45-3	**DePCR**	45	FLD0067	GACGTCTGCT	B	2.5	15066	14914	11837	3.28	11.37	7000	22, 23, S19
132	B4	214662459_Ankur068-B4-45-4_S68	B4-45-4	**DePCR**	45	FLD0068	GTACTCGCGA	B	2.5	30578	30236	23263	3.27	13.51	7000	22, 23, S19
133	B4	214647450_Ankur069-B4-45-5_S69	B4-45-5	**DePCR**	45	FLD0069	TCTGAGCGCA	B	2.5	19884	19682	15391	3.27	13.57	7000	22, 23, S19
134	B4	214653453_Ankur070-B4-45-6_S70	B4-45-6	**DePCR**	45	FLD0070	TAGACGTGCT	B	2.5	11388	11279	8488	NA	NA	7000	Removed from analysis
135	B4	214657454_Ankur071-B4-45-7_S71	B4-45-7	**DePCR**	45	FLD0071	GTGACTCGTC	B	2.5	33076	32763	26057	3.26	10.49	7000	22, 23, S19
136	B4	214661458_Ankur072-B4-45-8_S72	B4-45-8	**DePCR**	45	FLD0072	TCGAGTAGCG	B	2.5	13283	13118	9690	3.27	10.49	7000	22, 23, S19
137	B4	214663454_Ankur193-B4-55-1_S193	B4-55-1	**DePCR**	55	FLD0257	GTTCTTCGTA	B	2.5	40422	40022	32058	3.23	10.03	7000	22, 23, S19
138	B4	214660468_Ankur194-B4-55-2_S194	B4-55-2	**DePCR**	55	FLD0258	TTACACGTTC	B	2.5	35125	34787	27604	3.22	10.66	7000	22, 23, S19
139	B4	214647466_Ankur195-B4-55-3_S195	B4-55-3	**DePCR**	55	FLD0259	GTAGCCAGTA	B	2.5	32995	32631	25824	3.23	11.26	7000	22, 23, S19
140	B4	214657469_Ankur196-B4-55-4_S196	B4-55-4	**DePCR**	55	FLD0260	TGAGAAGGTA	B	2.5	48915	48489	38788	3.22	10.31	7000	22, 23, S19
141	B4	214653467_Ankur197-B4-55-5_S197	B4-55-5	**DePCR**	55	FLD0261	CCATATGATC	B	2.5	35094	34767	27347	3.22	12.46	7000	22, 23, S19
142	B4	214661477_Ankur198-B4-55-6_S198	B4-55-6	**DePCR**	55	FLD0262	CGATCCTATA	B	2.5	31844	31528	25041	3.21	12.43	7000	22, 23, S19
143	B4	214644461_Ankur199-B4-55-7_S199	B4-55-7	**DePCR**	55	FLD0263	TGACTAGCTT	B	2.5	39182	38881	31008	3.20	13.09	7000	22, 23, S19
144	B4	214661478_Ankur200-B4-55-8_S200	B4-55-8	**DePCR**	55	FLD0264	TAACTCTGCT	B	2.5	29606	29358	23423	3.22	10.00	7000	22, 23, S19

#	Exp	SRA Sample Name	SampleID	PCR Method	Annealing Temp	Barcode Name	Barcode Sequence	Input DNA Template	gDNA Conc (ng/ul)	Raw Reads	Merged Reads	Trimmed	Shannon	Ideal Score	Rarefaction	Ref. Figure
145	C1	214657455_Ankur073-C1-45-1_S73	C1-45-1	**DePCR**	45	FLD0073	CGTATGATGT	C	2.5	28081	27737	22751	0.27	21.22	7000	21, S20
146	C1	214639449_Ankur074-C1-45-2_S74	C1-45-2	**DePCR**	45	FLD0074	TAGTCTGTCA	C	2.5	26265	26003	21370	0.27	19.93	7000	21, S20
147	C1	214645446_Ankur075-C1-45-3_S75	C1-45-3	**DePCR**	45	FLD0075	TGTCTCTATC	C	2.5	28826	28509	23360	0.26	20.56	7000	21, S20
148	C1	214652456_Ankur076-C1-45-4_S76	C1-45-4	**DePCR**	45	FLD0076	CTAGAGTATC	C	2.5	25671	25376	20681	0.29	21.45	7000	21, S20
149	C1	214657456_Ankur077-C1-45-5_S77	C1-45-5	**DePCR**	45	FLD0077	TATCATGTGC	C	2.5	24377	24108	19249	0.30	18.96	7000	21, S20
150	C1	214660458_Ankur078-C1-45-6_S78	C1-45-6	**DePCR**	45	FLD0078	CATGAGTGTA	C	2.5	27715	27424	22466	0.29	22.08	7000	21, S20
151	C1	214640445_Ankur079-C1-45-7_S79	C1-45-7	**DePCR**	45	FLD0079	TGTCGTCATA	C	2.5	33451	33127	27241	0.30	23.96	7000	21, S20
152	C1	214660459_Ankur080-C1-45-8_S80	C1-45-8	**DePCR**	45	FLD0080	TATCTCATGC	C	2.5	23577	23340	18744	0.27	20.11	7000	21, S20
153	C1	214662471_Ankur201-C1-55-1_S201	C1-55-1	**DePCR**	55	FLD0265	TCGAATGTGC	C	2.5	27282	26968	21628	0.34	23.07	7000	21, S20
154	C1	214661479_Ankur202-C1-55-2_S202	C1-55-2	**DePCR**	55	FLD0266	TCGCTGAACA	C	2.5	30460	30144	24754	0.37	23.30	7000	21, S20
155	C1	214654457_Ankur203-C1-55-3_S203	C1-55-3	**DePCR**	55	FLD0267	GCGTTATTGC	C	2.5	45026	44509	36085	0.35	23.30	7000	21, S20
156	C1	214659466_Ankur204-C1-55-4_S204	C1-55-4	**DePCR**	55	FLD0268	GAACTATCAC	C	2.5	45291	44836	37046	0.33	23.64	7000	21, S20
157	C1	214652469_Ankur205-C1-55-5_S205	C1-55-5	**DePCR**	55	FLD0269	TCGAGGTACT	C	2.5	39693	39258	32100	0.35	23.67	7000	21, S20
158	C1	214662472_Ankur206-C1-55-6_S206	C1-55-6	**DePCR**	55	FLD0270	TGCGGATGGT	C	2.5	41053	40578	32814	0.41	24.16	7000	21, S20
159	C1	214662473_Ankur207-C1-55-7_S207	C1-55-7	**DePCR**	55	FLD0271	TTCGAGCTAT	C	2.5	26856	26578	21764	0.39	24.36	7000	21, S20
160	C1	214665453_Ankur208-C1-55-8_S208	C1-55-8	**DePCR**	55	FLD0272	GGTCTGGTGT	C	2.5	38729	38268	31382	0.37	25.44	7000	21, S20
161	C2	214654449_Ankur081-C2-45-1_S81	C2-45-1	**DePCR**	45	FLD0081	TGTGTCACTA	C	2.5	25995	25745	20883	2.35	17.23	7000	S21
162	C2	214656463_Ankur082-C2-45-2_S82	C2-45-2	**DePCR**	45	FLD0082	TATCGATGCT	C	2.5	24292	24027	19417	2.34	19.25	7000	S21
163	C2	214646456_Ankur083-C2-45-3_S83	C2-45-3	**DePCR**	45	FLD0083	TAGAGTCTGT	C	2.5	28561	28274	23009	2.32	19.31	7000	S21
164	C2	214654451_Ankur084-C2-45-4_S84	C2-45-4	**DePCR**	45	FLD0084	CATGCATCAT	C	2.5	20567	20363	16443	2.30	20.94	7000	S21
165	C2	214650455_Ankur085-C2-45-5_S85	C2-45-5	**DePCR**	45	FLD0085	TGATCAGTCA	C	2.5	28512	28236	22953	2.32	17.72	7000	S21
166	C2	214650456_Ankur086-C2-45-6_S86	C2-45-6	**DePCR**	45	FLD0086	CGTCTATGAT	C	2.5	18544	18334	14911	2.32	17.91	7000	S21
167	C2	214662460_Ankur087-C2-45-7_S87	C2-45-7	**DePCR**	45	FLD0087	GTGATACTGA	C	2.5	33422	33081	26819	2.33	19.85	7000	S21
168	C2	214657457_Ankur088-C2-45-8_S88	C2-45-8	**DePCR**	45	FLD0088	CTAGATCTGA	C	2.5	19064	18864	15301	2.33	20.37	7000	S21
169	C2	214653468_Ankur209-C2-55-1_S209	C2-55-1	**DePCR**	55	FLD0273	CTAAGTCATG	C	2.5	28027	27770	22489	2.31	15.00	7000	S21
170	C2	214645462_Ankur210-C2-55-2_S210	C2-55-2	**DePCR**	55	FLD0274	TTGCAGATCA	C	2.5	28084	27791	22653	2.32	16.32	7000	S21
171	C2	214645463_Ankur211-C2-55-3_S211	C2-55-3	**DePCR**	55	FLD0275	CTGCGAATGT	C	2.5	34290	33932	27410	2.31	15.38	7000	S21
172	C2	214657472_Ankur212-C2-55-4_S212	C2-55-4	**DePCR**	55	FLD0276	CTGTTCTAGC	C	2.5	29222	28930	22999	2.31	18.83	7000	S21
173	C2	214646477_Ankur213-C2-55-5_S213	C2-55-5	**DePCR**	55	FLD0277	CACTTGTGTG	C	2.5	29934	29634	23799	2.29	15.69	7000	S21
174	C2	214652470_Ankur214-C2-55-6_S214	C2-55-6	**DePCR**	55	FLD0278	TGGATGACAT	C	2.5	46853	46440	37793	2.29	14.35	7000	S21
175	C2	214648463_Ankur215-C2-55-7_S215	C2-55-7	**DePCR**	55	FLD0279	GATCCTGAGC	C	2.5	40867	40428	31909	2.31	16.89	7000	S21
176	C2	214654458_Ankur216-C2-55-8_S216	C2-55-8	**DePCR**	55	FLD0280	GTCGGTCTGA	C	2.5	36998	36635	29490	2.30	15.84	7000	S21
177	C3	214640447_Ankur089-C3-45-1_S89	C3-45-1	**DePCR**	45	FLD0089	TATCAGTCTG	C	2.5	23399	23177	18888	2.23	18.05	7000	S22
178	C3	214662461_Ankur090-C3-45-2_S90	C3-45-2	**DePCR**	45	FLD0090	TCAGATGCTA	C	2.5	28498	28234	22916	2.26	19.63	7000	S22
179	C3	214640448_Ankur091-C3-45-3_S91	C3-45-3	**DePCR**	45	FLD0091	TATGTACGTG	C	2.5	23609	23351	18716	2.24	20.92	7000	S22
180	C3	214644443_Ankur092-C3-45-4_S92	C3-45-4	**DePCR**	45	FLD0092	CTATACAGTG	C	2.5	32572	32283	26006	2.27	21.03	7000	S22
181	C3	214639451_Ankur093-C3-45-5_S93	C3-45-5	**DePCR**	45	FLD0093	TGATACTCTG	C	2.5	1911	1894	1535	NA	NA	7000	Removed from analysis
182	C3	214639452_Ankur094-C3-45-6_S94	C3-45-6	**DePCR**	45	FLD0094	TCAGCGATAT	C	2.5	20869	20631	16492	2.24	20.98	7000	S22
183	C3	214651452_Ankur095-C3-45-7_S95	C3-45-7	**DePCR**	45	FLD0095	CTACTGATGA	C	2.5	18607	18412	14836	2.25	22.32	7000	S22
184	C3	214650458_Ankur096-C3-45-8_S96	C3-45-8	**DePCR**	45	FLD0096	GTAGTACACA	C	2.5	38665	38351	31146	2.24	18.39	7000	S22
185	C3	214660471_Ankur217-C3-55-1_S217	C3-55-1	**DePCR**	55	FLD0281	TGTTACGATC	C	2.5	31585	31289	24825	2.24	15.12	7000	S22
186	C3	214651463_Ankur218-C3-55-2_S218	C3-55-2	**DePCR**	55	FLD0282	GTCTTGGCTC	C	2.5	31512	31209	24880	2.23	12.45	7000	S22
187	C3	214660472_Ankur219-C3-55-3_S219	C3-55-3	**DePCR**	55	FLD0283	GGTCGTGCAT	C	2.5	47160	46697	36436	2.22	15.07	7000	S22
188	C3	214663456_Ankur220-C3-55-4_S220	C3-55-4	**DePCR**	55	FLD0284	CAGGCTCAGT	C	2.5	35038	34719	27902	2.23	13.50	7000	S22
189	C3	214657473_Ankur221-C3-55-5_S221	C3-55-5	**DePCR**	55	FLD0285	TAGCTTCACT	C	2.5	35662	35332	28882	2.23	14.92	7000	S22
190	C3	214665456_Ankur222-C3-55-6_S222	C3-55-6	**DePCR**	55	FLD0286	CAGATGTCCT	C	2.5	36404	36083	29397	2.22	15.52	7000	S22
191	C3	214646479_Ankur223-C3-55-7_S223	C3-55-7	**DePCR**	55	FLD0287	TTACGCAGTG	C	2.5	26568	26316	21013	2.21	13.35	7000	S22
192	C3	214645465_Ankur224-C3-55-8_S224	C3-55-8	**DePCR**	55	FLD0288	TTCGTTCCTG	C	2.5	24228	23995	19495	2.22	14.52	7000	S22

#	Exp	SRA Sample Name	SampleID	PCR Method	Annealing Temp	Barcode Name	Barcode Sequence	Input DNA Template	gDNA Conc (ng/ul)	Raw Reads	Merged Reads	Trimmed	Shannon	Ideal Score	Rarefaction	Ref. Figure
193	D1	214657458_Ankur097-D1-45-1_S97	D1-45-1	**DePCR**	45	FLD0097	TGCTACATCA	D	2.5	26583	26343	21862	0.35	19.63	7000	21, S23
194	D1	214662463_Ankur098-D1-45-2_S98	D1-45-2	**DePCR**	45	FLD0098	AGTGTGTCTA	D	2.5	46973	46368	38114	0.38	21.11	7000	21, S23
195	D1	214646459_Ankur099-D1-45-3_S99	D1-45-3	**DePCR**	45	FLD0099	TCATATCGCG	D	2.5	21124	20926	16985	0.33	18.09	7000	21, S23
196	D1	214656464_Ankur100-D1-45-4_S100	D1-45-4	**DePCR**	45	FLD0100	TACGTATAGC	D	2.5	9509	9415	7339	NA	NA	7000	Removed from analysis
197	D1	214646461_Ankur101-D1-45-5_S101	D1-45-5	**DePCR**	45	FLD0101	CAGCTATAGC	D	2.5	14365	14225	11268	0.32	19.60	7000	21, S23
198	D1	214651453_Ankur102-D1-45-6_S102	D1-45-6	**DePCR**	45	FLD0102	TCGATGCGCT	D	2.5	25686	25408	20691	0.30	18.40	7000	21, S23
199	D1	214648449_Ankur103-D1-45-7_S103	D1-45-7	**DePCR**	45	FLD0103	GCACGCGTAT	D	2.5	19249	19016	15600	0.38	23.17	7000	21, S23
200	D1	214652457_Ankur104-D1-45-8_S104	D1-45-8	**DePCR**	45	FLD0104	GCAGTATGCG	D	2.5	31063	30702	24633	0.35	21.06	7000	21, S23
201	D1	214660473_Ankur225-D1-55-1_S225	D1-55-1	**DePCR**	55	FLD0289	CACTGCTTGA	D	2.5	24637	24413	20108	0.54	29.29	7000	21, S23
202	D1	214654459_Ankur226-D1-55-2_S226	D1-55-2	**DePCR**	55	FLD0290	TCTAGCGTGG	D	2.5	30792	30485	24576	0.54	27.60	7000	21, S23
203	D1	214662475_Ankur227-D1-55-3_S227	D1-55-3	**DePCR**	55	FLD0291	GCATAATCGC	D	2.5	36513	36117	29300	0.50	29.14	7000	21, S23
204	D1	214656470_Ankur228-D1-55-4_S228	D1-55-4	**DePCR**	55	FLD0292	GTCGTAACAC	D	2.5	36378	36008	29837	0.51	27.54	7000	21, S23
205	D1	214650470_Ankur229-D1-55-5_S229	D1-55-5	**DePCR**	55	FLD0293	GAGATTGCTA	D	2.5	39965	39621	32734	0.50	27.06	7000	21, S23
206	D1	214647469_Ankur230-D1-55-6_S230	D1-55-6	**DePCR**	55	FLD0294	GGACAGATGG	D	2.5	43163	42758	34977	0.52	26.60	7000	21, S23
207	D1	214651464_Ankur231-D1-55-7_S231	D1-55-7	**DePCR**	55	FLD0295	CTTACGTTGC	D	2.5	23263	23034	18464	0.58	26.69	7000	21, S23
208	D1	214661480_Ankur232-D1-55-8_S232	D1-55-8	**DePCR**	55	FLD0296	GTGTTCGGTC	D	2.5	37514	37084	30064	0.51	29.06	7000	21, S23
209	D2	214661465_Ankur105-D2-45-1_S105	D2-45-1	**DePCR**	45	FLD0105	TGATAGAGAG	D	2.5	22515	22329	18081	2.24	9.89	7000	S24
210	D2	214652458_Ankur106-D2-45-2_S106	D2-45-2	**DePCR**	45	FLD0106	GCTACTAGCG	D	2.5	24098	23845	19020	2.22	10.28	7000	S24
211	D2	214653455_Ankur107-D2-45-3_S107	D2-45-3	**DePCR**	45	FLD0107	TGCGAGACGT	D	2.5	22752	22534	17952	2.24	10.80	7000	S24
212	D2	214657460_Ankur108-D2-45-4_S108	D2-45-4	**DePCR**	45	FLD0108	CGATGACAGA	D	2.5	26538	26300	21134	2.26	10.06	7000	S24
213	D2	214652460_Ankur109-D2-45-5_S109	D2-45-5	**DePCR**	45	FLD0109	GACTCATGCT	D	2.5	24433	24218	19785	2.22	7.44	7000	S24
214	D2	214660464_Ankur110-D2-45-6_S110	D2-45-6	**DePCR**	45	FLD0110	GTCTGATACG	D	2.5	28027	27745	22285	2.24	8.83	7000	S24
215	D2	214661467_Ankur111-D2-45-7_S111	D2-45-7	**DePCR**	45	FLD0111	ACTAGCTGTC	D	2.5	42952	42435	34233	2.22	9.68	7000	S24
216	D2	214659458_Ankur112-D2-45-8_S112	D2-45-8	**DePCR**	45	FLD0112	GCGTAGACGA	D	2.5	19214	18691	14687	2.22	9.30	7000	S24
217	D2	214660474_Ankur233-D2-55-1_S233	D2-55-1	**DePCR**	55	FLD0297	CTCAAGAAGC	D	2.5	26441	26226	20809	1.89	15.65	7000	S24
218	D2	214657475_Ankur234-D2-55-2_S234	D2-55-2	**DePCR**	55	FLD0298	TCTCGGATAG	D	2.5	25810	25537	20371	1.92	17.30	7000	S24
219	D2	214647470_Ankur235-D2-55-3_S235	D2-55-3	**DePCR**	55	FLD0299	CTCTGGACGA	D	2.5	33346	33060	26216	1.87	17.16	7000	S24
220	D2	214652471_Ankur236-D2-55-4_S236	D2-55-4	**DePCR**	55	FLD0300	CGAGCATTGT	D	2.5	24922	24680	19709	1.90	16.53	7000	S24
221	D2	214645467_Ankur237-D2-55-5_S237	D2-55-5	**DePCR**	55	FLD0301	CCAAGAAGAA	D	2.5	39692	39325	31685	1.92	18.22	7000	S24
222	D2	214662476_Ankur238-D2-55-6_S238	D2-55-6	**DePCR**	55	FLD0302	TCCTTGTTCT	D	2.5	2874	2823	2233	NA	NA	7000	Removed from analysis
223	D2	214659467_Ankur239-D2-55-7_S239	D2-55-7	**DePCR**	55	FLD0303	GTAACGATGT	D	2.5	43015	42602	33211	1.92	20.08	7000	S24
224	D2	214654460_Ankur240-D2-55-8_S240	D2-55-8	**DePCR**	55	FLD0304	TGGACTCAGA	D	2.5	28699	28451	22793	1.90	17.19	7000	S24
225	E1	214647457_Ankur113-E1-45-1_S113	E1-45-1	**DePCR**	45	FLD0113	CTCAGCAGTG	E	2.5	17954	17755	14378	0.18	5.43	7000	21, S25
226	E1	214652462_Ankur114-E1-45-2_S114	E1-45-2	**DePCR**	45	FLD0114	CAGTCTACAT	E	2.5	24672	24411	20002	0.21	5.80	7000	21, S25
227	E1	214640451_Ankur115-E1-45-3_S115	E1-45-3	**DePCR**	45	FLD0115	TACTGCAGCG	E	2.5	21154	20919	16559	0.22	7.89	7000	21, S25
228	E1	214651455_Ankur116-E1-45-4_S116	E1-45-4	**DePCR**	45	FLD0116	TACACAGTAG	E	2.5	15133	14973	12007	0.21	7.66	7000	21, S25
229	E1	214657461_Ankur117-E1-45-5_S117	E1-45-5	**DePCR**	45	FLD0117	CACATACAGT	E	2.5	14755	14584	11636	0.19	8.83	7000	21, S25
230	E1	214661468_Ankur118-E1-45-6_S118	E1-45-6	**DePCR**	45	FLD0118	CACAGTGATG	E	2.5	24694	24445	19869	0.21	7.80	7000	21, S25
231	E1	214661469_Ankur119-E1-45-7_S119	E1-45-7	**DePCR**	45	FLD0119	CGAGCTAGCA	E	2.5	21029	20810	16711	0.20	6.40	7000	21, S25
232	E1	214659460_Ankur120-E1-45-8_S120	E1-45-8	**DePCR**	45	FLD0120	GAGACTATGC	E	2.5	9461	9213	7390	NA	NA	7000	Removed from analysis
233	E1	214657476_Ankur241-E1-55-1_S241	E1-55-1	**DePCR**	55	FLD0305	GGCATCATGC	E	2.5	34679	34280	26860	0.18	8.14	7000	21, S25
234	E1	214647471_Ankur242-E1-55-2_S242	E1-55-2	**DePCR**	55	FLD0306	GTATAACGCT	E	2.5	43759	43317	35609	0.18	9.09	7000	21, S25
235	E1	214646480_Ankur243-E1-55-3_S243	E1-55-3	**DePCR**	55	FLD0307	GCAGATAAGT	E	2.5	51624	51058	41729	0.20	7.03	7000	21, S25
236	E1	214662477_Ankur244-E1-55-4_S244	E1-55-4	**DePCR**	55	FLD0308	GTCGGCTCTA	E	2.5	32441	32088	25959	0.18	6.97	7000	21, S25
237	E1	214653469_Ankur245-E1-55-5_S245	E1-55-5	**DePCR**	55	FLD0309	TTCGATAGCA	E	2.5	41458	41067	33124	0.21	6.09	7000	21, S25
238	E1	214650471_Ankur246-E1-55-6_S246	E1-55-6	**DePCR**	55	FLD0310	GTCTAGCAGG	E	2.5	43875	43356	34807	0.20	8.60	7000	21, S25
239	E1	214665457_Ankur247-E1-55-7_S247	E1-55-7	**DePCR**	55	FLD0311	GGAACACAGG	E	2.5	44915	44436	36034	0.18	6.96	7000	21, S25
240	E1	214660477_Ankur248-E1-55-8_S248	E1-55-8	**DePCR**	55	FLD0312	TGGTTCGCTG	E	2.5	27323	27005	21769	0.21	6.86	7000	21, S25

#	Exp	SRA Sample Name	SampleID	PCR Method	Annealing Temp	Barcode Name	Barcode Sequence	Input DNA Template	gDNA Conc (ng/ul)	Raw Reads	Merged Reads	Trimmed	Shannon	Ideal Score	Rarefaction	Ref. Figure
241	E2	214657463_Ankur121-E2-45-1_S121	E2-45-1	**DePCR**	45	FLD0121	CAGAGCTAGT	E	2.5	13078	12947	10236	2.19	6.83	7000	S26
242	E2	214651456_Ankur122-E2-45-2_S122	E2-45-2	**DePCR**	45	FLD0122	CGCAGAGCAT	E	2.5	15626	15467	11986	2.17	4.70	7000	S26
243	E2	214653456_Ankur123-E2-45-3_S123	E2-45-3	**DePCR**	45	FLD0123	TGTACAGCGA	E	2.5	24247	24007	18837	2.18	6.00	7000	S26
244	E2	214651457_Ankur124-E2-45-4_S124	E2-45-4	**DePCR**	45	FLD0124	ACGTCAGTAT	E	2.5	50452	49850	39461	2.13	7.29	7000	S26
245	E2	214653457_Ankur125-E2-45-5_S125	E2-45-5	**DePCR**	45	FLD0125	TCACAGCATA	E	2.5	20356	20156	15962	2.18	6.17	7000	S26
246	E2	214656467_Ankur126-E2-45-6_S126	E2-45-6	**DePCR**	45	FLD0126	ACTGCGTGTC	E	2.5	46462	45967	35953	2.13	5.63	7000	S26
247	E2	214652463_Ankur127-E2-45-7_S127	E2-45-7	**DePCR**	45	FLD0127	CGATCGACTG	E	2.5	7750	7674	6031	NA	NA	7000	Removed from analysis
248	E2	214647458_Ankur128-E2-45-8_S128	E2-45-8	**DePCR**	45	FLD0128	GCGAGATGTA	E	2.5	18152	17778	13425	2.11	5.06	7000	S26
249	E2	214653470_Ankur249-E2-55-1_S249	E2-55-1	**DePCR**	55	FLD0313	CACATTAGCG	E	2.5	9997	9192	6594	NA	NA	7000	Removed from analysis
250	E2	214662478_Ankur250-E2-55-2_S250	E2-55-2	**DePCR**	55	FLD0314	GAAGCGCACT	E	2.5	17048	16423	12629	1.73	4.06	7000	S26
251	E2	214652472_Ankur251-E2-55-3_S251	E2-55-3	**DePCR**	55	FLD0315	GCATGCCAGT	E	2.5	47195	46724	37162	1.71	5.31	7000	S26
252	E2	214651465_Ankur252-E2-55-4_S252	E2-55-4	**DePCR**	55	FLD0316	GGAGACTGTA	E	2.5	46247	45852	36673	1.70	4.13	7000	S26
253	E2	214644464_Ankur253-E2-55-5_S253	E2-55-5	**DePCR**	55	FLD0317	TCGAACTGCA	E	2.5	30684	30402	24288	1.69	2.45	7000	S26
254	E2	214654461_Ankur254-E2-55-6_S254	E2-55-6	**DePCR**	55	FLD0318	GAGAGGACAT	E	2.5	39796	39445	31219	1.69	3.00	7000	S26
255	E2	214647472_Ankur255-E2-55-7_S255	E2-55-7	**DePCR**	55	FLD0319	GAGCACGGAA	E	2.5	35375	35047	27550	1.68	3.06	7000	S26
256	E2	214660478_Ankur256-E2-55-8_S256	E2-55-8	**DePCR**	55	FLD0320	GCTCTAACAT	E	2.5	43691	43195	34776	1.68	5.11	7000	S26
257	A1	217235192_Ankur001-A1S-45-1_S1	A1S-45-1	**TAS**	45	FLD0097	TGCTACATCA	A	2.5	21162	21069	18928	0.07	0.03	7000	20, 21, S11
258	A1	217243134_Ankur002-A1S-45-2_S2	A1S-45-2	**TAS**	45	FLD0098	AGTGTGTCTA	A	2.5	23554	23455	21071	0.09	0.06	7000	20, 21, S11
259	A1	217246043_Ankur003-A1S-45-3_S3	A1S-45-3	**TAS**	45	FLD0099	TCATATCGCG	A	2.5	18339	18241	16218	0.08	0.06	7000	20, 21, S11
260	A1	217235194_Ankur004-A1S-45-4_S4	A1S-45-4	**TAS**	45	FLD0100	TACGTATAGC	A	2.5	18628	18518	16500	0.09	0.03	7000	20, 21, S11
261	A1	217230264_Ankur005-A1S-45-5_S5	A1S-45-5	**TAS**	45	FLD0101	CAGCTATAGC	A	2.5	20856	20753	18500	0.09	0.09	7000	20, 21, S11
262	A1	217236201_Ankur006-A1S-45-6_S6	A1S-45-6	**TAS**	45	FLD0102	TCGATGCGCT	A	2.5	20893	20803	18596	0.07	0.11	7000	20, 21, S11
263	A1	217242129_Ankur007-A1S-45-7_S7	A1S-45-7	**TAS**	45	FLD0103	GCACGCGTAT	A	2.5	20679	20561	18366	0.09	0.00	7000	20, 21, S11
264	A1	217228282_Ankur008-A1S-45-8_S8	A1S-45-8	**TAS**	45	FLD0104	GCAGTATGCG	A	2.5	20538	20430	18104	0.08	0.00	7000	20, 21, S11
265	A1	217236213_Ankur041-A1S-55-1_S41	A1S-55-1	**TAS**	55	FLD0137	ACGTATCATC	A	2.5	21705	21594	19334	0.10	0.11	7000	20, 21, S11
266	A1	217244131_Ankur042-A1S-55-2_S42	A1S-55-2	**TAS**	55	FLD0138	AGTATCGTAC	A	2.5	22729	22606	20223	0.10	0.03	7000	20, 21, S11
267	A1	217245129_Ankur043-A1S-55-3_S43	A1S-55-3	**TAS**	55	FLD0139	GATACACTGA	A	2.5	20035	19943	17919	0.08	0.11	7000	20, 21, S11
268	A1	217236214_Ankur044-A1S-55-4_S44	A1S-55-4	**TAS**	55	FLD0140	GACTAGTCAG	A	2.5	20468	20364	18236	0.08	0.09	7000	20, 21, S11
269	A1	217232229_Ankur045-A1S-55-5_S45	A1S-55-5	**TAS**	55	FLD0141	GATGACTACG	A	2.5	19739	19639	17553	0.08	0.00	7000	20, 21, S11
270	A1	217231267_Ankur046-A1S-55-6_S46	A1S-55-6	**TAS**	55	FLD0142	CAGAGAGTCA	A	2.5	19005	18940	16992	0.08	0.06	7000	20, 21, S11
271	A1	217228293_Ankur047-A1S-55-7_S47	A1S-55-7	**TAS**	55	FLD0143	TCGATCGACA	A	2.5	19778	19678	17605	0.09	0.06	7000	20, 21, S11
272	A1	217248036_Ankur048-A1S-55-8_S48	A1S-55-8	**TAS**	55	FLD0144	ACTGATGTAG	A	2.5	20099	19969	17923	0.10	0.06	7000	20, 21, S11
273	A2	217235195_Ankur009-A2S-45-1_S9	A2S-45-1	**TAS**	45	FLD0105	TGATAGAGAG	A	2.5	20697	20591	18357	2.35	0.09	7000	20, S12
274	A2	217228284_Ankur010-A2S-45-2_S10	A2S-45-2	**TAS**	45	FLD0106	GCTACTAGCG	A	2.5	18696	18596	16441	2.34	0.03	7000	20, S12
275	A2	217246047_Ankur011-A2S-45-3_S11	A2S-45-3	**TAS**	45	FLD0107	TGCGAGACGT	A	2.5	18170	18081	15895	2.34	0.03	7000	20, S12
276	A2	217236205_Ankur012-A2S-45-4_S12	A2S-45-4	**TAS**	45	FLD0108	CGATGACAGA	A	2.5	19422	19315	17145	2.34	0.03	7000	20, S12
277	A2	217243137_Ankur013-A2S-45-5_S13	A2S-45-5	**TAS**	45	FLD0109	GACTCATGCT	A	2.5	19705	19618	17578	2.34	0.09	7000	20, S12
278	A2	217235197_Ankur014-A2S-45-6_S14	A2S-45-6	**TAS**	45	FLD0110	GTCTGATACG	A	2.5	17805	17722	15748	2.35	0.03	7000	20, S12
279	A2	217229260_Ankur015-A2S-45-7_S15	A2S-45-7	**TAS**	45	FLD0111	ACTAGCTGTC	A	2.5	21754	21645	19294	2.35	0.03	7000	20, S12
280	A2	217237205_Ankur016-A2S-45-8_S16	A2S-45-8	**TAS**	45	FLD0112	GCGTAGACGA	A	2.5	18668	18573	16453	2.34	0.00	7000	20, S12
281	A2	217246066_Ankur049-A2S-55-1_S49	A2S-55-1	**TAS**	55	FLD0145	ACTCGATAGT	A	2.5	22110	21990	19567	2.34	0.14	7000	20, S12
282	A2	217244134_Ankur050-A2S-55-2_S50	A2S-55-2	**TAS**	55	FLD0146	GACGATCGCA	A	2.5	20506	20398	18092	2.34	0.11	7000	20, S12
283	A2	217241177_Ankur051-A2S-55-3_S51	A2S-55-3	**TAS**	55	FLD0147	TCATCATGCG	A	2.5	19934	19848	17601	2.36	0.03	7000	20, S12
284	A2	217235205_Ankur052-A2S-55-4_S52	A2S-55-4	**TAS**	55	FLD0148	ACATGTCTGA	A	2.5	21140	21011	18679	2.35	0.14	7000	20, S12
285	A2	217242147_Ankur053-A2S-55-5_S53	A2S-55-5	**TAS**	55	FLD0149	AGTCATCGCA	A	2.5	22327	22210	19751	2.35	0.09	7000	20, S12
286	A2	217235206_Ankur054-A2S-55-6_S54	A2S-55-6	**TAS**	55	FLD0150	TAGCATACAG	A	2.5	16712	16616	14385	2.35	0.03	7000	20, S12
287	A2	217228294_Ankur055-A2S-55-7_S55	A2S-55-7	**TAS**	55	FLD0151	AGAGTCGCGT	A	2.5	20917	20823	18413	2.36	0.03	7000	20, S12
288	A2	217240178_Ankur056-A2S-55-8_S56	A2S-55-8	**TAS**	55	FLD0152	TCTACGACAT	A	2.5	19382	19294	17195	2.34	0.14	7000	20, S12

#	Exp	SRA Sample Name	SampleID	PCR Method	Annealing Temp	Barcode Name	Barcode Sequence	Input DNA Template	gDNA Conc (ng/ul)	Raw Reads	Merged Reads	Trimmed	Shannon	Ideal Score	Rarefaction	Ref. Figure
289	A3	217247032_Ankur017-A3S-45-1_S17	A3S-45-1	TAS	45	FLD0113	CTCAGCAGTG	A	2.5	20614	20508	18161	2.28	0.03	7000	20, S13
290	A3	217232226_Ankur018-A3S-45-2_S18	A3S-45-2	TAS	45	FLD0114	CAGTCTACAT	A	2.5	18757	18658	16689	2.29	0.06	7000	20, S13
291	A3	217232227_Ankur019-A3S-45-3_S19	A3S-45-3	TAS	45	FLD0115	TACTGCAGCG	A	2.5	19335	19252	17121	2.29	0.06	7000	20, S13
292	A3	217245122_Ankur020-A3S-45-4_S20	A3S-45-4	TAS	45	FLD0116	TACACAGTAG	A	2.5	17537	17451	15449	2.29	0.06	7000	20, S13
293	A3	217246053_Ankur021-A3S-45-5_S21	A3S-45-5	TAS	45	FLD0117	CACATACAGT	A	2.5	20631	20490	18209	2.27	0.03	7000	20, S13
294	A3	217238174_Ankur022-A3S-45-6_S22	A3S-45-6	TAS	45	FLD0118	CACAGTGATG	A	2.5	16833	16741	14791	2.28	0.00	7000	20, S13
295	A3	217243140_Ankur023-A3S-45-7_S23	A3S-45-7	TAS	45	FLD0119	CGAGCTAGCA	A	2.5	17134	17036	15120	2.29	0.09	7000	20, S13
296	A3	217241173_Ankur024-A3S-45-8_S24	A3S-45-8	TAS	45	FLD0120	GAGACTATGC	A	2.5	16176	16115	14231	2.29	0.09	7000	20, S13
297	A3	217232233_Ankur057-A3S-55-1_S57	A3S-55-1	TAS	55	FLD0153	CACGAGATGA	A	2.5	20339	20226	17853	2.29	0.00	7000	20, S13
298	A3	217228295_Ankur058-A3S-55-2_S58	A3S-55-2	TAS	55	FLD0154	ACGCACATAT	A	2.5	20961	20823	18510	2.29	0.06	7000	20, S13
299	A3	217237221_Ankur059-A3S-55-3_S59	A3S-55-3	TAS	55	FLD0155	ACGTGCTCTG	A	2.5	19456	19325	17212	2.29	0.00	7000	20, S13
300	A3	217237223_Ankur060-A3S-55-4_S60	A3S-55-4	TAS	55	FLD0156	ACGATCACAT	A	2.5	21133	21016	18691	2.31	0.29	7000	20, S13
301	A3	217231271_Ankur061-A3S-55-5_S61	A3S-55-5	TAS	55	FLD0157	AGTGTACTCA	A	2.5	22405	22282	19881	2.31	0.17	7000	20, S13
302	A3	217245133_Ankur062-A3S-55-6_S62	A3S-55-6	TAS	55	FLD0158	TGATGTATGT	A	2.5	18228	18147	16181	2.27	0.11	7000	20, S13
303	A3	217231272_Ankur063-A3S-55-7_S63	A3S-55-7	TAS	55	FLD0159	GATATATGTC	A	2.5	20822	20719	18577	2.29	0.09	7000	20, S13
304	A3	217238184_Ankur064-A3S-55-8_S64	A3S-55-8	TAS	55	FLD0160	TAGTACTAGA	A	2.5	18107	18029	16175	2.28	0.09	7000	20, S13
305	A4	217240168_Ankur025-A4S-45-1_S25	A4S-45-1	TAS	45	FLD0121	CAGAGCTAGT	A	2.5	21094	21005	18490	3.38	0.20	7000	20, S14
306	A4	217230267_Ankur026-A4S-45-2_S26	A4S-45-2	TAS	45	FLD0122	CGCAGAGCAT	A	2.5	20385	20277	17633	3.39	0.20	7000	20, S14
307	A4	217244125_Ankur027-A4S-45-3_S27	A4S-45-3	TAS	45	FLD0123	TGTACAGCGA	A	2.5	18246	18154	15859	3.36	0.03	7000	20, S14
308	A4	217244126_Ankur028-A4S-45-4_S28	A4S-45-4	TAS	45	FLD0124	ACGTCAGTAT	A	2.5	20656	20526	18114	3.38	0.09	7000	20, S14
309	A4	217246056_Ankur029-A4S-45-5_S29	A4S-45-5	TAS	45	FLD0125	TCACAGCATA	A	2.5	19735	19627	17225	3.37	0.14	7000	20, S14
310	A4	217237211_Ankur030-A4S-45-6_S30	A4S-45-6	TAS	45	FLD0126	ACTGCGTGTC	A	2.5	19061	18950	16618	3.37	0.06	7000	20, S14
311	A4	217242140_Ankur031-A4S-45-7_S31	A4S-45-7	TAS	45	FLD0127	CGATCGACTG	A	2.5	17785	17693	15600	3.35	0.03	7000	20, S14
312	A4	217239205_Ankur032-A4S-45-8_S32	A4S-45-8	TAS	45	FLD0128	GCGAGATGTA	A	2.5	16474	16393	14035	3.37	0.00	7000	20, S14
313	A4	217235212_Ankur065-A4S-55-1_S65	A4S-55-1	TAS	55	FLD0161	TATAGAGATC	A	2.5	19649	19582	17304	3.37	0.09	7000	20, S14
314	A4	217229270_Ankur066-A4S-55-2_S66	A4S-55-2	TAS	55	FLD0162	TCGATATCTA	A	2.5	18843	18765	16413	3.36	0.00	7000	20, S14
315	A4	217239213_Ankur067-A4S-55-3_S67	A4S-55-3	TAS	55	FLD0163	TACATGATAG	A	2.5	20400	20305	17918	3.37	0.11	7000	20, S14
316	A4	217245136_Ankur068-A4S-55-4_S68	A4S-55-4	TAS	55	FLD0164	TGAGATCATA	A	2.5	19875	19804	17546	3.37	0.09	7000	20, S14
317	A4	217235213_Ankur069-A4S-55-5_S69	A4S-55-5	TAS	55	FLD0165	CTACATACTA	A	2.5	17854	17766	15737	3.37	0.06	7000	20, S14
318	A4	217232237_Ankur070-A4S-55-6_S70	A4S-55-6	TAS	55	FLD0166	ATCAGTGTAT	A	2.5	20560	20455	18125	3.37	0.23	7000	20, S14
319	A4	217246070_Ankur071-A4S-55-7_S71	A4S-55-7	TAS	55	FLD0167	ATCATATCTC	A	2.5	19577	19480	17216	3.37	0.11	7000	20, S14
320	A4	217243151_Ankur072-A4S-55-8_S72	A4S-55-8	TAS	55	FLD0168	AGTAGATCAT	A	2.5	21133	21029	18644	3.36	0.17	7000	20, S14
321	A6	217231264_Ankur033-A6S-45-1_S33	A6S-45-1	TAS	45	FLD0129	CTGATGCAGA	A	2.5	19692	19594	17085	4.04	0.03	7000	20, S15
322	A6	217244130_Ankur034-A6S-45-2_S34	A6S-45-2	TAS	45	FLD0130	GTGACGTACG	A	2.5	18240	18152	15737	4.04	0.03	7000	20, S15
323	A6	217246062_Ankur035-A6S-45-3_S35	A6S-45-3	TAS	45	FLD0131	CGACGCTGAT	A	2.5	20266	20166	17728	4.05	0.03	7000	20, S15
324	A6	217232228_Ankur036-A6S-45-4_S36	A6S-45-4	TAS	45	FLD0132	CTACGATCAG	A	2.5	19419	19320	16974	4.04	0.06	7000	20, S15
325	A6	217229262_Ankur037-A6S-45-5_S37	A6S-45-5	TAS	45	FLD0133	GCACTAGACA	A	2.5	22448	22333	19724	4.03	0.03	7000	20, S15
326	A6	217243143_Ankur038-A6S-45-6_S38	A6S-45-6	TAS	45	FLD0134	CTAGCAGATG	A	2.5	20583	20490	17924	4.04	0.03	7000	20, S15
327	A6	217245127_Ankur039-A6S-45-7_S39	A6S-45-7	TAS	45	FLD0135	CATGATACGC	A	2.5	18434	18346	15717	4.03	0.06	7000	20, S15
328	A6	217246064_Ankur040-A6S-45-8_S40	A6S-45-8	TAS	45	FLD0136	GCAGCTGTCA	A	2.5	18760	18664	16357	4.04	0.03	7000	20, S15
329	A6	217229274_Ankur073-A6S-55-1_S73	A6S-55-1	TAS	55	FLD0169	ACATAGTATC	A	2.5	22380	22259	19747	4.01	0.29	7000	20, S15
330	A6	217237227_Ankur074-A6S-55-2_S74	A6S-55-2	TAS	55	FLD0170	ATGTATAGTC	A	2.5	21187	21084	18751	4.02	0.09	7000	20, S15
331	A6	217249037_Ankur075-A6S-55-3_S75	A6S-55-3	TAS	55	FLD0171	ACAGTCATAT	A	2.5	21662	21546	19066	4.02	0.00	7000	20, S15
332	A6	217246073_Ankur076-A6S-55-4_S76	A6S-55-4	TAS	55	FLD0172	ACATATACGT	A	2.5	20973	20838	18406	4.03	0.00	7000	20, S15
333	A6	217247053_Ankur077-A6S-55-5_S77	A6S-55-5	TAS	55	FLD0173	AGCATCTATA	A	2.5	20555	20458	18057	4.02	0.09	7000	20, S15
334	A6	217232242_Ankur078-A6S-55-6_S78	A6S-55-6	TAS	55	FLD0174	AGACTATATC	A	2.5	20963	20862	18480	4.02	0.06	7000	20, S15
335	A6	217229275_Ankur079-A6S-55-7_S79	A6S-55-7	TAS	55	FLD0175	CAGCATCTAG	A	2.5	17446	17357	15319	4.02	0.06	7000	20, S15
336	A6	217239216_Ankur080-A6S-55-8_S80	A6S-55-8	TAS	55	FLD0176	CGAGACGACA	A	2.5	16179	16090	14031	4.03	0.00	7000	20, S15

#	Exp	SRA Sample Name	SampleID	PCR Method	Annealing Temp	Barcode Name	Barcode Sequence	Input DNA Template	gDNA Conc (ng/ul)	Raw Reads	Merged Reads	Trimmed	Shannon	Ideal Score	Rarefaction	Ref. Figure
337	B1	217236223_Ankur081-B1S-45-1_S99	B1S-45-1	TAS	45	FLD0193	TTGTTGCTGT	B	2.5	31891	31751	28514	0.06	45.00	7000	21, 22, 23, S16
338	B1	217244143_Ankur082-B1S-45-2_S100	B1S-45-2	TAS	45	FLD0194	GTGTGGTTGT	B	2.5	27795	27644	24885	0.05	46.71	7000	21, 22, 23, S16
339	B1	217244144_Ankur083-B1S-45-3_S101	B1S-45-3	TAS	45	FLD0195	TAGGTGGAAT	B	2.5	30980	30836	27749	0.05	44.63	7000	21, 22, 23, S16
340	B1	217240186_Ankur084-B1S-45-4_S102	B1S-45-4	TAS	45	FLD0196	TGTAGGTGGA	B	2.5	28957	28836	25875	0.07	44.74	7000	21, 22, 23, S16
341	B1	217237232_Ankur085-B1S-45-5_S103	B1S-45-5	TAS	45	FLD0197	TTAGTGGTGA	B	2.5	27494	27373	24519	0.07	45.40	7000	21, 22, 23, S16
342	B1	217246076_Ankur086-B1S-45-6_S104	B1S-45-6	TAS	45	FLD0198	GTGAAGGTAA	B	2.5	27184	27070	24282	0.08	46.83	7000	21, 22, 23, S16
343	B1	217231276_Ankur087-B1S-45-7_S105	B1S-45-7	TAS	45	FLD0199	TGTTGTGGTA	B	2.5	27739	27607	24836	0.05	46.09	7000	21, 22, 23, S16
344	B1	217231277_Ankur088-B1S-45-8_S106	B1S-45-8	TAS	45	FLD0200	GTTGATGAGT	B	2.5	32569	32431	29359	0.06	45.60	7000	21, 22, 23, S16
345	B1	217245168_Ankur169-B1S-55-1_S187	B1S-55-1	TAS	55	FLD0289	CACTGCTTGA	B	2.5	28100	27979	25225	0.06	54.57	7000	21, 22, 23, S16
346	B1	217238210_Ankur170-B1S-55-2_S188	B1S-55-2	TAS	55	FLD0290	TCTAGCGTGG	B	2.5	30343	30201	26900	0.05	55.31	7000	21, 22, 23, S16
347	B1	217232270_Ankur171-B1S-55-3_S189	B1S-55-3	TAS	55	FLD0291	GCATAATCGC	B	2.5	30101	29968	26800	0.06	56.00	7000	21, 22, 23, S16
348	B1	217239243_Ankur172-B1S-55-4_S190	B1S-55-4	TAS	55	FLD0292	GTCGTAACAC	B	2.5	35040	34905	31510	0.07	56.40	7000	21, 22, 23, S16
349	B1	217245169_Ankur173-B1S-55-5_S191	B1S-55-5	TAS	55	FLD0293	GAGATTGCTA	B	2.5	31919	31784	28519	0.06	57.29	7000	21, 22, 23, S16
350	B1	217245170_Ankur174-B1S-55-6_S192	B1S-55-6	TAS	55	FLD0294	GGACAGATGG	B	2.5	30967	30852	27817	0.05	55.20	7000	21, 22, 23, S16
351	B1	217242176_Ankur175-B1S-55-7_S193	B1S-55-7	TAS	55	FLD0295	CTTACGTTGC	B	2.5	27753	27637	24684	0.08	57.80	7000	21, 22, 23, S16
352	B1	217249068_Ankur176-B1S-55-8_S194	B1S-55-8	TAS	55	FLD0296	GTGTTCGGTC	B	2.5	27031	26914	24068	0.06	55.17	7000	21, 22, 23, S16
353	B2	217229278_Ankur089-B2S-45-1_S107	B2S-45-1	TAS	45	FLD0201	GGTCAGTGTA	B	2.5	29476	29330	26283	2.33	45.83	7000	22, 23, S17
354	B2	217244147_Ankur090-B2S-45-2_S108	B2S-45-2	TAS	45	FLD0202	GTAATGGAGT	B	2.5	26804	26643	23933	2.33	44.37	7000	22, 23, S17
355	B2	217236229_Ankur091-B2S-45-3_S109	B2S-45-3	TAS	45	FLD0203	CTCGTTATTC	B	2.5	25757	25632	22960	2.34	44.94	7000	22, 23, S17
356	B2	217241185_Ankur092-B2S-45-4_S110	B2S-45-4	TAS	45	FLD0204	GGAAGTAAGG	B	2.5	27800	27654	24753	2.34	45.54	7000	22, 23, S17
357	B2	217229280_Ankur093-B2S-45-5_S111	B2S-45-5	TAS	45	FLD0205	CGGTGTGTGT	B	2.5	27912	27762	24656	2.32	45.89	7000	22, 23, S17
358	B2	217244148_Ankur094-B2S-45-6_S112	B2S-45-6	TAS	45	FLD0206	CGTCTTCTTA	B	2.5	27102	26996	24272	2.33	47.26	7000	22, 23, S17
359	B2	217231282_Ankur095-B2S-45-7_S113	B2S-45-7	TAS	45	FLD0207	TGTGAATCTC	B	2.5	26913	26804	24166	2.33	45.37	7000	22, 23, S17
360	B2	217235219_Ankur096-B2S-45-8_S114	B2S-45-8	TAS	45	FLD0208	CTAATCGTGT	B	2.5	28337	28220	25136	2.33	44.60	7000	22, 23, S17
361	B2	217232272_Ankur177-B2S-55-1_S195	B2S-55-1	TAS	55	FLD0297	CTCAAGAAGC	B	2.5	23860	23739	21067	2.35	28.46	7000	22, 23, S17
362	B2	217231297_Ankur178-B2S-55-2_S196	B2S-55-2	TAS	55	FLD0298	TCTCGGATAG	B	2.5	24257	24129	21549	2.35	31.46	7000	22, 23, S17
363	B2	217243179_Ankur179-B2S-55-3_S197	B2S-55-3	TAS	55	FLD0299	CTCTGGACGA	B	2.5	25987	25878	23044	2.36	32.83	7000	22, 23, S17
364	B2	217245172_Ankur180-B2S-55-4_S198	B2S-55-4	TAS	55	FLD0300	CGAGCATTGT	B	2.5	29077	28959	25847	2.34	30.66	7000	22, 23, S17
365	B2	217246098_Ankur181-B2S-55-5_S199	B2S-55-5	TAS	55	FLD0301	CCAAGAAGAA	B	2.5	28880	28710	25561	2.33	34.34	7000	22, 23, S17
366	B2	217251042_Ankur182-B2S-55-6_S200	B2S-55-6	TAS	55	FLD0302	TCCTTGTTCT	B	2.5	26160	26056	23530	2.35	32.14	7000	22, 23, S17
367	B2	217247085_Ankur183-B2S-55-7_S201	B2S-55-7	TAS	55	FLD0303	GTAACGATGT	B	2.5	27869	27754	24377	2.33	31.40	7000	22, 23, S17
368	B2	217247086_Ankur184-B2S-55-8_S202	B2S-55-8	TAS	55	FLD0304	TGGACTCAGA	B	2.5	22989	22914	20435	2.33	32.57	7000	22, 23, S17
369	B3	217241187_Ankur097-B3S-45-1_S115	B3S-45-1	TAS	45	FLD0209	CTCTTAGTTC	B	2.5	25897	25766	23054	2.27	44.23	7000	22, 23, S18
370	B3	217247058_Ankur098-B3S-45-2_S116	B3S-45-2	TAS	45	FLD0210	GGATAGGATC	B	2.5	25898	25746	22957	2.27	45.40	7000	22, 23, S18
371	B3	217243159_Ankur099-B3S-45-3_S117	B3S-45-3	TAS	45	FLD0211	GGTGTCTTGT	B	2.5	26495	26361	23510	2.29	46.89	7000	22, 23, S18
372	B3	217241188_Ankur100-B3S-45-4_S118	B3S-45-4	TAS	45	FLD0212	GATGGTTGTA	B	2.5	24558	24466	21916	2.30	44.06	7000	22, 23, S18
373	B3	217249044_Ankur101-B3S-45-5_S119	B3S-45-5	TAS	45	FLD0213	CCTCGTTGTT	B	2.5	25462	25333	22774	2.28	44.23	7000	22, 23, S18
374	B3	217249045_Ankur102-B3S-45-6_S120	B3S-45-6	TAS	45	FLD0214	GGTTGGAGTT	B	2.5	30916	30748	27392	2.27	45.23	7000	22, 23, S18
375	B3	217241189_Ankur103-B3S-45-7_S121	B3S-45-7	TAS	45	FLD0215	TGGTGTCCGT	B	2.5	23603	23507	20818	2.27	46.86	7000	22, 23, S18
376	B3	217232254_Ankur104-B3S-45-8_S122	B3S-45-8	TAS	45	FLD0216	CGTTAGCGTA	B	2.5	25558	25444	22713	2.27	45.03	7000	22, 23, S18
377	B3	217236243_Ankur185-B3S-55-1_S203	B3S-55-1	TAS	55	FLD0305	GGCATCATGC	B	2.5	22733	22623	19783	2.27	31.89	7000	22, 23, S18
378	B3	217240211_Ankur186-B3S-55-2_S204	B3S-55-2	TAS	55	FLD0306	GTATAACGCT	B	2.5	25509	25401	22881	2.29	31.51	7000	22, 23, S18
379	B3	217245174_Ankur187-B3S-55-3_S205	B3S-55-3	TAS	55	FLD0307	GCAGATAAGT	B	2.5	27262	27129	24222	2.29	30.49	7000	22, 23, S18
380	B3	217236246_Ankur188-B3S-55-4_S206	B3S-55-4	TAS	55	FLD0308	GTCGGCTCTA	B	2.5	23791	23520	20975	2.27	32.03	7000	22, 23, S18
381	B3	217244176_Ankur189-B3S-55-5_S207	B3S-55-5	TAS	55	FLD0309	TTCGATAGCA	B	2.5	21434	21328	18975	2.28	32.80	7000	22, 23, S18
382	B3	217238216_Ankur190-B3S-55-6_S208	B3S-55-6	TAS	55	FLD0310	GTCTAGCAGG	B	2.5	24593	24465	21648	2.29	30.97	7000	22, 23, S18
383	B3	217241210_Ankur191-B3S-55-7_S209	B3S-55-7	TAS	55	FLD0311	GGAACACAGG	B	2.5	21223	21120	18778	2.28	32.20	7000	22, 23, S18
384	B3	217243182_Ankur192-B3S-55-8_S210	B3S-55-8	TAS	55	FLD0312	TGGTTCGCTG	B	2.5	29956	29832	26542	2.31	30.29	7000	22, 23, S18

#	Exp	SRA Sample Name	SampleID	PCR Method	Annealing Temp	Barcode Name	Barcode Sequence	Input DNA Template	gDNA Conc (ng/ul)	Raw Reads	Merged Reads	Trimmed	Shannon	Ideal Score	Rarefaction	Ref. Figure
385	B4	217242157_Ankur105-B4S-45-1_S123	B4S-45-1	TAS	45	FLD0217	TACTAGGATC	B	2.5	26589	26460	23466	3.36	42.71	7000	22, 23, S19
386	B4	217239224_Ankur106-B4S-45-2_S124	B4S-45-2	TAS	45	FLD0218	GTCTCAATGT	B	2.5	24979	24870	22194	3.35	43.91	7000	22, 23, S19
387	B4	217245150_Ankur107-B4S-45-3_S125	B4S-45-3	TAS	45	FLD0219	GATGAGGTAT	B	2.5	25966	25831	22933	3.34	44.69	7000	22, 23, S19
388	B4	217237244_Ankur108-B4S-45-4_S126	B4S-45-4	TAS	45	FLD0220	GGTGTTAGTG	B	2.5	25978	25853	22975	3.33	40.37	7000	22, 23, S19
389	B4	217232256_Ankur109-B4S-45-5_S127	B4S-45-5	TAS	45	FLD0221	CATTCTCTGA	B	2.5	26578	26437	23353	3.34	40.26	7000	22, 23, S19
390	B4	217249051_Ankur110-B4S-45-6_S128	B4S-45-6	TAS	45	FLD0222	CATCTGGAGT	B	2.5	29045	28914	25764	3.36	40.94	7000	22, 23, S19
391	B4	217238194_Ankur111-B4S-45-7_S129	B4S-45-7	TAS	45	FLD0223	GAATGGAAGA	B	2.5	29743	29637	26236	3.34	42.60	7000	22, 23, S19
392	B4	217232257_Ankur112-B4S-45-8_S130	B4S-45-8	TAS	45	FLD0224	GGCTGTGATC	B	2.5	30838	30716	27175	3.34	41.60	7000	22, 23, S19
393	B4	217237273_Ankur193-B4S-55-1_S211	B4S-55-1	TAS	55	FLD0313	CACATTAGCG	B	2.5	23953	23852	21019	3.36	34.57	7000	22, 23, S19
394	B4	217247088_Ankur194-B4S-55-2_S212	B4S-55-2	TAS	55	FLD0314	GAAGCGCACT	B	2.5	30814	30705	27120	3.34	35.40	7000	22, 23, S19
395	B4	217232281_Ankur195-B4S-55-3_S213	B4S-55-3	TAS	55	FLD0315	GCATGCCAGT	B	2.5	25699	25568	22541	3.35	31.49	7000	22, 23, S19
396	B4	217243184_Ankur196-B4S-55-4_S214	B4S-55-4	TAS	55	FLD0316	GGAGACTGTA	B	2.5	25432	25318	22530	3.35	35.06	7000	22, 23, S19
397	B4	217231305_Ankur197-B4S-55-5_S215	B4S-55-5	TAS	55	FLD0317	TCGAACTGCA	B	2.5	23279	23155	20574	3.37	35.14	7000	22, 23, S19
398	B4	217237274_Ankur198-B4S-55-6_S216	B4S-55-6	TAS	55	FLD0318	GAGAGGACAT	B	2.5	26815	26708	23652	3.36	34.09	7000	22, 23, S19
399	B4	217231306_Ankur199-B4S-55-7_S217	B4S-55-7	TAS	55	FLD0319	GAGCACGGAA	B	2.5	24421	24324	21438	3.34	32.43	7000	22, 23, S19
400	B4	217249076_Ankur200-B4S-55-8_S218	B4S-55-8	TAS	55	FLD0320	GCTCTAACAT	B	2.5	25918	25798	22963	3.38	34.86	7000	22, 23, S19
401	C1	217228312_Ankur113-C1S-45-1_S131	C1S-45-1	TAS	45	FLD0225	TGGTGCTGGA	C	2.5	30543	30393	27045	0.06	47.54	7000	21, S20
402	C1	217232259_Ankur114-C1S-45-2_S132	C1S-45-2	TAS	45	FLD0226	TATGGTAAGG	C	2.5	30375	30255	27178	0.06	48.94	7000	21, S20
403	C1	217229289_Ankur115-C1S-45-3_S133	C1S-45-3	TAS	45	FLD0227	GTTCGATTGT	C	2.5	29675	29536	26741	0.06	49.54	7000	21, S20
404	C1	217247063_Ankur116-C1S-45-4_S134	C1S-45-4	TAS	45	FLD0228	GGTAGAATGA	C	2.5	30298	30190	26996	0.06	47.59	7000	21, S20
405	C1	217243166_Ankur117-C1S-45-5_S135	C1S-45-5	TAS	45	FLD0229	TTCTCATCGT	C	2.5	28073	27970	25263	0.06	48.77	7000	21, S20
406	C1	217237248_Ankur118-C1S-45-6_S136	C1S-45-6	TAS	45	FLD0230	CTCAATCGTA	C	2.5	30807	30661	27683	0.05	49.48	7000	21, S20
407	C1	217243168_Ankur119-C1S-45-7_S137	C1S-45-7	TAS	45	FLD0231	CGCTAATGTA	C	2.5	26845	26725	23981	0.07	48.91	7000	21, S20
408	C1	217247065_Ankur120-C1S-45-8_S138	C1S-45-8	TAS	45	FLD0232	GCGTCTGAAT	C	2.5	30448	30324	27446	0.06	49.02	7000	21, S20
409	C1	217232283_Ankur201-C1S-55-1_S219	C1S-55-1	TAS	55	FLD0321	TGCTGGCTTG	C	2.5	26555	26444	23803	0.07	60.11	7000	21, S20
410	C1	217251051_Ankur202-C1S-55-2_S220	C1S-55-2	TAS	55	FLD0322	TGCATGGAGC	C	2.5	22741	22626	20063	0.07	61.12	7000	21, S20
411	C1	217240213_Ankur203-C1S-55-3_S221	C1S-55-3	TAS	55	FLD0323	GTACTAAGAG	C	2.5	30520	30395	27449	0.06	58.43	7000	21, S20
412	C1	217244179_Ankur204-C1S-55-4_S222	C1S-55-4	TAS	55	FLD0324	GAAGTCAAGC	C	2.5	23311	23212	20698	0.06	59.03	7000	21, S20
413	C1	217239248_Ankur205-C1S-55-5_S223	C1S-55-5	TAS	55	FLD0325	GCGCATTATG	C	2.5	26913	26785	24182	0.07	60.38	7000	21, S20
414	C1	217238219_Ankur206-C1S-55-6_S224	C1S-55-6	TAS	55	FLD0326	GTCCAGACAT	C	2.5	33716	33576	30395	0.05	60.14	7000	21, S20
415	C1	217250048_Ankur207-C1S-55-7_S225	C1S-55-7	TAS	55	FLD0327	GAGACCTCTA	C	2.5	29354	29261	26378	0.06	60.20	7000	21, S20
416	C1	217245179_Ankur208-C1S-55-8_S226	C1S-55-8	TAS	55	FLD0328	TTGCACTCAG	C	2.5	26981	26897	24319	0.06	59.85	7000	21, S20
417	C2	217249056_Ankur121-C2S-45-1_S139	C2S-45-1	TAS	45	FLD0233	TTCTGTTGCC	C	2.5	28720	28581	25591	2.34	47.80	7000	S21
418	C2	217240196_Ankur122-C2S-45-2_S140	C2S-45-2	TAS	45	FLD0234	TTGTCCTTGC	C	2.5	31796	31664	28129	2.34	48.83	7000	S21
419	C2	217239226_Ankur123-C2S-45-3_S141	C2S-45-3	TAS	45	FLD0235	CCTGTGTAGA	C	2.5	31608	31449	27916	2.34	49.17	7000	S21
420	C2	217241194_Ankur124-C2S-45-4_S142	C2S-45-4	TAS	45	FLD0236	GATAAGAAGG	C	2.5	28726	28590	25569	2.33	49.72	7000	S21
421	C2	217238197_Ankur125-C2S-45-5_S143	C2S-45-5	TAS	45	FLD0237	CAGGTCACAT	C	2.5	27069	26947	24025	2.33	49.20	7000	S21
422	C2	217232261_Ankur126-C2S-45-6_S144	C2S-45-6	TAS	45	FLD0238	GCCATGTCAT	C	2.5	28695	28557	25552	2.34	48.68	7000	S21
423	C2	217242164_Ankur127-C2S-45-7_S145	C2S-45-7	TAS	45	FLD0239	TCTGCCTATA	C	2.5	25509	25415	22744	2.34	50.08	7000	S21
424	C2	217238198_Ankur128-C2S-45-8_S146	C2S-45-8	TAS	45	FLD0240	CTTAGTTCGC	C	2.5	27923	27778	24622	2.32	49.91	7000	S21
425	C2	217235256_Ankur209-C2S-55-1_S227	C2S-55-1	TAS	55	FLD0329	TGCGGCGATA	C	2.5	24127	24039	21125	2.34	34.88	7000	S21
426	C2	217235257_Ankur210-C2S-55-2_S228	C2S-55-2	TAS	55	FLD0330	AGTTGCTAGT	C	2.5	26456	26332	23513	2.35	32.28	7000	S21
427	C2	217243188_Ankur211-C2S-55-3_S229	C2S-55-3	TAS	55	FLD0331	AGGATTGAGG	C	2.5	30845	30687	27493	2.35	33.78	7000	S21
428	C2	217239251_Ankur212-C2S-55-4_S230	C2S-55-4	TAS	55	FLD0332	CCAGAACAGA	C	2.5	24465	24312	21548	2.34	35.20	7000	S21
429	C2	217243189_Ankur213-C2S-55-5_S231	C2S-55-5	TAS	55	FLD0333	CGTCAAGCAT	C	2.5	25619	25485	22649	2.34	33.97	7000	S21
430	C2	217244185_Ankur214-C2S-55-6_S232	C2S-55-6	TAS	55	FLD0334	TTGTCGAGAC	C	2.5	27223	27129	24353	2.34	32.71	7000	S21
431	C2	217239252_Ankur215-C2S-55-7_S233	C2S-55-7	TAS	55	FLD0335	GACAGGTGAC	C	2.5	25662	25560	22837	2.34	33.48	7000	S21
432	C2	217232285_Ankur216-C2S-55-8_S234	C2S-55-8	TAS	55	FLD0336	CTGACAAGTG	C	2.5	26914	26754	24055	2.34	34.03	7000	S21

#	Exp	SRA Sample Name	SampleID	PCR Method	Annealing Temp	Barcode Name	Barcode Sequence	Input DNA Template	gDNA Conc (ng/ul)	Raw Reads	Merged Reads	Trimmed	Shannon	Ideal Score	Rarefaction	Ref. Figure
433	C3	217244156_Ankur129-C3S-45-1_S147	C3S-45-1	TAS	45	FLD0241	CGTAATGAGC	C	2.5	31613	31450	27908	2.30	47.82	7000	S22
434	C3	217242165_Ankur130-C3S-45-2_S148	C3S-45-2	TAS	45	FLD0242	TTGCTTAGTC	C	2.5	27782	27671	24814	2.27	47.80	7000	S22
435	C3	217239229_Ankur131-C3S-45-3_S149	C3S-45-3	TAS	45	FLD0243	TCTTGTTCAC	C	2.5	30387	30250	27052	2.29	48.51	7000	S22
436	C3	217250034_Ankur132-C3S-45-4_S150	C3S-45-4	TAS	45	FLD0244	GTGGCTTCGT	C	2.5	26226	26072	23115	2.28	51.14	7000	S22
437	C3	217242167_Ankur133-C3S-45-5_S151	C3S-45-5	TAS	45	FLD0245	TGTTCGATAG	C	2.5	29212	29095	25973	2.29	49.08	7000	S22
438	C3	217236235_Ankur134-C3S-45-6_S152	C3S-45-6	TAS	45	FLD0246	TCATTCAGTG	C	2.5	26186	26050	23360	2.28	48.94	7000	S22
439	C3	217240198_Ankur135-C3S-45-7_S153	C3S-45-7	TAS	45	FLD0247	GTGGAGAGCT	C	2.5	26579	26444	23752	2.27	49.40	7000	S22
440	C3	217250036_Ankur136-C3S-45-8_S154	C3S-45-8	TAS	45	FLD0248	GTAGAAGTGG	C	2.5	28115	27978	24891	2.27	47.80	7000	S22
441	C3	217237284_Ankur217-C3S-55-1_S235	C3S-55-1	TAS	55	FLD0337	CACGAAGAGC	C	2.5	24438	24311	21358	2.29	31.80	7000	S22
442	C3	217235259_Ankur218-C3S-55-2_S236	C3S-55-2	TAS	55	FLD0338	CATACCTGAT	C	2.5	26552	26411	23605	2.28	33.45	7000	S22
443	C3	217239255_Ankur219-C3S-55-3_S237	C3S-55-3	TAS	55	FLD0339	GACGTGCTTC	C	2.5	20386	20268	17976	2.29	35.06	7000	S22
444	C3	217245182_Ankur220-C3S-55-4_S238	C3S-55-4	TAS	55	FLD0340	ATTGTGGAGT	C	2.5	26420	26262	23497	2.29	33.43	7000	S22
445	C3	217236251_Ankur221-C3S-55-5_S239	C3S-55-5	TAS	55	FLD0341	TCTGGTCTCA	C	2.5	21560	21445	19187	2.30	33.11	7000	S22
446	C3	217237288_Ankur222-C3S-55-6_S240	C3S-55-6	TAS	55	FLD0342	AGGTAAGAGG	C	2.5	29203	29059	25873	2.29	37.06	7000	S22
447	C3	217231312_Ankur223-C3S-55-7_S241	C3S-55-7	TAS	55	FLD0343	TCCTGACAGA	C	2.5	25320	25215	22386	2.29	34.94	7000	S22
448	C3	217231313_Ankur224-C3S-55-8_S242	C3S-55-8	TAS	55	FLD0344	GCACTGTTGC	C	2.5	24251	24138	21399	2.28	35.05	7000	S22
449	D1	217235232_Ankur137-D1S-45-1_S155	D1S-45-1	TAS	45	FLD0249	TGGAGCATGT	D	2.5	31201	31056	28216	0.06	66.31	7000	21, S23
450	D1	217247068_Ankur138-D1S-45-2_S156	D1S-45-2	TAS	45	FLD0250	GAAGGAGATA	D	2.5	29454	29336	26666	0.07	65.77	7000	21, S23
451	D1	217239232_Ankur139-D1S-45-3_S157	D1S-45-3	TAS	45	FLD0251	CGAATGTATG	D	2.5	33468	33330	30150	0.05	66.20	7000	21, S23
452	D1	217235235_Ankur140-D1S-45-4_S158	D1S-45-4	TAS	45	FLD0252	TCGTGAATGA	D	2.5	29655	29535	26771	0.07	65.49	7000	21, S23
453	D1	217242168_Ankur141-D1S-45-5_S159	D1S-45-5	TAS	45	FLD0253	GAATAGCTGA	D	2.5	32384	32261	29212	0.07	65.94	7000	21, S23
454	D1	217239233_Ankur142-D1S-45-6_S160	D1S-45-6	TAS	45	FLD0254	TTGTCACATC	D	2.5	29672	29559	26717	0.06	65.91	7000	21, S23
455	D1	217239234_Ankur143-D1S-45-7_S161	D1S-45-7	TAS	45	FLD0255	CTGGAGGCTA	D	2.5	27414	27272	24393	0.06	66.54	7000	21, S23
456	D1	217241199_Ankur144-D1S-45-8_S162	D1S-45-8	TAS	45	FLD0256	TGTCAGCTTA	D	2.5	30279	30173	27329	0.06	65.46	7000	21, S23
457	D1	217243192_Ankur225-D1S-55-1_S243	D1S-55-1	TAS	55	FLD0345	ACCATGAGTC	D	2.5	27267	27135	24618	0.06	71.31	7000	21, S23
458	D1	217240217_Ankur226-D1S-55-2_S244	D1S-55-2	TAS	55	FLD0346	AATGCAGTGT	D	2.5	30957	30809	27899	0.07	71.34	7000	21, S23
459	D1	217236252_Ankur227-D1S-55-3_S245	D1S-55-3	TAS	55	FLD0347	ATATGGTGGA	D	2.5	28039	27883	25078	0.06	70.89	7000	21, S23
460	D1	217244191_Ankur228-D1S-55-4_S246	D1S-55-4	TAS	55	FLD0348	ACTCAGTTAC	D	2.5	31942	31785	28713	0.07	70.57	7000	21, S23
461	D1	217245186_Ankur229-D1S-55-5_S247	D1S-55-5	TAS	55	FLD0349	AAGTGCGATG	D	2.5	24473	24320	21912	0.08	71.06	7000	21, S23
462	D1	217232290_Ankur230-D1S-55-6_S248	D1S-55-6	TAS	55	FLD0350	CCACAGAGTG	D	2.5	27490	27339	24772	0.06	70.91	7000	21, S23
463	D1	217243195_Ankur231-D1S-55-7_S249	D1S-55-7	TAS	55	FLD0351	AGTGGTGATC	D	2.5	26453	26322	23854	0.06	70.80	7000	21, S23
464	D1	217244192_Ankur232-D1S-55-8_S250	D1S-55-8	TAS	55	FLD0352	ACTTCTTAGC	D	2.5	26285	26161	23549	0.07	70.83	7000	21, S23
465	D2	217243174_Ankur145-D2S-45-1_S163	D2S-45-1	TAS	45	FLD0257	GTTCTTCGTA	D	2.5	30771	30609	27733	2.35	44.14	7000	S24
466	D2	217228321_Ankur146-D2S-45-2_S164	D2S-45-2	TAS	45	FLD0258	TTACACGTTC	D	2.5	29168	29054	26288	2.36	45.03	7000	S24
467	D2	217232265_Ankur147-D2S-45-3_S165	D2S-45-3	TAS	45	FLD0259	GTAGCCAGTA	D	2.5	26246	26141	23434	2.35	44.57	7000	S24
468	D2	217243175_Ankur148-D2S-45-4_S166	D2S-45-4	TAS	45	FLD0260	TGAGAAGGTA	D	2.5	31827	31693	28625	2.35	45.00	7000	S24
469	D2	217232266_Ankur149-D2S-45-5_S167	D2S-45-5	TAS	45	FLD0261	CCATATGATC	D	2.5	25754	25620	22911	2.35	45.71	7000	S24
470	D2	217244162_Ankur150-D2S-45-6_S168	D2S-45-6	TAS	45	FLD0262	CGATCCTATA	D	2.5	24648	24538	22063	2.36	45.20	7000	S24
471	D2	217242171_Ankur151-D2S-45-7_S169	D2S-45-7	TAS	45	FLD0263	TGACTAGCTT	D	2.5	24909	24800	22402	2.35	45.57	7000	S24
472	D2	217245164_Ankur152-D2S-45-8_S170	D2S-45-8	TAS	45	FLD0264	TAACTCTGCT	D	2.5	25293	25176	22744	2.35	46.11	7000	S24
473	D2	217242196_Ankur233-D2S-55-1_S251	D2S-55-1	TAS	55	FLD0353	GCCACATATA	D	2.5	25601	25491	22937	2.36	36.11	7000	S24
474	D2	217245187_Ankur234-D2S-55-2_S252	D2S-55-2	TAS	55	FLD0354	ACGCAGGAGT	D	2.5	29209	29014	26001	2.36	36.49	7000	S24
475	D2	217235263_Ankur235-D2S-55-3_S253	D2S-55-3	TAS	55	FLD0355	AATATGCTGC	D	2.5	20625	20502	18239	2.36	37.00	7000	S24
476	D2	217232292_Ankur236-D2S-55-4_S254	D2S-55-4	TAS	55	FLD0356	AAGCGTAGAA	D	2.5	6210	6189	5573	NA	NA	7000	Removed from analysis
477	D2	217251063_Ankur237-D2S-55-5_S255	D2S-55-5	TAS	55	FLD0357	GACAGCAAGC	D	2.5	21433	21335	18894	2.36	38.09	7000	S24
478	D2	217241219_Ankur238-D2S-55-6_S256	D2S-55-6	TAS	55	FLD0358	CTGACCGAGA	D	2.5	24865	24736	22073	2.37	36.34	7000	S24
479	D2	217240219_Ankur239-D2S-55-7_S257	D2S-55-7	TAS	55	FLD0359	CGCGACTTGT	D	2.5	27110	27000	24659	2.36	37.23	7000	S24
480	D2	217245190_Ankur240-D2S-55-8_S258	D2S-55-8	TAS	55	FLD0360	CATCAACATG	D	2.5	26631	26497	23933	2.35	34.77	7000	S24

#	Exp	SRA Sample Name	SampleID	PCR Method	Annealing Temp	Barcode Name	Barcode Sequence	Input DNA Template	gDNA Conc (ng/ul)	Raw Reads	Merged Reads	Trimmed	Shannon	Ideal Score	Rarefaction	Ref. Figure
481	E1	217249065_Ankur153-E1S-45-1_S171	E1S-45-1	TAS	45	FLD0265	TCGAATGTGC	E	2.5	26996	26830	23882	0.05	20.90	7000	21, S25
482	E1	217247075_Ankur154-E1S-45-2_S172	E1S-45-2	TAS	45	FLD0266	TCGCTGAACA	E	2.5	27815	27685	24954	0.06	20.07	7000	21, S25
483	E1	217244165_Ankur155-E1S-45-3_S173	E1S-45-3	TAS	45	FLD0267	GCGTTATTGC	E	2.5	30288	30132	26985	0.07	20.32	7000	21, S25
484	E1	217235239_Ankur156-E1S-45-4_S174	E1S-45-4	TAS	45	FLD0268	GAACTATCAC	E	2.5	28376	28269	25602	0.06	18.98	7000	21, S25
485	E1	217237256_Ankur157-E1S-45-5_S175	E1S-45-5	TAS	45	FLD0269	TCGAGGTACT	E	2.5	30608	30478	27475	0.06	20.07	7000	21, S25
486	E1	217244166_Ankur158-E1S-45-6_S176	E1S-45-6	TAS	45	FLD0270	TGCGGATGGT	E	2.5	28866	28702	25789	0.06	22.55	7000	21, S25
487	E1	217244167_Ankur159-E1S-45-7_S177	E1S-45-7	TAS	45	FLD0271	TTCGAGCTAT	E	2.5	24171	24063	21674	0.06	19.61	7000	21, S25
488	E1	217247080_Ankur160-E1S-45-8_S178	E1S-45-8	TAS	45	FLD0272	GGTCTGGTGT	E	2.5	25606	25469	22985	0.08	19.72	7000	21, S25
489	E1	217251065_Ankur241-E1S-55-1_S259	E1S-55-1	TAS	55	FLD0361	TGGCTACGCT	E	2.5	23840	23758	21124	0.06	20.15	7000	21, S25
490	E1	217249088_Ankur242-E1S-55-2_S260	E1S-55-2	TAS	55	FLD0362	ACGCGGACTA	E	2.5	33675	33455	29973	0.09	20.90	7000	21, S25
491	E1	217246112_Ankur243-E1S-55-3_S261	E1S-55-3	TAS	55	FLD0363	AGAGGTCGGA	E	2.5	28128	27990	24862	0.07	20.87	7000	21, S25
492	E1	217244195_Ankur244-E1S-55-4_S262	E1S-55-4	TAS	55	FLD0364	AATCGAGCGT	E	2.5	29128	28969	25990	0.07	20.14	7000	21, S25
493	E1	217250053_Ankur245-E1S-55-5_S263	E1S-55-5	TAS	55	FLD0365	AAGTACACTC	E	2.5	27961	27830	25169	0.07	20.21	7000	21, S25
494	E1	217245192_Ankur246-E1S-55-6_S264	E1S-55-6	TAS	55	FLD0366	AGCTGAATGA	E	2.5	28106	27970	25141	0.07	21.21	7000	21, S25
495	E1	217238232_Ankur247-E1S-55-7_S265	E1S-55-7	TAS	55	FLD0367	ATGCCTATCA	E	2.5	26632	26486	23883	0.05	21.58	7000	21, S25
496	E1	217238235_Ankur248-E1S-55-8_S266	E1S-55-8	TAS	55	FLD0368	ACTGTAGGAC	E	2.5	29200	29045	26132	0.07	21.21	7000	21, S25
497	E2	217251035_Ankur161-E2S-45-1_S179	E2S-45-1	TAS	45	FLD0273	CTAAGTCATG	E	2.5	24418	24278	21624	2.34	21.54	7000	S26
498	E2	217240205_Ankur162-E2S-45-2_S180	E2S-45-2	TAS	45	FLD0274	TTGCAGATCA	E	2.5	25648	25505	22880	2.33	22.11	7000	S26
499	E2	217251036_Ankur163-E2S-45-3_S181	E2S-45-3	TAS	45	FLD0275	CTGCGAATGT	E	2.5	27482	27324	24389	2.33	21.86	7000	S26
500	E2	217245166_Ankur164-E2S-45-4_S182	E2S-45-4	TAS	45	FLD0276	CTGTTCTAGC	E	2.5	21522	21398	18919	2.33	21.69	7000	S26
501	E2	217235243_Ankur165-E2S-45-5_S183	E2S-45-5	TAS	45	FLD0277	CACTTGTGTG	E	2.5	24946	24834	22080	2.34	23.11	7000	S26
502	E2	217249067_Ankur166-E2S-45-6_S184	E2S-45-6	TAS	45	FLD0278	TGGATGACAT	E	2.5	28964	28839	25810	2.33	21.77	7000	S26
503	E2	217247082_Ankur167-E2S-45-7_S185	E2S-45-7	TAS	45	FLD0279	GATCCTGAGC	E	2.5	25218	25107	22117	2.32	21.57	7000	S26
504	E2	217244170_Ankur168-E2S-45-8_S186	E2S-45-8	TAS	45	FLD0280	GTCGGTCTGA	E	2.5	27437	27297	24399	2.33	19.60	7000	S26
505	E2	217242204_Ankur249-E2S-55-1_S267	E2S-55-1	TAS	55	FLD0369	ATAGCCGTGT	E	2.5	24368	24235	21449	2.33	21.77	7000	S26
506	E2	217242205_Ankur250-E2S-55-2_S268	E2S-55-2	TAS	55	FLD0370	TCACGACGAA	E	2.5	23957	23826	21072	2.33	21.31	7000	S26
507	E2	217231323_Ankur251-E2S-55-3_S269	E2S-55-3	TAS	55	FLD0371	ATCTGTCCAT	E	2.5	25882	25764	22984	2.33	24.00	7000	S26
508	E2	217240222_Ankur252-E2S-55-4_S270	E2S-55-4	TAS	55	FLD0372	ACTTAGAGAG	E	2.5	28828	28656	25681	2.33	20.86	7000	S26
509	E2	217242207_Ankur253-E2S-55-5_S271	E2S-55-5	TAS	55	FLD0373	AGTGGCAGGT	E	2.5	30349	30189	26729	2.31	21.03	7000	S26
510	E2	217244197_Ankur254-E2S-55-6_S272	E2S-55-6	TAS	55	FLD0374	ATGAGGTCGT	E	2.5	28277	28126	25067	2.32	20.80	7000	S26
511	E2	217251068_Ankur255-E2S-55-7_S273	E2S-55-7	TAS	55	FLD0375	AGGAGAAGGA	E	2.5	29573	29425	26299	2.32	19.54	7000	S26
512	E2	217247097_Ankur256-E2S-55-8_S274	E2S-55-8	TAS	55	FLD0376	ACAACTGCAA	E	2.5	26922	26763	23625	2.35	21.46	7000	S26

Supplemental Material 1: Description of synthetic DNA template design and template sequences

>*Rhodanobacter denitrificans* strain 2APBS1 16S ribosomal RNA gene (NR_102497)
CTGGCTCAGATTGAACGCTGGCGGCATGCCTAACACATGCAAGTCGAACGGCAGCACAGCAGTAGCAATACTGTGGGTGGCGAGTGGCGGACGGGTGAGTAA
TGCATCGGGATCTACCCTGACGTGGGGGATAACCTCGGGAAACCGGGACTAATACCGCATACGTCCTACGGGAGAAAGCGGGGGACCTTCGGGCCTCGCGCG
GCAGGACGAACCGATGTGCGATTAGCTAGTTGGCGGGGTAATGGCCCACCAAGGCGACGATCGCTAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTG
AGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGT
TGTAAAGCACTTTTATCAGGAGCGAAATACCACGGGTTAATACCCTATGGGGCTGACGGTACCTGAGGAATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCG
GTAATACGAAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGGGTGCGTAGGCGGTTACTTAAGTCTGTCGTGAAATCCCCGGGCTCAACCTGGGAATG
GCGATGGATACTGGGTGGCTAGAGTGTGTCAGAGGATGGTGGAATTCCCGGTGTAGCGGTGAAATGCGTAGAGATCGGGAGGAACATCAGTGGCGAAGGCG
GCCATCTGGGACAACACTGACGCTGAAGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGCGAACTGGATGTTG
GTCTCAACTCGGAGATCAGTGTCGAAGCTAACGCGTTAAGTTCGCCGCCTGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACA
AGCGGTGGAGTATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTGGCCTTGACATGTCCGGAATCCTGCAGAGATGCGGGAGTGCCTTCGGGAATCGG
AACACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCCTTAGTTGCCAGCACGTAATGGTGG
GAACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCATCATGGCCCTTACGGCCAGGGCTACACACGTACTACAATGGTCG
GTACAGAGGGTTGCAATACCGCGAGGTGGAGCCAATCCCAGAAAGCCGATCCCAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGT
AATCGCGGATCAGCTATGCCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCTCCAGAAGGCGTTAGTCTAACCGC
AAGGGGGACGACGCCCACGGAGTGGTCCATGACTGGGGTGAAGTCGTAACAAGGTAGCCGTATCGGAAGGTGCGGCTGGATCACCTCCTTT

Standard 515F primer site: GTGCCAGCAGCCGCGGTAA [515F-"Parada": GTGYCAGCMGCCGCGGTAA]
Standard 806R primer site: ATTAGATACCCTGGTAGTCC [806R-"Apprill": GGACTACNVGGGTWTCTAAT]

>Inverse complement of *Rhodanobacter denitrificans* strain 2APBS1 16S ribosomal RNA gene (NR_102497)
AAAGGAGGTGATCCAGCCGCACCTTCCGATACGGCTACCTTGTTACGACTTCACCCCAGTCATGGACCACTCCGTGGGCGTCGTCCCCCTTGCGGTTAGACTAAC
GCCTTCTGGAGCAACCCACTCCCATGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCGGCATAGCTGATCCGCGATTACTAGCGATTCCGA
CTTCATGGAGTCGAGTTGCAGACTCCAATCCGGACTGGGATCGGCTTTCTGGGATTGGCTCCACCTCGCGGTATTGCAACCCTCTGTACCGACCATTGTAGTACG
TGTGTAGCCCTGGCCGTAAGGGCCATGATGACTTGACGTCATCCCCACCTTCCTCCGGTTTGTCACCGGCAGTCTCCTTAGAGTTCCCACCATTACGTGCTGGCAA
CTAAGGACAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAGCCATGCAGCACCTGTGTTCCGATTCCCGAAGGCACTCC
CGCATCTCTGCAGGATTCCGGACATGTCAAGGCCAGGTAAGGTTCTTCGCGTTGCATCGAATTAAACCACATACTCCACCGCTTGTGCGGGCCCCCGTCAATTCC
TTTGAGTTTC**AGTCTTGCGACCGTACTCCCCAGGCGGCGAACTTAACGCGTTAGCTTCGACACTGATCTCCGAGTTGAGACCAACATCCAGTTCGCATCGTTTA
GGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGTGCTTCAGCGTCAGTGTTGTCCCAGATGGCCGCCTTCGCCACTGATGTTCCTCCCG
ATCTCTACGCATTTCACCGCTACACCGGGAATTCCACCATCCTCTGACACACTCTAGCCACCCAGTATCCATCGCCATTCCCAGGTTGAGCCCGGGGATTTCACG
ACAGACTTAAGTAACCGCCTACGCACCCTTTACGCCCAGTAATTCCGATTAACGCTTGCACCCTTCGTATTACCGCGGCTGCTGGCACGAAGTTAGCCGGTGCT
TATTCCTCAGGTACCGTCAGCCCCATAGGGTATTAACCCGTGGT**ATTTCGCTCCTGATAAAAGTGCTTTACAACCCGAAGGCCTTCTTCACACACGCGGCATTGC
TGGATCAGGCTTGCGCCCATTGTCCAATATTCCCCACTGCTGCCTCCCGTAGGAGTCTGGGCCGTGTCTCAGTCCCAGTGTGGCTGATCATCCTCTCAGACCAGCT
AGCGATCGTCGCCTTGGTGGGCCATTACCCCGCCAACTAGCTAATCGCACATCGGTTCGTCCTGCCGCGCGAGGCCCGAAGGTCCCCCGCTTTCTCCCGTAGGAC
GTATGCGGTATTAGTCCCGGTTTCCCGAGGTTATCCCCCACGTCAGGGTAGATCCCGATGCATTACTCACCCGTCCGCCACTCGCCACCCACAGTATTGCTACTGC
TGTGCTGCCGTTCGACTTGCATGTGTTAGGCATGCCGCCAGCGTTCAATCTGAGCCAG

***Underlined and bold region indicates area of the gene used for synthetic gBLOCK synthesis, modified as shown below.**

Synthetic Template Design

AGTCTTGCGACCGTACTCCCCAGGCGGCGAACTTAACGCGTTAGCTTCGACACTGATCTCCGAGTTGAGACCAACATCCAGTTCGCATCGTTTAGGGCGTGGACT
ACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGTGCTTCAGCGTCAGTGTTGTCCCAGATGGCCGCCTTCGCCACTGATGTTCCTCCCGATCTCTACGCATT
TCACCGCTACACCGGGAATTCCACCATCCTCTGACACACTCTAGCCACCCAGTATCCATCGCCATTCCCAGGTTGAGCCCGGGGATTTCACGACAGACTTAAGTA
ACCGCCTACGCACCCTTTACGCCCAGTAATTCCGATTAACGCTTGCACCCTTCGTATTACCGCGGCTGCTGGCACGAAGTTAGCCGGTGCTTATTCCTCAGGTACC
GTCAGCCCCATAGGGTATTAACCCGTGGT

Step 1: Select region of *R. denitrificans* 2APBS 16S rRNA gene surrounding the standard EMP 515F/806R primer pair which generate a 292 bp amplicon with this template. Total length of fragment is 452 bp.

Step 2: Use the sequence for the 806R primer as the default "forward" primer (GGACTACCAGGGTATCTAAT). The standard 806R primer for the EMP is 24-fold degenerate. We use only the variant matching *R. denitrificans* 2APBS1 for the default primer sequence (Rh806Syn_1). The primer site is shown in light blue, above. In other templates that template 1, this primer site is altered.

Step 3: Identify a 'recognition' sequence in the DNA that will be varied from template to template. The same nucleotides are used to maintain GC content, but the sequence is scrambled so that many mismatches between the recognition sequences are present. The recognition sequence for template 1 is simply a 12 base region of the original DNA from *R. denitrificans* 2APBS1 (CGATCTCTACGC), and is highlighted in red.

Step 4: A new reverse primer site is developed to decrease the size of the amplicon to allow for better merging with 2x153 base sequencing on the Illumina MiniSeq. No degeneracies are used here, and all synthetic templates retain this sequence. Primer modifications are only performed at the '806R' primer site. The chosen primer design was: CGGAATTACTGGGCGTAAAGG (inverse complement = CCTTTACGCCCAGTAATTCCG), and the annealing location is highlighted in grey. The entire amplicon is 251 bp in size.

>ST1
AGTCTTGCGACCGTACTCCCCAGGCGGCGAACTTAACGCGTTAGCTTCGACACTGATCTCCGAGTTGAGACCAACATCCAGTTCGCATCGTTTAGGGCGTGGACT
ACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGTGCTTCAGCGTCAGTGTTGTCCCAGATGGCCGCCTTCGCCACTGATGTTCCTCCCGATCTCTACGCATT
TCACCGCTACACCGGGAATTCCACCATCCTCTGACACACTCTAGCCACCCAGTATCCATCGCCATTCCCAGGTTGAGCCCGGGGATTTCACGACAGACTTAAGTA
ACCGCCTACGCACCCTTTACGCCCAGTAATTCCGATTAACGCTTGCACCCTTCGTATTACCGCGGCTGCTGGCACGAAGTTAGCCGGTGCTTATTCCTCAGGTACC
GTCAGCCCCATAGGGTATTAACCCGTGGT

Rh806 variant: GGACTACCAGGGTATCTAAT
Recognition sequence: CGATCTCTACGC
Rh555 sequence: CGGAATTACTGGGCGTAAAGG (IC: CCTTTACGCCCAGTAATTCCG)

>ST4
AGTCTTGCGACCGTACTCCCCAGGCGGCGAACTTAACGCGTTAGCTTCGACACTGATCTCCGAGTTGAGACCAACATCCAGTTCGCATCGTTTAGGGCGTGGACT
ACCAGGGCATCTAATCCTGTTTGCTCCCCACGCTTTCGTGCTTCAGCGTCAGTGTTGTCCCAGATGGCCGCCTTCGCCACTGATGTTCCTCCTGCGCTCCAACTATT
TCACCGCTACACCGGGAATTCCACCATCCTCTGACACACTCTAGCCACCCAGTATCCATCGCCATTCCCAGGTTGAGCCCGGGGATTTCACGACAGACTTAAGTA
ACCGCCTACGCACCCTTTACGCCCAGTAATTCCGATTAACGCTTGCACCCTTCGTATTACCGCGGCTGCTGGCACGAAGTTAGCCGGTGCTTATTCCTCAGGTACC
GTCAGCCCCATAGGGTATTAACCCGTGGT

Rh806 variant: GGACTACCAGGGCATCTAAT
Recognition sequence: TGCGCTCCAACT
Rh555 sequence: CGGAATTACTGGGCGTAAAGG (IC: CCTTTACGCCCAGTAATTCCG)

>ST6
AGTCTTGCGACCGTACTCCCCAGGCGGCGAACTTAACGCGTTAGCTTCGACACTGATCTCCGAGTTGAGACCAACATCCAGTTCGCATCGTTTAGGGCGTGGACT
ACCAGGGTATCTACTCCTGTTTGCTCCCCACGCTTTCGTGCTTCAGCGTCAGTGTTGTCCCAGATGGCCGCCTTCGCCACTGATGTTCCTCCCTTAGCATGCCCATT
TCACCGCTACACCGGGAATTCCACCATCCTCTGACACACTCTAGCCACCCAGTATCCATCGCCATTCCCAGGTTGAGCCCGGGGATTTCACGACAGACTTAAGTA
ACCGCCTACGCACCCTTTACGCCCAGTAATTCCGATTAACGCTTGCACCCTTCGTATTACCGCGGCTGCTGGCACGAAGTTAGCCGGTGCTTATTCCTCAGGTACC
GTCAGCCCCATAGGGTATTAACCCGTGGT

Rh806 variant: GGACTACCAGGGTATCTACT
Recognition sequence: CTTAGCATGCCC
Rh555 sequence: CGGAATTACTGGGCGTAAAGG (IC: CCTTTACGCCCAGTAATTCCG)

>ST7
AGTCTTGCGACCGTACTCCCCAGGCGGCGAACTTAACGCGTTAGCTTCGACACTGATCTCCGAGTTGAGACCAACATCCAGTTCGCATCGTTTAGGGCGTGGACT
ACCAGGGTATCTATTCCTGTTTGCTCCCCACGCTTTCGTGCTTCAGCGTCAGTGTTGTCCCAGATGGCCGCCTTCGCCACTGATGTTCCTCCCGTCCATCTACGATT
TCACCGCTACACCGGGAATTCCACCATCCTCTGACACACTCTAGCCACCCAGTATCCATCGCCATTCCCAGGTTGAGCCCGGGGATTTCACGACAGACTTAAGTA
ACCGCCTACGCACCCTTTACGCCCAGTAATTCCGATTAACGCTTGCACCCTTCGTATTACCGCGGCTGCTGGCACGAAGTTAGCCGGTGCTTATTCCTCAGGTACC
GTCAGCCCCATAGGGTATTAACCCGTGGT

Rh806 variant: GGACTACCAGGGTATCTATT
Recognition sequence: CGTCCATCTACG
Rh555 sequence: CGGAATTACTGGGCGTAAAGG (IC: CCTTTACGCCCAGTAATTCCG)

>ST8
AGTCTTGCGACCGTACTCCCCAGGCGGCGAACTTAACGCGTTAGCTTCGACACTGATCTCCGAGTTGAGACCAACATCCAGTTCGCATCGTTTAGGGCGTGGACT
ACCAGGGTATCTAGTCCTGTTTGCTCCCCACGCTTTCGTGCTTCAGCGTCAGTGTTGTCCCAGATGGCCGCCTTCGCCACTGATGTTCCTCCTCTACCGATGCCATT
TCACCGCTACACCGGGAATTCCACCATCCTCTGACACACTCTAGCCACCCAGTATCCATCGCCATTCCCAGGTTGAGCCCGGGGATTTCACGACAGACTTAAGTA
ACCGCCTACGCACCCTTTACGCCCAGTAATTCCGATTAACGCTTGCACCCTTCGTATTACCGCGGCTGCTGGCACGAAGTTAGCCGGTGCTTATTCCTCAGGTACC
GTCAGCCCCATAGGGTATTAACCCGTGGT

Rh806 variant: GGACTACCAGGGTATCTAGT
Recognition sequence: TCTACCGATGCC
Rh555 sequence: CGGAATTACTGGGCGTAAAGG (IC: CCTTTACGCCCAGTAATTCCG)

>ST11
AGTCTTGCGACCGTACTCCCCAGGCGGCGAACTTAACGCGTTAGCTTCGACACTGATCTCCGAGTTGAGACCAACATCCAGTTCGCATCGTTTAGGGCGTGGACT
ACCAGGGAATCTAATCCTGTTTGCTCCCCACGCTTTCGTGCTTCAGCGTCAGTGTTGTCCCAGATGGCCGCCTTCGCCACTGATGTTCCTCCGAACCTTTCCCGATT
TCACCGCTACACCGGGAATTCCACCATCCTCTGACACACTCTAGCCACCCAGTATCCATCGCCATTCCCAGGTTGAGCCCGGGGATTTCACGACAGACTTAAGTA
ACCGCCTACGCACCCTTTACGCCCAGTAATTCCGATTAACGCTTGCACCCTTCGTATTACCGCGGCTGCTGGCACGAAGTTAGCCGGTGCTTATTCCTCAGGTACC
GTCAGCCCCATAGGGTATTAACCCGTGGT

Rh806 variant: GGACTACCAGGGAATCTAAT
Recognition sequence: GAACCTTTCCCG
Rh555 sequence: CGGAATTACTGGGCGTAAAGG (IC: CCTTTACGCCCAGTAATTCCG)

>ST15
AGTCTTGCGACCGTACTCCCCAGGCGGCGAACTTAACGCGTTAGCTTCGACACTGATCTCCGAGTTGAGACCAACATCCAGTTCGCATCGTTTAGGGCGTGGACT
ACCAGGGGATCTAATCCTGTTTGCTCCCCACGCTTTCGTGCTTCAGCGTCAGTGTTGTCCCAGATGGCCGCCTTCGCCACTGATGTTCCTCCGACCCTAGCTTCAT
TTCACCGCTACACCGGGAATTCCACCATCCTCTGACACACTCTAGCCACCCAGTATCCATCGCCATTCCCAGGTTGAGCCCGGGGATTTCACGACAGACTTAAGTA
ACCGCCTACGCACCCTTTACGCCCAGTAATTCCGATTAACGCTTGCACCCTTCGTATTACCGCGGCTGCTGGCACGAAGTTAGCCGGTGCTTATTCCTCAGGTACC
GTCAGCCCCATAGGGTATTAACCCGTGGT

Rh806 variant: GGACTACCAGGGGATCTAAT
Recognition sequence: GACCCTAGCTTC
Rh555 sequence: CGGAATTACTGGGCGTAAAGG (IC: CCTTTACGCCCAGTAATTCCG)

>ST23
AGTCTTGCGACCGTACTCCCCAGGCGGCGAACTTAACGCGTTAGCTTCGACACTGATCTCCGAGTTGAGACCAACATCCAGTTCGCATCGTTTAGGGCGTGGACT
ATCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGTGCTTCAGCGTCAGTGTTGTCCCAGATGGCCGCCTTCGCCACTGATGTTCCTCCTGCCAGCCCTATATT
TCACCGCTACACCGGGAATTCCACCATCCTCTGACACACTCTAGCCACCCAGTATCCATCGCCATTCCCAGGTTGAGCCCGGGGATTTCACGACAGACTTAAGTA
ACCGCCTACGCACCCTTTACGCCCAGTAATTCCGATTAACGCTTGCACCCTTCGTATTACCGCGGCTGCTGGCACGAAGTTAGCCGGTGCTTATTCCTCAGGTACC
GTCAGCCCCATAGGGTATTAACCCGTGGT

Rh806 variant: GGACTATCAGGGTATCTAAT
Recognition sequence: TGCCAGCCCTAT
Rh555 sequence: CGGAATTACTGGGCGTAAAGG (IC: CCTTTACGCCCAGTAATTCCG)

>ST39
AGTCTTGCGACCGTACTCCCCAGGCGGCGAACTTAACGCGTTAGCTTCGACACTGATCTCCGAGTTGAGACCAACATCCAGTTCGCATCGTTTAGGGCGTGGACT
AACAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGTGCTTCAGCGTCAGTGTTGTCCCAGATGGCCGCCTTCGCCACTGATGTTCCTCCACATCGCTCGTCATT
TCACCGCTACACCGGGAATTCCACCATCCTCTGACACACTCTAGCCACCCAGTATCCATCGCCATTCCCAGGTTGAGCCCGGGGATTTCACGACAGACTTAAGTA
ACCGCCTACGCACCCTTTACGCCCAGTAATTCCGATTAACGCTTGCACCCTTCGTATTACCGCGGCTGCTGGCACGAAGTTAGCCGGTGCTTATTCCTCAGGTACC
GTCAGCCCCATAGGGTATTAACCCGTGGT

Rh806 variant: GGACTAACAGGGTATCTAAT
Recognition sequence: ACATCGCTCGTC
Rh555 sequence: CGGAATTACTGGGCGTAAAGG (IC: CCTTTACGCCCAGTAATTCCG)

>ST55
AGTCTTGCGACCGTACTCCCCAGGCGGCGAACTTAACGCGTTAGCTTCGACACTGATCTCCGAGTTGAGACCAACATCCAGTTCGCATCGTTTAGGGCGTGGACT
AGCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGTGCTTCAGCGTCAGTGTTGTCCCAGATGGCCGCCTTCGCCACTGATGTTCCTCCCTTACCAGTCGCATT
TCACCGCTACACCGGGAATTCCACCATCCTCTGACACACTCTAGCCACCCAGTATCCATCGCCATTCCCAGGTTGAGCCCGGGGATTTCACGACAGACTTAAGTA
ACCGCCTACGCACCCTTTACGCCCAGTAATTCCGATTAACGCTTGCACCCTTCGTATTACCGCGGCTGCTGGCACGAAGTTAGCCGGTGCTTATTCCTCAGGTACC
GTCAGCCCCATAGGGTATTAACCCGTGGT

Rh806 variant: GGACTAGCAGGGTATCTAAT
Recognition sequence: CTTACCAGTCGC
Rh555 sequence: CGGAATTACTGGGCGTAAAGG (IC: CCTTTACGCCCAGTAATTCCG)

Supplemental Material 2: Script used for generation of BIOM files

```
#!/bin/sh
#PBS -j oe
#PBS -l walltime=5:00:00
#PBS -l mem=25gb
#PBS -m bea
#PBS -M anaqib2@uic.edu


cd /MergedFASTAFiles
#Folder where all the merged FASTA files have been stored
FILES=/MergedFASTAFiles/*.fna
#Only files with extension *.fna to be selected

mkdir PrimerTemplateCounts
mkdir OutputFolder

for T in $FILES
do
 Orgname=`basename $T .fna`
 echo $Orgname
 num=2
 while [ $num -gt 1 ] && [ $num -lt 642 ]
    do
     Read1=`awk -v i=$num 'FNR == i {print $1}' PrimerTemplateCombinations.txt`
     Read2=`awk -v j=$num 'FNR == j {print $2}' PrimerTemplateCombinations.txt | sed s'/.$//'`
     PrimerCount=`grep -Ec $Read1.*$Read2 $T`
     echo "$Read1:$Read2:$PrimerCount" >> PrimerTemplateCounts/$Orgname.PrimersTemplateCount.txt
     echo "$PrimerCount" >> OutputFolder/$Orgname.PrimersTemplateCount.txt
     num=`expr $num + 1`
    done
done

paste OutputFolder/*.txt | column -s $'\t' -t > FinalCount.txt
```

Chapter VI: Conclusion

Major advances

Fundamental primer-template interactions have been the focus of this thesis, with an aim to reduce selection bias in PCRs contain complex DNA templates and complex pools of primers. PCR amplification of complex microbial genomic DNA templates with degenerate primers can lead to distortion of the underlying community structure through many mechanisms, and due to inefficient primer-template interactions. In addition to all the reviewed PCR biases, one of the first and major findings of this study has been the identification of a previously undescribed form of PCR bias that changes during the different cycles of PCR amplification [18]. This thesis has theorized that primer-template and primer-amplicon interactions do not operate at the same efficiency, and that primer-template interactions can substantially impact amplification results for up to 12 cycles. In attempting to circumvent the negative interactive effects of simultaneous primer-template and primer-amplicon interactions, a new methodological approach to PCR amplification was developed. This thesis describes this novel technique and demonstrates the efficacy of the deconstructed PCR method for reducing selection bias – both from simultaneous primer-template and primer-amplicon interactions and from the compound effects of low efficiency primer-template interactions occurring over 30 cycles of PCR amplification. Modeling of selection bias in PCR has been effective in simple systems, and the necessary parameters include amplification efficiency and a factor representing competition between primers and amplicons for annealing to templates. However, during PCR amplification, changes in the concentrations of various reactants can lead to alteration of the parameters for amplification efficiency and for primer/amplicon competition. In addition, the concentration of available primers decreases during amplification, as does the composition of the primer pool in reactions with degenerate primers. Furthermore, the percent contribution of linear copying decreases

continuously, but linear copies still contribute >1% of all final PCR products for up to 12 cycles. In this dynamic system, modeling can be difficult, and all these factors contribute to distortion of the underlying template distribution. The success of the deconstructed PCR method is due to the overall simplification of the system; locus-specific degenerate primers are only employed for 2 cycles of amplification, while non-degenerate primers targeting linker sequences perform exponential amplification. However, here the composition of the primer pool does not change over the 30 cycles of PCR amplification. Furthermore, linear copying from gDNA is no longer possible in DePCR due to the removal of all locus-specific primers. One potential downside of the method is an increase in stochastic variation between replicates (PCR drift). Since the genomic DNA is copied only during the first two cycles rather than continuously throughout the reaction, it is possible that the DePCR method could be more sensitive to minor variations in template, annealing temperature, primer-template annealing, etc., thereby leading to increased variation. In fact, this is observed – particularly at lower annealing temperatures. Reproducibility of technical replicates with the DePCR is generally somewhat lower relative to standard PCR. Conversely, it may be that standard PCR is more controlled by selection bias, leading to more overall similar observed community structure and greater distortion of the underlying microbial community. Regardless, the results shown here demonstrate that indeed the two defined types of primer-template interactions within PCR (*i.e.,* natural and artificial interactions with genomic DNA and amplicons, respectively) can and should be separated when degenerate primers are used or when mismatches with the template are anticipated. The analysis here of an artificially synthesized mock community demonstrates the strong potential for a degenerate primer pool of oligonucleotides of varying melting temperatures to preferentially select templates based on sequence variations in the primer site. The DePCR strategy limits the gDNA template-primer

interaction to two cycles, with all subsequent amplification cycles employing non-degenerate, non-template interactions.

As part of this work, both mock communities and environmental samples have been interrogated. Future studies should examine more closely the activity of the DePCR method in environmental samples where the microbial community structure has been well characterized using PCR-independent methodologies. Improvements in sequencing technology should allow for deep sequencing and partial *de novo* assembly of microbial genomes from a sample of initially unknown composition. Thus, for example, comparison of TAS and DePCR methodologies with mammalian fecal gDNA could be better interrogated to determine which method improves the representation of the underlying microbial community structure. Mock systems, although critical for the development of the DePCR methodology, are not perfect proxies for analysis of environmental samples. First, the studies conducted herein have been using synthetic DNA of a very short size, and in an overall low-complexity system with template at a high concentration. The 2.5 ng input concentrations, for example, in the final mock community analyses represent on the scale of 10^8 targets per microliter in the PCR reactions. Microbial genomic DNA, conversely, is less than 1% target, and provides a broad range of alternate priming sites that could confound PCR reactions. In reactions run at very low annealing temperatures, primer mis-annealing in gDNA samples created sequencing output that was heavily dominated by non-16S rRNA genes and led to distortions in the observed microbial community structure. Thus, although PCR annealing temperature leads to improved tolerance of mismatches, this benefit cannot be extended to increasingly low annealing temperatures with environmental samples. Nonetheless, the ten template x 64 primer mock community system was able to generate substantial diversity in primer-template interactions, and could easily be scaled upward – most easily with additional

primers. The combination of mock community and environmental samples provides the ability to determine that similar processes are operating with DePCR in both systems. For example, the relationship between annealing temperature and evenness of primer utilization profiles (PUPs) is one of the most important findings of this study and was observed in both environmental samples and mock community samples.

In addition to reducing PCR selection bias, the DePCR methodology has two other important advantages relative to standard PCR. The first is the dramatic decrease in the formation of chimeric sequences when employing the DePCR method with environmental samples. The likely possibilities for this phenomenon include: (a) reduction in input DNA concentration for exponential amplification due to the double-purification step, (b) higher annealing temperature for the exponential amplification due to targeting of P5/P7 Illumina adapters –potentially reducing the re-annealing of PCR products to other products, and (c) long elongation times during the first cycles, reducing the formation of incomplete molecules during the first stages of PCR. Conceivably, chimera formation with DePCR could be reduced further; in this study, in general 30 cycles of amplification to generate robust PCR yields for sequencing. Although not shown in this work, the amplification of the pool of amplicons during the second stage PCR can be titrated across different numbers of cycles, and the reaction with the fewest numbers of cycles yielding sufficient DNA for sequencing could be employed. In addition, it is critical to remember that the rate of chimera formation represents only the rate of detectable chimera formation, and that chimeras generated from closely related sequences are not only likely to occur at higher rates but also essentially undetectable by chimera detection software. Finally, it should be noted that amplification of fecal gDNA with the TAS protocols resulted in higher observed diversity relative to the same sample amplified with DePCR protocols, and this

could represent the residual presence of chimeras that were not removed. It would not be the first instance in which high observed microbial diversity was later determined to be the result of artifact generation [100].

The second additional advantage of the DePCR methodology is the most crucial for interrogating primer-template interactions. This is the ability to determine which primers in a degenerate primer pool interact with DNA templates, ultimately leading to so-called 'Primer utilization profiles' (PUPs). Since only two cycles of locus-specific primer-template annealing are allowed in DePCR, the oligonucleotide primers incorporated into growing strands represent the actual primer-template interaction. After these two cycles, all further amplification is performed with linker primers that do not interact with the DNA templates, nor do they interact with the locus-specific region of the generated amplicons. This is fundamentally different from standard PCR, where the locus-specific region of primers is used for further annealing and copying during all subsequent cycles of PCR. Because of this, the final sequence data show which primer annealed to the DNA template – and this is novel information, as the primer sequences are typically discarded from downstream analyses because they do not represent the true sequence of the template. Here, the primer sequence is used for a different type of information – regarding template-primer interactions – and in mock communities, the primer annealing site sequence is known. This allows the systematic interrogation of perfect match and mismatch annealing interactions that are at the heart of the final manuscript; such information has never previously been available. Importantly, the interrogation of primer-template interactions is performed under conditions (*i.e.*, cycles 1 and 2) where primers are not limiting in any sense and where no distortion of the evenness of the primer pool is present. If interrogation

of the system occurred later in PCR cycles, certain primers may be selectively consumed by amplification cycles, leading to greater uncertainty in the observed signal.

Using this novel source of information, this study demonstrated how evenness of primer annealing from degenerate primer pools correlates with observed microbial community structure. A robust negative quadratic relationship between DePCR annealing temperature and evenness of primer utilization was observed. As annealing temperature increases, the DePCR system increasingly favors perfect match annealing and elongation, and this leads to fewer and fewer different primers annealing and elongating, and finally leads to shifts in observed community structure. Some of the most critical observations were that although perfect match annealing is important in copying of DNA templates, when degenerate primer pools are available, mismatch amplifications are the dominant form of copying. This is true even during cycle 1 and 2 of PCR reactions when no primer variant should be limiting. Remarkably, when perfect match primers were removed from mock template experiments, a high diversity of mismatch primers were able to reproduce the expected results better than with fewer but perfect match primers. This demonstrates that when using DePCR, a heavily degenerate pool is acceptable and perfect match variants for every target organism is not necessary.

This study also addressed the role of primer melting temperature variation in degenerate primer pools. Since degenerate primers can have a broad range of melting temperatures due to varying nucleotide positions, individual primers might be favored simply due to higher melting temperature. In a systematic analysis using multiple complex environmental samples, no strong effect of primer melting temperature was observed. Likewise, in analysis of average theoretical melting temperature in mock community analyses, no strong selection for high melting temperature primers were observed. Thus, it is likely that sequence matching and mismatch

tolerance drive primer-template annealing in PCR rather than melting temperature considerations, at least under the melting temperature ranges targeting 16S rRNA genes. Degenerate primer pools for protein-coding genes, with extraordinarily high degeneracy, might show stronger melting temperature effects.

In summary, this study provides a totally novel strategy for deeply exploring primer-template interactions, providing information previously inaccessible. Some primer-template interaction phenomena are confirmed – 3' mismatches are destabilizing and perfect matches favored. Other phenomena are novel: perfect matches may be favored, but are the dominant type of interaction, and non-perfect match amplification starts immediately during the first cycles of PCR, not in later cycles. Furthermore, amplification using the DePCR methodology better tolerates multiple mismatches leading to improved recovery of expected community structure. In addition, this study establishes an experimental system for interrogating primer-template interactions, by providing a mechanism for identifying perfect match and mismatch primer-template interactions. Such an experimental system has broad applicability and will provide empirical evidence for future degenerate pool primer design.

Ultimately, the study sought to better understand the relationship between primers and templates, particularly with regard to mismatch tolerance, to help improve the design of complex primer pools for amplification of complex environmental samples. This study contributes to the scientific community with a new source of error, a novel method to better evaluate and reduce that error, a systematic protocol to generate empirical evidence of primer-template interactions (PUPs and template profiling), and accessory data signifying relationship between primer degeneracy, annealing temperature, and primer-template interactions in mock community and complex environmental systems.

Cited Literature

1. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI: **QIIME allows analysis of high-throughput community sequencing data**. *Nature methods* 2010, **7**(5):335.
2. Cho I, Blaser MJ: **The human microbiome: at the interface of health and disease**. *Nature Reviews Genetics* 2012, **13**(4):260.
3. Janssens Y, Nielandt J, Bronselaer A, Debunne N, Verbeke F, Wynendaele E, Van Immerseel F, Vandewynckel Y-P, De Tré G, De Spiegeleer B: **Disbiome database: linking the microbiome to disease**. *Bmc Microbiol* 2018, **18**(1):50.
4. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR: **Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses**. *Proceedings of the National Academy of Sciences* 1985, **82**(20):6955-6959.
5. Větrovský T, Baldrian P: **The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses**. *PloS one* 2013, **8**(2):e57923.
6. Lee ZM-P, Bussema III C, Schmidt TM: **rrn DB: Documenting the number of rRNA and tRNA genes in bacteria and archaea**. *Nucleic acids research* 2008, **37**(suppl_1):D489-D493.
7. Liesack W, Weyland H, Stackebrandt E: **Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed-culture of strict barophilic bacteria**. *Microbial Ecology* 1991, **21**(1):191-198.
8. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N: **Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia**. *Science* 1985, **230**(4732):1350-1354.

9. Schmidt TM, DeLong EF, Pace NR: **Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing**. *Journal of bacteriology* 1991, **173**(14):4371-4378.

10. Muyzer G, De Waal EC, Uitterlinden AG: **Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA**. *Applied environmental microbiology* 1993, **59**(3):695-700.

11. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ: **Microbial diversity in the deep sea and the underexplored "rare biosphere"**. *Proceedings of the National Academy of Sciences* 2006, **103**(32):12115-12120.

12. Amann RI, Ludwig W, Schleifer K-H: **Phylogenetic identification and in situ detection of individual microbial cells without cultivation**. *Microbiological reviews* 1995, **59**(1):143-169.

13. Reysenbach A-L, Giver LJ, Wickham GS, Pace NR: **Differential amplification of rRNA genes by polymerase chain reaction**. *Applied Environmental Microbiology* 1992, **58**(10):3417-3418.

14. Wagner A, Blackstone N, Cartwright P, Dick M, Misof B, Snow P, Wagner GP, Bartels J, Murtha M, Pendleton J: **Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift**. *Systematic Biology* 1994, **43**(2):250-261.

15. Polz MF, Cavanaugh CM: **Bias in template-to-product ratios in multitemplate PCR**. *Applied environmental Microbiology* 1998, **64**(10):3724-3730.

16. Pinto AJ, Raskin L: **PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets**. *PloS one* 2012, **7**(8):e43093.

17. Sipos R, Székely AJ, Palatinszky M, Révész S, Márialigeti K, Nikolausz M: **Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis**. *FEMS Microbiology Ecology* 2007, **60**(2):341-350.

18. Green SJ, Venkatramanan R, Naqib A: **Deconstructing the polymerase chain reaction: understanding and correcting bias associated with primer degeneracies and primer-template mismatches**. *PloS one* 2015, **10**(5):e0128122.

19. Ogino S, Wilson RB: **Quantification of PCR bias caused by a single nucleotide polymorphism in SMN gene dosage analysis**. *The Journal of molecular diagnostics* 2002, **4**(4):185-190.

20. Hong S, Bunge J, Leslin C, Jeon S, Epstein SS: **Polymerase chain reaction primers miss half of rRNA microbial diversity**. *The ISME Journal* 2009, **3**(12):1365.

21. Gohl DM, Vangay P, Garbe J, MacLean A, Hauge A, Becker A, Gould TJ, Clayton JB, Johnson TJ, Hunter R: **Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies**. *Nature biotechnology* 2016, **34**(9):942.

22. Hansen MC, Tolker-Nielsen T, Givskov M, Molin S: **Biased 16S rDNA PCR amplification caused by interference from DNA flanking the template region**. *FEMS Microbiology Ecology* 1998, **26**(2):141-149.

23. Wilson KH, Blitchington RB: **Human colonic biota studied by ribosomal DNA sequence analysis**. *Applied Environmental Microbiology* 1996, **62**(7):2273-2278.

24. Chandler D, Fredrickson J, Brockman F: **Effect of PCR template concentration on the composition and distribution of total community 16S rDNA clone libraries**. *Molecular Ecology* 1997, **6**(5):475-482.

25. Kennedy K, Hall MW, Lynch MD, Moreno-Hagelsieb G, Neufeld JD: **Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles**. *Applied environmental microbiology* 2014:AEM. 01451-01414.

26. D'Amore R, Ijaz UZ, Schirmer M, Kenny JG, Gregory R, Darby AC, Shakya M, Podar M, Quince C, Hall N: **A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling**. *BMC genomics* 2016, **17**(1):55.

27. Ishii K, Fukui M: **Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR**. *Applied Environmental Microbiology* 2001, **67**(8):3753-3755.

28. Suzuki MT, Giovannoni SJ: **Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR**. *Applied environmental microbiology* 1996, **62**(2):625-630.

29. Warnecke PM, Stirzaker C, Melki JR, Millar DS, Paul CL, Clark SJ: **Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA**. *Nucleic acids research* 1997, **25**(21):4422-4426.

30. Mathieu-Daudé F, Welsh J, Vogt T, McClelland M: **DNA rehybridization during PCR: the 'C o t effect'and its consequences**. *Nucleic acids research* 1996, **24**(11):2080-2086.

31. Becker S, Boger P, Oehlmann R, Ernst A: **PCR bias in ecological analysis: A case study for quantitative Taq nuclease assays in analyses of microbial communities**. *Appl Environ Microb* 2000, **66**(11):4945-+.

32. Kurata S, Kanagawa T, Magariyama Y, Takatsu K, Yamada K, Yokomaku T, Kamagata Y: **Reevaluation and reduction of a PCR bias caused by reannealing of templates**. *Applied environmental microbiology* 2004, **70**(12):7545-7549.

33. Innis MA, Gelfand DH, Sninsky JJ, White TJ: **PCR protocols: a guide to methods and applications**: Academic press; 2012.

34. Wang GC, Wang Y: **The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species**. *Microbiology* 1996, **142**(5):1107-1114.

35. Wang G, Wang Y: **Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes**. *Applied environmental microbiology* 1997, **63**(12):4645-4650.

36. Schloss PD, Gevers D, Westcott SL: **Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies**. *PloS one* 2011, **6**(12):e27310.

37. Qiu X, Wu L, Huang H, McDonel PE, Palumbo AV, Tiedje JM, Zhou J: **Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning**. *Applied environmental microbiology* 2001, **67**(2):880-887.

38. Naqib A, Poggi S, Wang W, Hyde M, Kunstman K, Green SJ: **Making and Sequencing Heavily Multiplexed, High-Throughput 16S Ribosomal RNA Gene Amplicon Libraries Using a Flexible, Two-Stage PCR Protocol**. In: *Gene Expression Analysis.* Springer; 2018: 149-169.

39. Ionescu D, Overholt WA, Lynch MD, Neufeld JD, Naqib A, Green SJ: **Microbial community analysis using high-throughput amplicon sequencing**. In: *Manual of*

Environmental Microbiology, Fourth Edition. American Society of Microbiology; 2016: 2.4. 2-1-2.4. 2-26.

40. Hongoh Y, Yuzawa H, Ohkuma M, Kudo T: **Evaluation of primers and PCR conditions for the analysis of 16S rRNA genes from a natural environment**. *FEMS Microbiology Letters* 2003, **221**(2):299-304.

41. Shen L, Guo Y, Chen X, Ahmed S, Issa J-PJ: **Optimizing annealing temperature overcomes bias in bisulfite PCR methylation analysis**. *Biotechniques* 2007, **42**(1):48-58.

42. Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, Wilkins MJ, Wrighton KC, Williams KH, Banfield JF: **Unusual biology across a group comprising more than 15% of domain Bacteria**. *Nature* 2015, **523**(7559):208.

43. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K: **A new view of the tree of life**. *Nature microbiology* 2016, **1**(5):16048.

44. Isenbarger TA, Finney M, Ríos-Velázquez C, Handelsman J, Ruvkun G: **Miniprimer PCR, a new lens for viewing the microbial world**. *Applied environmental microbiology* 2008, **74**(3):840-849.

45. Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kauserud H: **ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases**. *Bmc Microbiol* 2010, **10**(1):189.

46. Hayashi H, Sakamoto M, Benno Y: **Evaluation of three different forward primers by terminal restriction fragment length polymorphism analysis for determination of fecal Bifidobacterium spp. in healthy subjects**. *Microbiology and immunology*

2004, **48**(1):1-6.

47. Lane D: **16S/23S rRNA sequencing**. *Nucleic acid techniques in bacterial systematics* 1991:115-175.

48. Marchesi JR, Sato T, Weightman AJ, Martin TA, Fry JC, Hiom SJ, Wade WG: **Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA**. *Applied environmental microbiology* 1998, **64**(2):795-799.

49. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M: **Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms**. *The ISME journal* 2012, **6**(8):1621.

50. v. Wintzingerode F, Göbel UB, Stackebrandt E: **Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis**. *FEMS microbiology reviews* 1997, **21**(3):213-229.

51. Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF: **PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample**. *Applied environmental microbiology* 2005, **71**(12):8966-8969.

52. Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ: **Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes**. *Applied environmental microbiology* 2008, **74**(8):2461-2470.

53. Lee CK, Herbold CW, Polson SW, Wommack KE, Williamson SJ, McDonald IR, Cary SC: **Groundtruthing next-gen sequencing for microbial ecology–biases and errors in community structure estimates from PCR amplicon pyrosequencing**. *PloS one* 2012, **7**(9):e44224.

54. Kembel SW, Wu M, Eisen JA, Green JL: **Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance**. *PLoS computational biology* 2012, **8**(10):e1002743.

55. Mao DP, Zhou Q, Chen CY, Quan ZX: **Coverage evaluation of universal bacterial primers using the metagenomic datasets**. *Bmc Microbiol* 2012, **12**.

56. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R: **Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample**. *Proceedings of the national academy of sciences* 2011, **108**(Supplement 1):4516-4522.

57. Prakash O, Green SJ, Jasrotia P, Overholt WA, Canion A, Watson DB, Brooks SC, Kostka JE: **Rhodanobacter denitrificans sp. nov., isolated from nitrate-rich zones of a contaminated aquifer**. *International journal of systematic and evolutionary microbiology* 2012, **62**(10):2457-2462.

58. Kostka JE, Green SJ, Rishishwar L, Prakash O, Katz LS, Mariño-Ramírez L, Jordan IK, Munk C, Ivanova N, Mikhailova N: **Genome sequences for six Rhodanobacter strains, isolated from soils and the terrestrial subsurface, with variable denitrification capabilities**. *Journal of bacteriology* 2012, **194**(16):4461-4462.

59. Bybee SM, Bracken-Grissom H, Haynes BD, Hermansen RA, Byers RL, Clement MJ, Udall JA, Wilcox ER, Crandall KA: **Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics**. *Genome biology evolution* 2011, **3**:1312-1323.

60. de Cárcer DA, Denman SE, McSweeney C, Morrison M: **A strategy for modular tagged high-throughput amplicon sequencing**. *Applied environmental microbiology* 2011:AEM. 05146-05111.

61. Moonsamy P, Williams T, Bonella P, Holcomb C, Höglund B, Hillman G, Goodridge D, Turenchalk G, Blake L, Daigle D: **High throughput HLA genotyping using 454 sequencing and the Fluidigm Access Array™ system for simplified amplicon library preparation**. *Tissue antigens* 2013, **81**(3):141-149.

62. Loakes D, Brown DM, Linde S, Hill F: **3-Nitropyrrole and 5-nitroindole as universal bases in primers for DNA sequencing and PCR**. *Nucleic acids research* 1995, **23**(13):2361-2366.

63. Gihring TM, Green SJ, Schadt CW: **Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes**. *Environmental microbiology* 2012, **14**(2):285-290.

64. Wang Q, Garrity GM, Tiedje JM, Cole JR: **Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy**. *Applied environmental microbiology* 2007, **73**(16):5261-5267.

65. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P: **An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea**. *The ISME journal* 2012, **6**(3):610.

66. Clarke KR: **Non-parametric multivariate analyses of changes in community structure**. *Australian journal of ecology* 1993, **18**(1):117-143.

67. Wu JH, Hong PY, Liu WT: **Quantitative effects of position and type of single mismatch on single base primer extension**. *J Microbiol Meth* 2009, **77**(3):267-275.

68. Bru D, Martin-Laurent F, Philippot L: **Quantification of the detrimental effect of a single primer-template mismatch by real-time PCR using the 16S rRNA gene as an example**. *Applied environmental microbiology* 2008, **74**(5):1660-1663.

69. Crosby LD, Criddle CS: **Gene capture and random amplification for quantitative recovery of homologous genes**. *Molecular cellular probes* 2007, **21**(2):140-147.

70. Chang S-S, Hsu H-L, Cheng J-C, Tseng C-P: **An efficient strategy for broad-range detection of low abundance bacteria without DNA decontamination of PCR reagents**. *PloS one* 2011, **6**(5):e20303.

71. Klappenbach JA, Saxman PR, Cole JR, Schmidt TM: **rrndb: the ribosomal RNA operon copy number database**. *Nucleic acids research* 2001, **29**(1):181-184.

72. Angly FE, Dennis PG, Skarshewski A, Vanwonterghem I, Hugenholtz P, Tyson GW: **CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction**. *Microbiome* 2014, **2**(1):11.

73. Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, Gilbert JA, Jansson JK, Caporaso JG, Fuhrman JA: **Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys**. *Msystems* 2016, **1**(1):e00009-00015.

74. Owczarzy R, Tataurov AV, Wu Y, Manthey JA, McQuisten KA, Almabrazi HG, Pedersen KF, Lin Y, Garretson J, McEntaggart NO: **IDT SciTools: a suite for analysis**

and design of nucleic acid oligomers. *Nucleic acids research* 2008, **36**(suppl_2):W163-W169.

75. Zhang J, Kobert K, Flouri T, Stamatakis A: **PEAR: a fast and accurate Illumina Paired-End reAd mergeR**. *Bioinformatics* 2013, **30**(5):614-620.

76. Edgar R: **Usearch**. In.: Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA (United States); 2010.

77. McDonald D, Clemente JC, Kuczynski J, Rideout JR, Stombaugh J, Wendel D, Wilke A, Huse S, Hufnagle J, Meyer F: **The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome**. *GigaScience* 2012, **1**(1):7.

78. Clarke K, Gorley R: **Getting started with PRIMER v7**. *PRIMER-E: Plymouth, Plymouth Marine Laboratory* 2015.

79. Team RC: **R: A language and environment for statistical computing**. 2013.

80. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara R, Simpson GL, Solymos P, Stevens MHH, Wagner H: **vegan: Community ecology package**. *R package version* 2011:117-118.

81. Wickham H: **ggplot2: elegant graphics for data analysis**. *Stat Softw* 2010, **35**(1):65-88.

82. Apprill A, McNally S, Parsons R, Weber L: **Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton**. *Aquatic Microbial Ecology* 2015, **75**(2):129-137.

83. Parada AE, Needham DM, Fuhrman JA: **Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples**. *Environmental microbiology* 2016, **18**(5):1403-1414.

84. Hugenholtz P, Huber T: **Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases**. *International journal of systematic and evolutionary microbiology* 2003, **53**(1):289-293.

85. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R: **UCHIME improves sensitivity and speed of chimera detection**. *Bioinformatics* 2011, **27**(16):2194-2200.

86. Lahr DJ, Katz LA: **Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase**. *Biotechniques* 2009, **47**(4):857-866.

87. Fonseca V, Nichols B, Lallias D, Quince C, Carvalho G, Power D, Creer S: **Sample richness and genetic diversity as drivers of chimera formation in nSSU metagenetic analyses**. *Nucleic Acids Research* 2012, **40**(9):e66-e66.

88. Kanagawa T: **Bias and artifacts in multitemplate polymerase chain reactions (PCR)**. *Journal of bioscience and bioengineering* 2003, **96**(4):317-323.

89. Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ: **Removing noise from pyrosequenced amplicons**. *BMC bioinformatics* 2011, **12**(1):38.

90. Lundberg DS, Yourstone S, Mieczkowski P, Jones CD, Dangl JL: **Practical innovations for high-throughput amplicon sequencing**. *Nature methods* 2013, **10**(10):999.

91. Fox J, Weisberg S: **Multivariate linear models in R**. *An R Companion to Applied Regression Los Angeles: Thousand Oaks* 2011.

92. Parks DH, Tyson GW, Hugenholtz P, Beiko RG: **STAMP: statistical analysis of taxonomic and functional profiles**. *Bioinformatics* 2014, **30**(21):3123-3124.

93. White JR, Nagarajan N, Pop M: **Statistical methods for detecting differentially abundant features in clinical metagenomic samples**. *PLoS computational biology* 2009, **5**(4):e1000352.

94. Hochberg Y, Benjamini Y: **More powerful procedures for multiple significance testing**. *Statistics in medicine* 1990, **9**(7):811-818.

95. Kwok S, Kellogg D, McKinney N, Spasic D, Goda L, Levenson C, Sninsky J: **Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies**. *Nucleic acids research* 1990, **18**(4):999-1005.

96. Naqib A, Poggi S, Green SJ: **Deconstructing the Polymerase Chain Reaction II: An improved workflow and effects on artifact formation and primer degeneracy**. In.: PeerJ Preprints; 2019.

97. Ahn J-H, Kim B-Y, Song J, Weon H-Y: **Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities**. *Journal of Microbiology* 2012, **50**(6):1071-1074.

98. Pomp D, Medrano J: **Organic solvents as facilitators of polymerase chain reaction**. *Biotechniques* 1991, **10**(1):58-59.

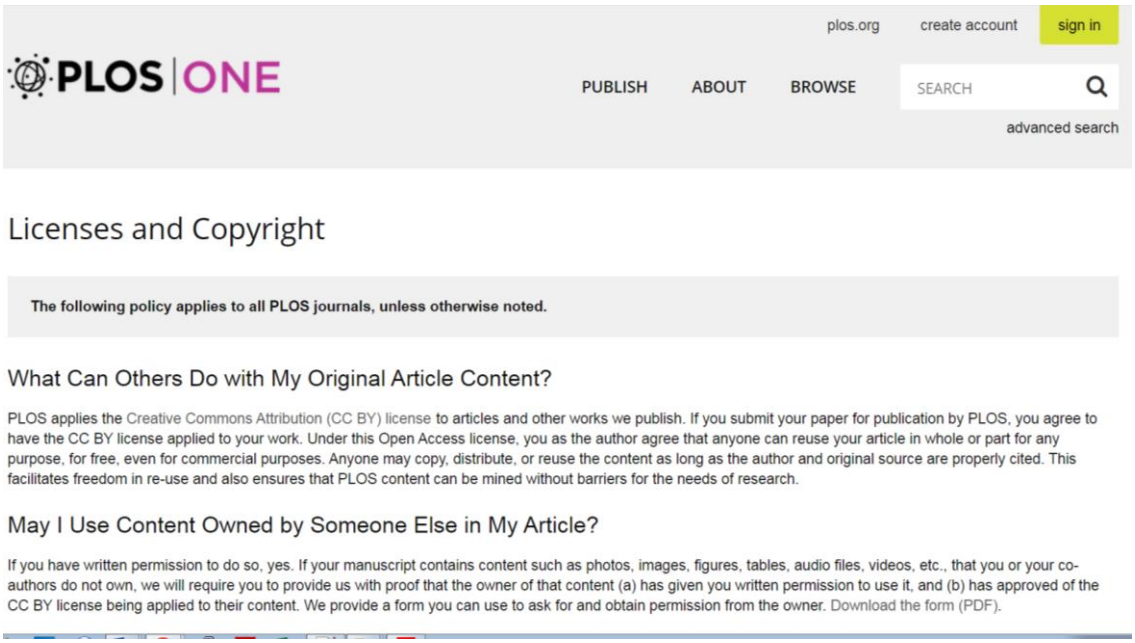
99. Naqib A, Jeon T, Kunstman K, Wang W, Shen Y, Sweeney D, Hyde M, Green SJ: **PCR effects of melting temperature adjustment of individual primers in degenerate primer pools**. *In preparation*.

100. Huse SM, Welch DM, Morrison HG, Sogin ML: **Ironing out the wrinkles in the rare biosphere through improved OTU clustering**. *Environmental microbiology* 2010, **12**(7):1889-1898.

101. Liu W-T, Marsh TL, Cheng H, Forney LJ: **Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA**. *Applied environmental microbiology* 1997, **63**(11):4516-4522.

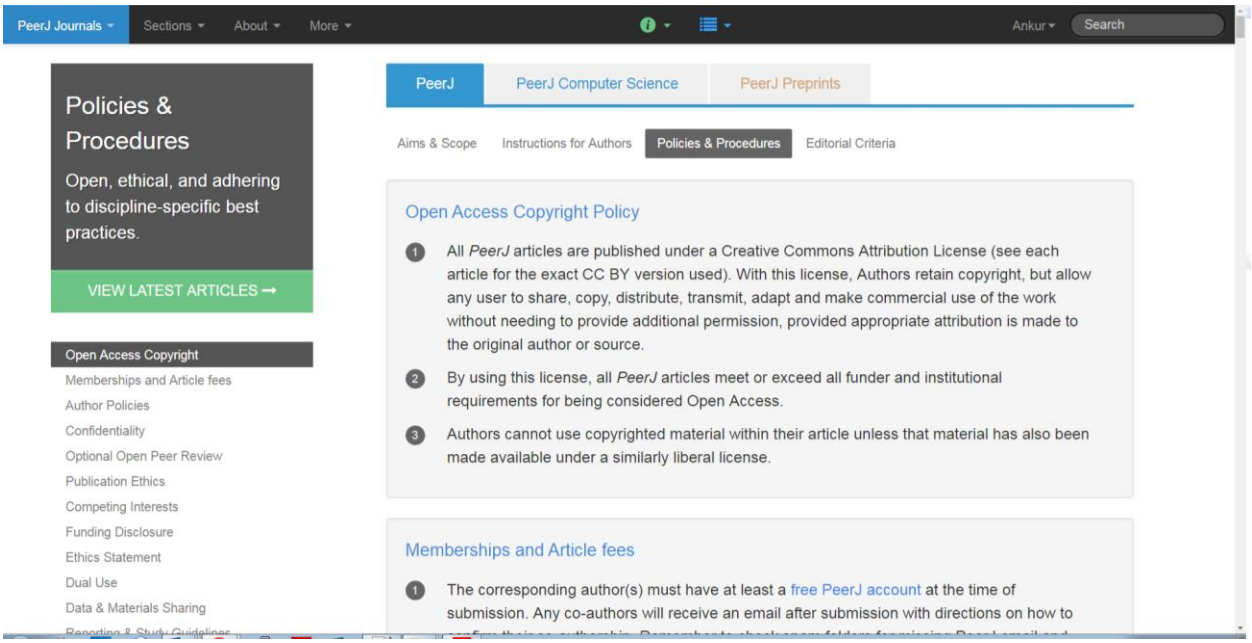
APPENDIX-I

The work presented in Chapter II of this thesis is published in PLOS ONE. I am an author for this published article. The screenshot below is from the PLOS ONE website. As PLOS ONE is an open-access journal, I retain the rights to use my published work for my thesis.



APPENDIX-II

The work presented in Chapter IV of this thesis is published in PeerJ open access journal. I am the first author of this published article. The screenshot below is from the PeerJ website. As PeerJ is an open-access journal, I retain the rights to use my published work for my thesis.



VITA

Ankur Naqib

Education

- **Doctor of Philosophy, Bioengineering** Fall 2014 – Spring 2019
  University of Illinois at Chicago, Chicago, USA.
- **Bachelor of Technology, Bioinformatics** Fall 2004 – Spring 2008
  Dr. D. Y. Patil University, Maharashtra, India.

Professional and Academic Experience

- **Graduate Research Assistant, Sequencing Core** Jan 2013 – Jan 2019
  University of Illinois at Chicago, Chicago, IL
- **Project Engineer, Wipro Technologies Pvt. Ltd** Nov 2009 – July 2012
  Bangalore, India

Technical Skills

- **Programming Languages.**
  Perl | Python | R | Shell scripting
- **Bioinformatics Environments**.
  QIIME | GATK | Primer7 | Picard | ANNOVAR | CLC Genomics | STAMP

Published Journal Articles and Book Chapters

20. Naqib A, Jeon T, Kunstman K, Wang W, Shen Y, Sweeney D, Hyde M, Green SJ. 2019. PCR effects of melting temperature adjustment of individual primers in degenerate primer pools. PeerJ 7:e6570.

19. Dodiya HB, Forsyth CB, Voigt RM, Engen PA, Patel J, Shaikh M, Green SJ, **Naqib A**, Roy A, Kordower JH, Pahan K. Chronic stress-induced gut dysfunction exacerbates Parkinson's disease phenotype and pathology in a rotenone-induced mouse model of Parkinson's disease. Neurobiology of Disease. 2018 Dec 21.

18. Perez-Pardo P, Dodiya HB, Engen PA, Forsyth CB, Huschens AM, Shaikh M, Voigt RM, **Naqib A**, Green SJ, Kordower JH, Shannon KM. Role of TLR4 in the gut-brain axis in Parkinson's disease: a translational study from men to mice. Gut. 2018 Dec 15:gutjnl-2018.

17. Perez-Pardo, P., Dodiya, H. B., Engen, P. A., **Naqib, A.**, Forsyth, C. B., Green, S. J., ... & Kraneveld, A. D. (2018). Gut bacterial composition in a mouse model of Parkinson's disease. Beneficial microbes, 9(5), 799-814.

16. Yang, H. J., LoSavio, P. S., Engen, P. A., **Naqib, A.**, Mehta, A., Kota, R., ... & Keshavarzian, A. (2018). Association of nasal microbiome and asthma control in patients with chronic rhinosinusitis. Clinical & Experimental Allergy.

15. Naqib, A., Poggi, S., Wang, W., Hyde, M., Kunstman, K., & Green, S. J. (2018). Making and Sequencing Heavily Multiplexed, High-Throughput 16S Ribosomal RNA Gene Amplicon Libraries Using a Flexible, Two-Stage PCR Protocol. In Gene Expression Analysis (pp. 149-169). Humana Press, New York, NY.

14. Adami, G. R., Tangney, C. C., Tang, J. L., Zhou, Y., Ghaffari, S., **Naqib, A.**, ... & Schwartz, J. L. (2018). Effects of green tea on miRNA and microbiome of oral epithelium. Scientific reports, 8(1), 5873.

13. Xiao, L., van't Land, B., Engen, P. A., **Naqib, A.**, Green, S. J., Nato, A., ... & Folkerts, G. (2018). Human milk oligosaccharides protect against the development of autoimmune diabetes in NOD-mice. Scientific reports, 8(1), 3829**.**

12. Bishehsari, F., Engen, P. A., Preite, N. Z., Tuncil, Y. E., **Naqib, A.**, Shaikh, M., ... & Khazaie, K. (2018). Dietary Fiber Treatment Corrects the Composition of Gut Microbiota, Promotes SCFA Production, and Suppresses Colon Carcinogenesis. Genes, 9(2), 102.

11. Mahdavinia, M., Engen, P. A., LoSavio, P. S., **Naqib, A.**, Khan, R. J., Tobin, M. C., ... & Tajudeen, B. A. (2018). The nasal microbiome in chronic rhinosinusitis: analyzing the effects of atopy and bacterial functional pathways in 111 patients. Journal of Allergy and Clinical Immunology.

10. Collado, M. C., Engen, P. A., Bandín, C., Cabrera-Rubio, R., Voigt, R. M., **Naqib, A.**, Green, S. J., ... & Garaulet, M. (2017). Timing of food intake impacts daily rhythms of huan salivary microbiota: a randomized, crossover study. The FASEB Journal, fj-201700697RR.

9. Mehta, S. D., Pradhan, A. K., Green, S. J., **Naqib, A.**, Odoyo-June, E., Gaydos, C. A., ... & Bailey, R. C. (2017). Microbial Diversity of Genital Ulcers of HSV-2 Seropositive Women. Scientific Reports, 7(1), 15475.

8. Krarup, A. R., Abdel-Mohsen, M., Schleimann, M. H., **Naqib, A.**, Vibholm, L., Engen, P. A., Dige, A. ... & Keshavarzian, A. (2017). The TLR9 agonist MGN1703 triggers a potent type I interferon response in the sigmoid colon. Mucosal immunology.

7. Mahdavinia, M., Rasmussen, H. E., Engen, P., Van den Berg, J. P., Davis, E., **Naqib, A.**, Engen, K., ... & Lunjani, N. (2017). Atopic dermatitis and food sensitization in South African toddlers: role of fiber and gut microbiota. Annals of Allergy, Asthma & Immunology, 118(6), 742-743.

6. Engen, P. A., Dodiya, H. B., **Naqib, A.**, Forsyth, C. B., Green, S. J., Voigt, R. M., ... & Keshavarzian, A. (2017). The potential role of gut-derived inflammation in multiple system atrophy. Journal of Parkinson's disease, 7(2), 331-346.

5. Voigt, R. M., Summa, K. C., Forsyth, C. B., Green, S. J., Engen, P., **Naqib, A.**, ... & Keshavarzian, A. (2016). The circadian clock mutation promotes intestinal dysbiosis. Alcoholism: Clinical and Experimental Research, 40(2), 335-347.

4. Ionescu, D., Overholt, W. A., Lynch, M. D., Neufeld, J. D., **Naqib, A.**, & Green, S. J. (2016). Microbial community analysis using high-throughput amplicon sequencing. In Manual of Environmental Microbiology, Fourth Edition (pp. 2-4). American Society of Microbiology.

3. Keshavarzian, A., Green, S. J., Engen, P. A., Voigt, R. M., **Naqib, A.**, Forsyth, C. B., ... & Shannon, K. M. (2015). Colonic bacterial composition in Parkinson's disease. Movement Disorders, 30(10), 1351-1360.

2. Earley, Z. M., Akhtar, S., Green, S. J., **Naqib, A.**, Khan, O., Cannon, A. R. & Gamelli, R. L. (2015). Burn injury alters the intestinal microbiome and increases gut permeability and bacterial translocation. PloS one, 10(7), e0129996.

1. Green, S. J., Venkatramanan, R., & **Naqib, A.** (2015). Deconstructing the polymerase chain reaction: understanding and correcting bias associated with primer degeneracies and primer-template mismatches. PloS one, 10(5), e0128122.

* All authors contributed equally to the manuscript.

Posters

2. Milani, B. Y., **Naqib, A.**, Green, S. J., & Maumenee, I. H. (2015). Mutation Analysis In Eight Families With Congenital Motor Nystagmus. Investigative Ophthalmology & Visual Science, 56(7), 5995-5995.

1. Earley, Z. M., Akhtar, S., Green, S., **Naqib, A.**, Khan, O., Cannon, A., & Gamelli, R. L. (2015, June). Dysbiosis Of Intestinal Microbiome After Severe Burn Injury. In SHOCK (Vol. 43, No. 6, pp. 22-23). TWO COMMERCE SQ, 2001 MARKET ST, PHILADELPHIA, PA 19103 USA: LIPPINCOTT WILLIAMS & WILKINS.

References

- **Dr. Stefan J. Green (greendna@uic.edu)** – Director (Sequencing Core), Associate Director (Research Resources Center), University of Illinois at Chicago. Chicago, IL.
  Dr. Ali Keshavarzian (ali_keshavarzian@rush.edu) – Director (Division of Digestive Diseases and Nutrition), Professor at Department of Internal Medicine, Rush Medical College. Chicago, IL.
- **Dr. Supriya Mehta (supriyad@uic.edu)** – Associate Professor of Epidemiology, University of Illinois at Chicago. Chicago, IL
- **Dr. Irene H. Maumenee (maumenee@uic.edu)** – Professor of Ophthalmology, Columbia University Irving Medical Center. New York City, NY