Quantitating primer-template interactions using a deconstructed PCR methodology

Ankur Naqib [1, 2], **Trisha Jeon** [1], **Stefan J Green** [Corresp. 1]

[1] Sequencing Core, Research Resources Center, University of Illinois at Chicago, Chicago, IL, United States

[2] Rush University, Chicago, Illinois, United States

Corresponding Author: Stefan J Green
Email address: GreenDNA@uic.edu

When the polymerase chain reaction (PCR) is used to amplify simultaneously multiple templates, preferential amplification of certain templates (PCR bias) leads to a distorted representation of the original templates in the 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 uncertainty when attempting to optimize PCR reactions and degenerate 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 previously described deconstructed PCR (DePCR) methodology, we generated empirical data showing individual primer interactions with templates in complex template-primer amplification reactions. 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 that mismatched primer annealing and polymerase copying starts immediately during the first two cycle of PCR. In reactions with degenerate primer pools, multiple mismatches between primer and template are tolerated, and these do not have a strong effect on observed template ratios after amplification when employing the DePCR methodology. When employing the DePCR methodology, mismatched primer-template interactions were able to amplify source templates with significantly lower distortion relative to standard PCR. We establish here a quantitative experimental system for interrogating primer-template interactions and demonstrate the efficacy of the DePCR method for amplification of complex template mixtures with complex primer pools.

Quantitating primer-template interactions using a deconstructed PCR methodology

Running title: Primer-template interactions in DePCR

Ankur Naqib[1^], Trisha Jeon[1], Stefan J. Green[1#]

Author affiliations: [1]Sequencing Core, Research Resources Center, University of Illinois at Chicago

^ Current affiliation: Rush University, Chicago, Illinois

Corresponding author: Stefan J. Green, PhD (GreenDNA@uic.edu); 835 S. Wolcott, A-310, Chicago, Illinois, 60612

Abstract

When the polymerase chain reaction (PCR) is used to amplify simultaneously multiple templates, preferential amplification of certain templates (PCR bias) leads to a distorted representation of the original templates in the 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 uncertainty when attempting to optimize PCR reactions and degenerate 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 previously described deconstructed PCR (DePCR) methodology, we generated empirical data showing individual primer interactions with templates in complex template-primer amplification reactions. 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 that mismatched primer annealing and polymerase copying starts immediately during the first two cycle of PCR. In reactions with degenerate primer pools, multiple mismatches between primer and template are tolerated, and these do not have a strong effect on observed template ratios after amplification when employing the DePCR methodology. When employing the DePCR methodology, mismatched primer-template interactions were able to amplify source templates with significantly lower distortion relative to standard PCR. We establish here a quantitative experimental system for interrogating primer-template interactions and demonstrate the efficacy of the DePCR method for amplification of complex template mixtures with complex primer pools.

Introduction

The polymerase chain reaction (PCR) is a well-established tool for amplification of regions of DNA [1, 2] 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 [3-5]. Wagner et al. [5] 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 [6, 7], mismatches between template and primer [8, 9], location of mismatches between template and primer [10], interference from flanking regions during initial stages of PCR [11], too many PCR cycles [12], input DNA concentration [13-15], preferential amplification of low GC templates in a mixture [16], higher GC content in primer region/differences in primer binding energy [3, 17], template saturation at the plateau phase of PCR [2], preferential formation of primer dimers from some primer variants when working with degenerate pools of primers [3], preferential amplification of unmethylated DNA [18], re-annealing of PCR copies to templates leading to reduced amplification efficiency [19, 20], temperature ramp during thermocycling allowing for formation of homoduplexes [21], and combinatorial effects of linear copying of gDNA and exponential amplification of PCR products occurring simultaneously and at different efficiencies [7].

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) [3, 5]. To reduce PCR drift, multiple reactions are typically combined. However, Suzuki and Giovannoni [2] suggested that PCR selection was the primary driver of PCR bias, though low input gDNA could lead to higher stochastic effects [3]. A third category of bias should also be considered – the generation of PCR

artifacts, such as chimeras [21]. 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 [22]. Reducing cycle number is always recommended with regards to decreasing chimera formation [10, 23-25].

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 [16], DMSO and glycerol [26], running fewer cycles of PCR [2, 27, 28], reducing degeneracies in primers whenever possible [3], increasing ramp rates for transitions between temperatures [21], and use of long elongation times and/or use of highly processive polymerases to ensure complete copying during each cycle [25]. In some systems, higher annealing temperatures are recommended to reduce effects of secondary structure [29], while in complex template systems such as microbial DNA, lower annealing temperatures are recommended to improve tolerance for mismatch annealing [30]. We have also introduced the "deconstructed PCR" (DePCR) method [7, 31] to reduce PCR bias by addressing several issues simultaneously. First, locus-specific primers are only employed for two cycles in DePCR, 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; [32]) or denaturing gradient gel electrophoresis (DGGE; [33]). 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 later, to identify which primers in a degenerate primer pool interact with each template.

As part of this study, 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). 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. Finally, high-throughput amplicon sequencing was performed on an Illumina MiniSeq sequencer, enabling us to generate thousands of sequences per sample for robust quantitation of amplicons. Our study 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 were used only for 2 cycles – therefore locus-specific primer limitations were avoided.

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 1, Table S2** 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 1/10th 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 S1**). The 64 primers ("806F" primers) were grouped into four categories relative to each template: (i) primer with no mismatches (1 primer per template), (ii) primers with one mismatch (9 primers per template), (iii) primers with two mismatches (27 primers per template), and (iv) primers with three mismatches (27 primers per template) (**Table S2**). For each template mixture (A-E), a separate experiment was conducted using one of five primer pools (**Table 1**). 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 (**Table S3**). Primer theoretical melting temperatures were calculated using the OligoAnalyzer3.1 calculator [34], 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 [35]. 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 [36]. 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°C 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. One 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, and the pooled library was purified using an AMPure XP cleanup protocol (0.7X, vol/vol; Agencourt, Beckmann-Coulter) to remove short fragments. Pooled and cleaned amplicons were sequenced on an Illumina MiniSeq mid-output flow cell with 2x153 base reads, and with an approximate 30% phiX spike-in due to the extreme low complexity of the amplicons.

Deconstructed PCR (DePCR) Protocol

153 A two-stage deconstructed PCR (DePCR) method [7, 31] was also used to generate amplicons for next-
154 generation sequencing (**Figure 1)**. In this protocol, four primers are added to the first stage reaction,
155 including locus-specific primer pools containing 5' CS1 and CS2 linkers (pools i, ii, iii and iv as described;
156 each pool was added at 200 nM concentration), as well as Fluidigm Access Array Barcode Library
157 primers, containing Illumina sequencing adapters, a sample-specific 10 nucleotide barcode, and CS1 and
158 CS2 linkers at the 3' ends (added at 400 nM concentration). 2.5 ng of synthetic ST mixtures (A-E,
159 described above) was used for each 10 µL reaction. All reactions were performed using 2× MyTaq HS
160 Mix and reactions were conducted in 96-well plates. First stage thermocycling conditions were: initial
161 denaturation at 95°C for 5 minutes, followed by two cycles of 95°C for 30 seconds and either 45°C and
162 55°C for 20 minutes, followed by two cycles of 95°C for 30 seconds and 60°C for 2 minutes.
163 Subsequently, technical replicates from each experiment (*e.g.*, A1, A2, A3) were pooled together from
164 both annealing temperatures (16 reactions per pool). Pooled replicates were purified twice sequentially
165 using an AMPure XP cleanup protocol (0.7X, vol/vol) and eluted in 50 µL. Of this eluate, 20 µL were used
166 as template for amplification in the second stage reaction with P5 and P7 primers. Final volume for each
167 amplification reaction was 50 µL. Thermocycling conditions were 95°C for 5 minutes and 30 cycles of
168 95°C for 30 seconds, 60°C for 45 seconds, and 72°C for 90 seconds. Amplicons generated from second
169 stage reactions were again purified using an AMPure XP cleanup protocol (0.7X, vol/vol). Pooled and
170 purified amplicons from each experiment were quantified using Qubit fluorimetry (Qubit 4.0, Thermo
171 Fisher Scientific), and further pooled together to generate a final library. Pooled, cleaned amplicons
172 were sequenced on an Illumina MiniSeq mid-output flow cell with 2x153 base reads, and with an
173 approximate 30% phiX spike-in. Library preparation and sequencing were performed at the UIC
174 Sequencing Core (UICSQC).

175 ***Sequence Data Analysis***

Raw FASTQ files were merged using the software package PEAR [37] 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 [38]. 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 S1**. All possible primer-template interactions are shown in **Tables S2 and S3**. To calculate utilization profiles for all the samples, a mapping file, containing all possible unique combinations of 806F primers and recognition sequences were generated (**Table S4**). 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) [39]. The BIOM was rarefied to a depth of 7,000 counts per sample in the R programming environment [40] 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 [41] 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 [42] 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 [43] (Primer-E, Plymouth, UK). Ideal score (IS) analysis was performed using the *vegan* R package. The IS analysis was slightly modified from the formula described previously [7] 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. The IS has a range from 0 (perfect representation of the input template distribution) to 200.

Data Archive

Raw sequence data files were submitted to 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 S5**.

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 most reactions, various combinations of the 64 primers were used. When multiple primers were used, they were present in equimolar concentration. A full list of experimental conditions is shown in **Table 1**. 7-8 technical replicates were generated for each experimental condition.

The primary template was designed in a similar manner to synthetic templates described previously [7]. Briefly, the synthetic DNA sequences were based on the 16S rRNA gene sequence from a Gammaproteobacterium, *Rhodanobacter denitrificans* [44]. 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 ("Forward"). Synthetic template sequences at the second primer site ("Reverse") were identical for all reference templates and targeted by the 555R primer (**Table S1**). 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 2**). Each template, when compared to other templates, has variants in up to 3 positions, located at -2, -8, and -14 base positions, counting from the 3' end of the 806F primer annealing site. The -2, -8, and -14 positions represent 3', middle, and 5' mismatches, respectively (**Figure 2**, **Table S2**). 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 2**). Using DePCR, the sequence of the primer annealing to templates is retained during exponential amplification [7, 31], and in this experimental system is linked to a recognition sequence. In this manner, NGS amplicon sequence data were used to identify which templates were amplified and which primer annealed to each template molecule. 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 S1-S16**). 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 primer utilization profiles. In addition, analysis of similarity (ANOSIM; 9999 permutations) calculations were performed to determine if primer utilization 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, but only 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 (ANOVA) 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 after amplification 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) generated a template distribution profile most similar to that of the input template, as well as which annealing temperature within each method generated 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 1; Figures 3, S1-S5**), 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 S1-S5**).

Primer sequences (variants 1-64) were identified in each generated 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 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 both to genomic DNA templates and then later to DNA copies [7]. Due to tolerance of 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 input templates is lost. This is observed in all experiments with greater than one forward primer variant

(**Figures 5, 6, and S1-S16**). 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 1**).

Several patterns were observed when amplifying the single ST1 template with various primer pools (**Figure 3**). 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 (**Table S3**; 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. 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 the higher annealing temperature (**Figure 3**). 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 (**Table S3**).

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 3, 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 3**, 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 in experiments conducted using an annealing temperature of 45°C.

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 3**).

In the A1 experiment, only a single primer perfectly matching the ST1 template was included. However, we observed that approximately 4% of the DePCR reads 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 1**) 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 4**). DePCR was superior to the TAS for reproducing the expected template distribution in all experiments except for A1 (**Figure 4**). This was determined by calculation of the Ideal Score (IS), which represents a summation of the 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 S1-S16**). Higher IS 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 4**).

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 primer pool was used (**Figure 4**). 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 4**; **Table S2**). 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 1**) 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 5, 6 and S6-S9**. Results from the "C" experiment, with 10 unique templates but with ST1 at 1/10th concentration are shown in **Figures S10-S12**. Results from the "D" experiment, with four unique templates (including ST1 and three 3' single mismatch templates) at graduated concentrations are shown in **Figures S13-S14**. 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 S15-S16**.

Amplification method (DePCR or TAS) yielded significantly different PUPs in "B" experiments with 10 templates and varying number of primers (**Figure 5).** 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 5**). 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 5**). 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% of all reads, with the higher value occurring at the 55°C annealing temperature (**Figure 5**). 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.

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 6**). 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 6, S17, S18**). 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 6**). 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 6**). 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 6**). 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 5 and 6**). 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 6**).

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 5** represent average primer utilization across all templates. PUPs presented in **Figure 7** present primer utilization for each of the 10 templates in experiment B2 at 45° and 55°C annealing temperatures. In experiment B2, two patterns were observed in template-specific PUPs: (a) dominant annealing of perfect match primers and one or two other primers to templates (*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 7**). 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 (specifically, of the ten available primers in experiment B2, one primer matched the ST1 template perfectly, while the remaining nine primers each had a single mismatch with ST1). Conversely, for all other templates, there was a mixture of one perfect match, three single mismatch, and six 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 S2**).

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 [3] – 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 [16, 45]. As has been shown previously, templates with mismatches to primers can be difficult to detect, and mismatches close to 3' ends are particularly damaging [46, 47]. We previously developed a novel method for reducing PCR bias [7, 31], 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 [7, 31]. In the original study [7], we identified a novel source of PCR bias – namely, the concurrent 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 copy both original DNA templates and PCR-generated 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 [7]. 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 [31]. 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 Giovannoni [2] demonstrated that the evenness of amplification 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 Giovannoni [2]). 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 [7]. 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 conserved primer regions [48, 49]. 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 interactions and variable amplification efficiency by cycle number [7]. 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 [31].

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 [46]. Among other findings, Bru et al. [46] 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 [50]. 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. [47] 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 [45].

We demonstrate here that primer-template interactions favor perfect matches, but not overwhelmingly so. In fact, most annealing and copying in DePCR experiments was performed by 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 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.

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 [31]. 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 a 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 [51]. 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 acknowledge the support of the members of the University of Illinois at Chicago Sequencing Core for assistance in the study.

Figure Legends

Figure 1. 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 reactions 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. In the list of reaction mixture components, added primers are indicated by an asterisk.

Figure 2. 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 1**).

Figure 3. 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 were varied (**Table 1**). Shown are one-way clustered heatmaps of untransformed primer variant utilization during amplification with 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.

Figure 4. Effect of PCR methodology and annealing temperature on template profiles in amplification reactions utilizing a single primer. One-way clustered heatmaps of untransformed template 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 and **Table 1**. 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.

Figure 5. Effect of PCR methodology, annealing temperature, and primer pool on PUPs in experiments with ten templates. In experiments B1-4, all ten synthetic DNA templates were added to amplification reactions at equimolar concentrations, while primer pools varied (**Table 1**). One-way clustered heatmaps of untransformed PUPs during amplification with 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.

Figure 6. 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 1**. 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.

Figure 7. 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 6.** Primer utilization is shown for annealing temperatures 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.

Figures S1-S16. 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 1**. 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.

Figures S17-18. 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 S17** = 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 S18** = 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 1**. 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 Legends

Table 1. Description of templates and primers used in experiments conducted as part of this study.

Table S1. **Locus-specific primer sequences used in this study**

Table S2. 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 S3. **Rarefied biological observation matrix for all experiments.** Data were rarefied to 7,000 sequences per sample, and each experimental condition has 7-8 replicates. A total of 640 possible interactions are listed (10 templates x 64 primers), and numbers represent the numbers of reads matching each of the combinations. For each row of the BIOM, the number of mismatches between primer and template are shown, along with the position of mismatch, the mismatch sequence pairing, and the theoretical melting temperature of the primer. Reactions conducted with DePCR are highlighted in blue; no highlighting is used for TAS amplification reactions.

Table S4. **Mapping file used for creation of biological observation matrices.** This mapping files is used by the script described in the text and provided in Supplemental Materials 2.

Table S5. Metadata associated with all samples used in this study

Supplemental Materials

Supplemental Materials 1. Description of synthetic DNA template design and template sequences.

718 **Supplemental Materials 2. Script used for generation of BIOM files.**

719

720 **References**

721 1. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N: **Enzymatic amplification**
722 **of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell**
723 **anemia**. *Science* 1985, **230**(4732):1350-1354.
724 2. Suzuki MT, Giovannoni SJ: **Bias caused by template annealing in the amplification of mixtures**
725 **of 16S rRNA genes by PCR**. *Applied environmental microbiology* 1996, **62**(2):625-630.
726 3. Polz MF, Cavanaugh CM: **Bias in template-to-product ratios in multitemplate PCR**. *Applied*
727 *environmental Microbiology* 1998, **64**(10):3724-3730.
728 4. Pinto AJ, Raskin L: **PCR biases distort bacterial and archaeal community structure in**
729 **pyrosequencing datasets**. *PloS one* 2012, **7**(8):e43093.
730 5. Wagner A, Blackstone N, Cartwright P, Dick M, Misof B, Snow P, Wagner GP, Bartels J, Murtha
731 M, Pendleton J: **Surveys of gene families using polymerase chain reaction: PCR selection and**
732 **PCR drift**. *Systematic Biology* 1994, **43**(2):250-261.
733 6. Sipos R, Székely AJ, Palatinszky M, Révész S, Márialigeti K, Nikolausz M: **Effect of primer**
734 **mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting**
735 **bacterial community analysis**. *FEMS Microbiology Ecology* 2007, **60**(2):341-350.
736 7. Green SJ, Venkatramanan R, Naqib A: **Deconstructing the polymerase chain reaction:**
737 **understanding and correcting bias associated with primer degeneracies and primer-template**
738 **mismatches**. *PloS one* 2015, **10**(5):e0128122.
739 8. Ogino S, Wilson RB: **Quantification of PCR bias caused by a single nucleotide polymorphism in**
740 **SMN gene dosage analysis**. *The Journal of molecular diagnostics* 2002, **4**(4):185-190.
741 9. Hong S, Bunge J, Leslin C, Jeon S, Epstein SS: **Polymerase chain reaction primers miss half of**
742 **rRNA microbial diversity**. *The ISME Journal* 2009, **3**(12):1365.
743 10. Gohl DM, Vangay P, Garbe J, MacLean A, Hauge A, Becker A, Gould TJ, Clayton JB, Johnson TJ,
744 Hunter R: **Systematic improvement of amplicon marker gene methods for increased accuracy**
745 **in microbiome studies**. *Nature biotechnology* 2016, **34**(9):942.
746 11. Hansen MC, Tolker-Nielsen T, Givskov M, Molin S: **Biased 16S rDNA PCR amplification caused by**
747 **interference from DNA flanking the template region**. *FEMS Microbiology Ecology* 1998,
748 **26**(2):141-149.
749 12. Wilson KH, Blitchington RB: **Human colonic biota studied by ribosomal DNA sequence analysis**.
750 *Applied Environmental Microbiology* 1996, **62**(7):2273-2278.
751 13. Chandler D, Fredrickson J, Brockman F: **Effect of PCR template concentration on the**
752 **composition and distribution of total community 16S rDNA clone libraries**. *Molecular Ecology*
753 1997, **6**(5):475-482.
754 14. D'Amore R, Ijaz UZ, Schirmer M, Kenny JG, Gregory R, Darby AC, Shakya M, Podar M, Quince C,
755 Hall N: **A comprehensive benchmarking study of protocols and sequencing platforms for 16S**
756 **rRNA community profiling**. *BMC genomics* 2016, **17**(1):55.
757 15. Kennedy K, Hall MW, Lynch MD, Moreno-Hagelsieb G, Neufeld JD: **Evaluating bias of Illumina-**
758 **based bacterial 16S rRNA gene profiles**. *Applied environmental microbiology* 2014:AEM. 01451-
759 01414.

16. 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.
17. 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.
18. 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.
19. 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.
20. 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-+.
21. 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.
22. 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.
23. 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.
24. 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.
25. 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.
26. Pomp D, Medrano J: **Organic solvents as facilitators of polymerase chain reaction**. *Biotechniques* 1991, **10**(1):58-59.
27. Kanagawa T: **Bias and artifacts in multitemplate polymerase chain reactions (PCR)**. *Journal of bioscience and bioengineering* 2003, **96**(4):317-323.
28. 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.
29. 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.
30. 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.
31. Naqib A, Poggi S, Green SJ: **Deconstructing the Polymerase Chain Reaction II: an improved workflow and effects on artifact formation and primer degeneracy**. *PeerJ* 2019, **7**:e7121.
32. 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.
33. 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.

34. 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.
35. 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.
36. 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.
37. Zhang J, Kobert K, Flouri T, Stamatakis A: **PEAR: a fast and accurate Illumina Paired-End reAd mergeR**. *Bioinformatics* 2013, **30**(5):614-620.
38. 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.
39. 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.
40. Team RC: **R: A language and environment for statistical computing**. 2013.
41. 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.
42. Wickham H: **ggplot2: elegant graphics for data analysis**. *Stat Softw* 2010, **35**(1):65-88.
43. Clarke K, Gorley R: **Getting started with PRIMER v7**. *PRIMER-E: Plymouth, Plymouth Marine Laboratory* 2015.
44. 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.
45. Mao DP, Zhou Q, Chen CY, Quan ZX: **Coverage evaluation of universal bacterial primers using the metagenomic datasets**. *Bmc Microbiol* 2012, **12**.
46. 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.
47. 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.
48. 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.
49. 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.
50. Kwok S, Kellogg D, McKinney N, Spasic D, Goda L, Levenson C, Sninsky JJNar: **Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies**. 1990, **18**(4):999-1005.
51. Innis MA, Gelfand DH, Sninsky JJ, White TJ: **PCR protocols: a guide to methods and applications**: Academic press; 2012.

Figure 1(on next page)

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 reactions 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. In the list of reaction mixture components, added primers are indicated by an asterisk.

CS1_806F_v1 5′-ACACTGACGACATGGTTCTACAGGACTACCAGGGTATCTAAT-3′
CS2_555R 5′-TACGGTAGCAGAGACTTGGTCTCGGAATTACTGGGCGTAAAGG-3′
PE1-CS1 5′-AATGATACGGCGACCACCGAGATCTACACTGACGACATGGTTCTACA-3′
PE2-[BC]-CS2 5′-CAAGCAGAAGACGGCATACGAGATXXXXXXXXXXTACGGTAGCAGAGACTTGGTCT-3′
P5 5′-AATGATACGGCGACCACCGA-3′
P7 5′-CAAGCAGAAGACGGCATACGA-3′

Linker sequences [CS1 and CS2 linkers shown]
Locus-specific primer
Sample-specific barcode **Sequencing adapters**

Template DNA
Copied DNA from prior cycle
Copied DNA generated during current cycle
Inverse complement

Stage A: Polymerase copying with locus-specific primers containing linkers

2X Buffer
DNA template
*CS1_806F_pool
*CS2_555R
*PE-CS1
*PE-[BC]-CS2

1 1 1
1 1 1
1 1 1
1 1 1

95°C – 30”
45°C or 55°C – 20’ **2X**

95°C – 30”
60°C – 2’ **2X**

Pool all reactions together. Perform AMPure XP cleanup, 2X

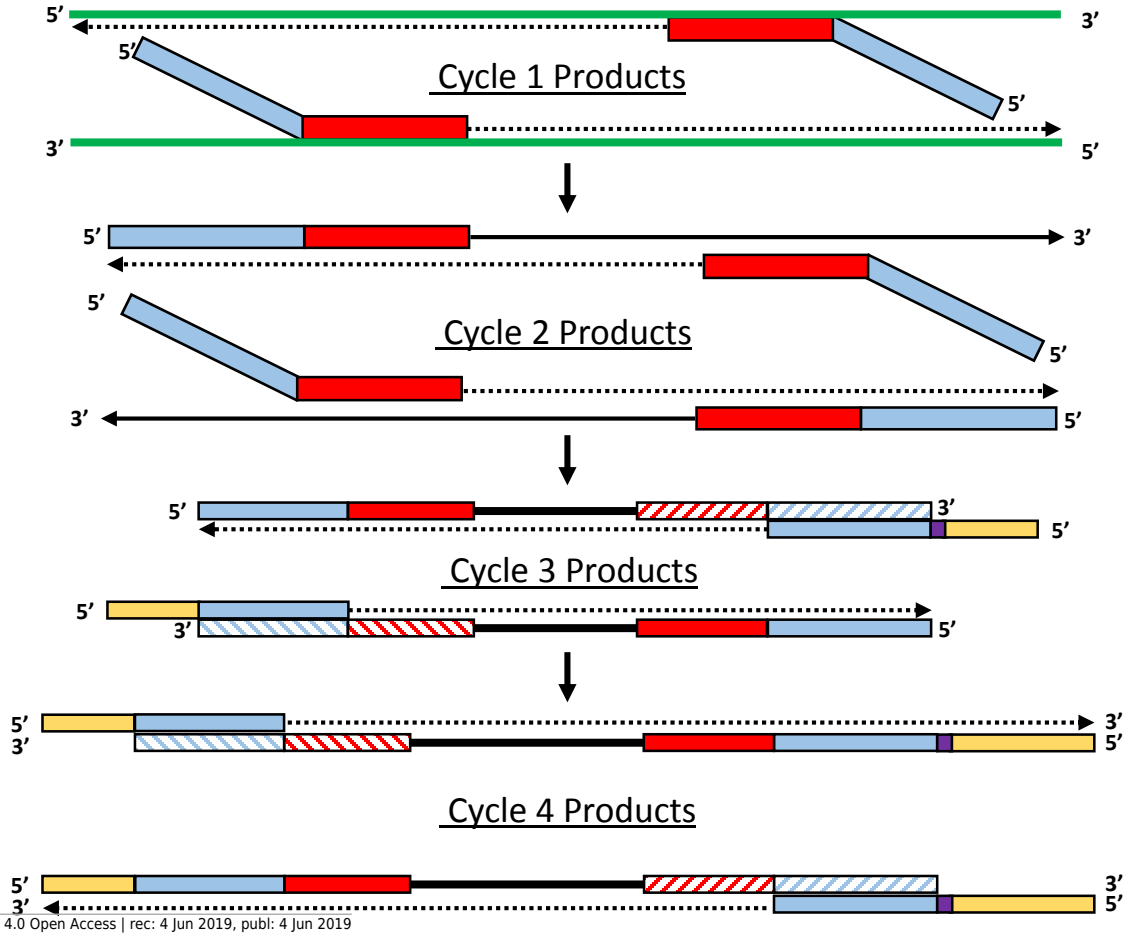
Stage B: PCR amplification with Illumina P5 and P7 primers

2X Buffer
Pooled template from step 1

*P5
*P7

2

95°C – 5’
95°C – 30”
60°C– 45” **30X**
72°C – 60”



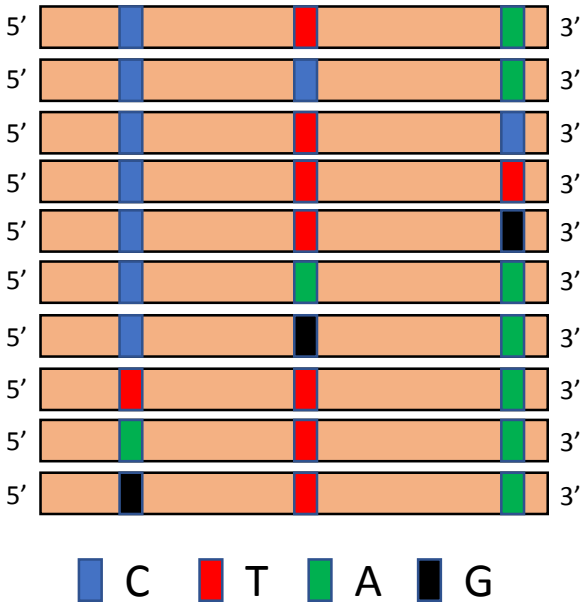
PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27781v1 | CC BY 4.0 Open Access | rec: 4 Jun 2019, publ: 4 Jun 2019

Figure 2(on next page)

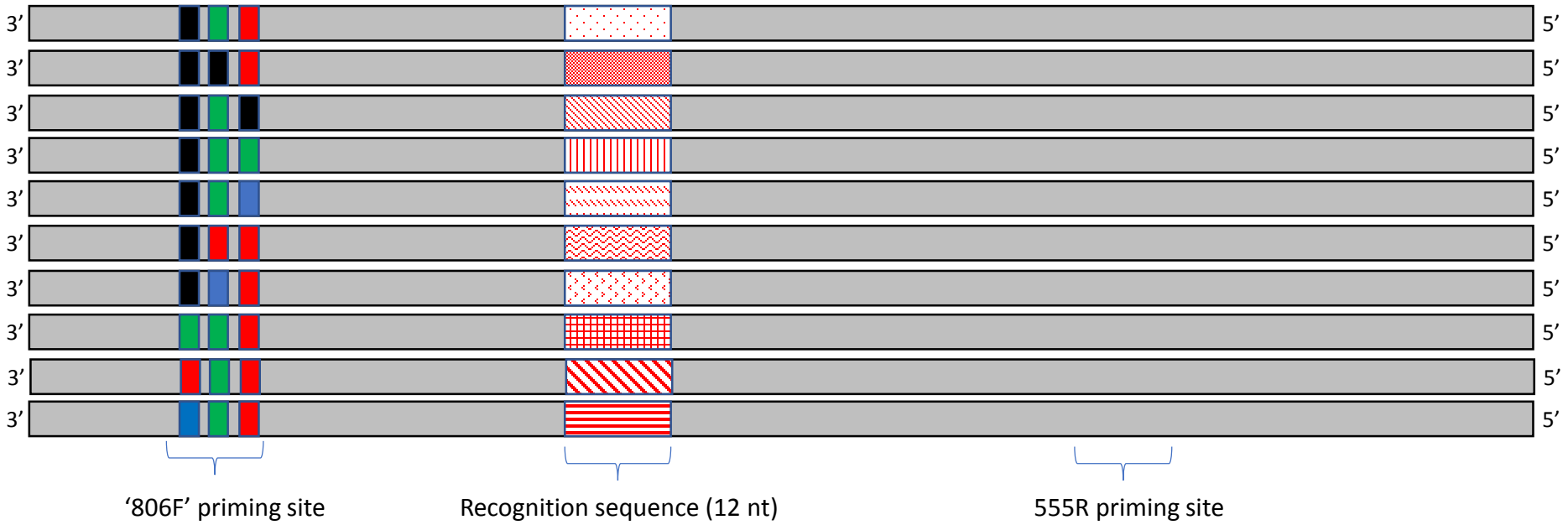
Primer, template and experimental design

(**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 1**).

Primer Design

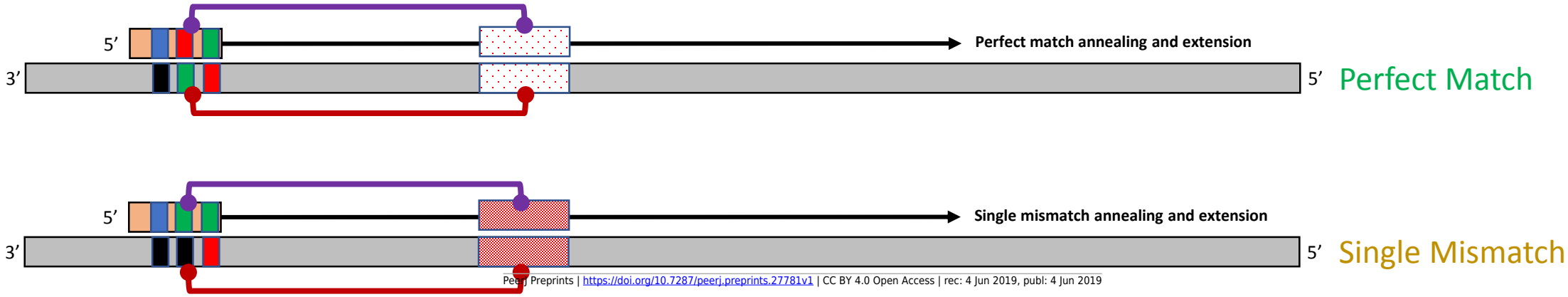


Template Design



Primer-Template Interaction

Primer used anneal to template is linked to a template recognition sequence by polymerase elongation.

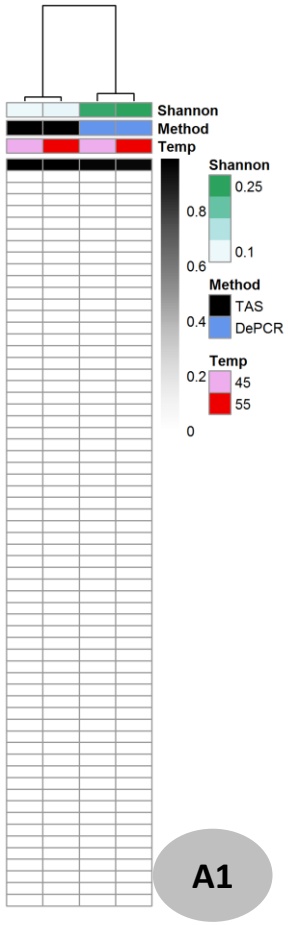


PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27781v1 | CC BY 4.0 Open Access | rec: 4 Jun 2019, publ: 4 Jun 2019

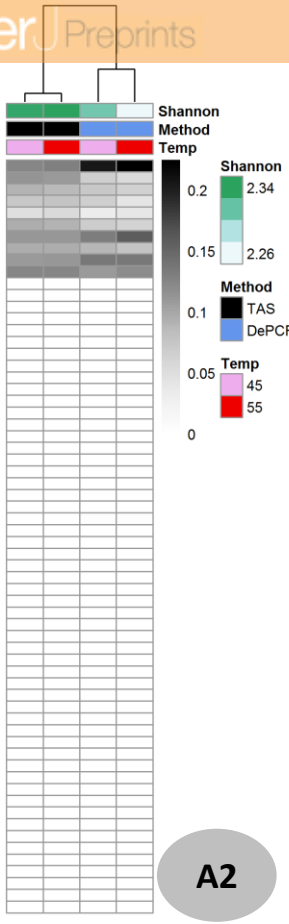
Figure 3(on next page)

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

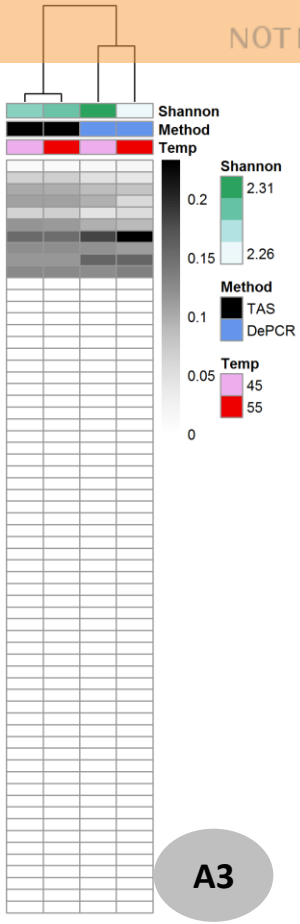
In experiments A1-6, only template ST1 was added to amplification reactions, while primer pools were varied (**Table 1**). Shown are one-way clustered heatmaps of untransformed primer variant utilization during amplification with varying primer pools ("A1" = 1 primer, "A2" = 10 primers, "A3" = 9 primers, "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.



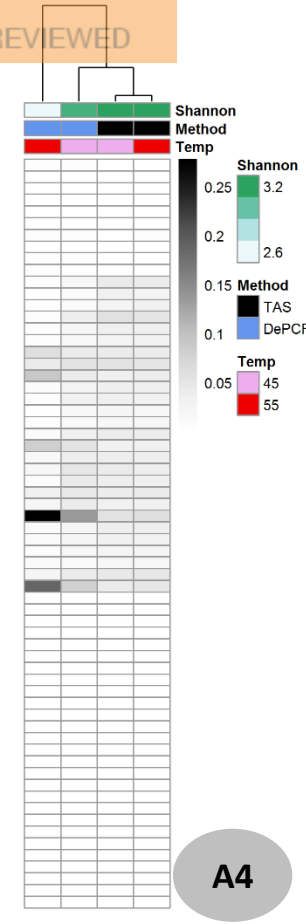
A1



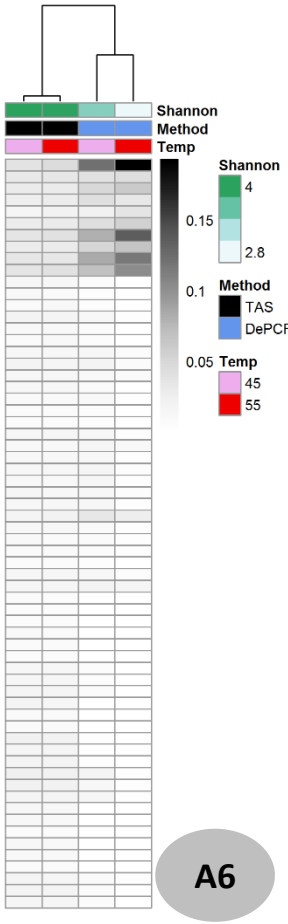
A2



A3



A4



A6

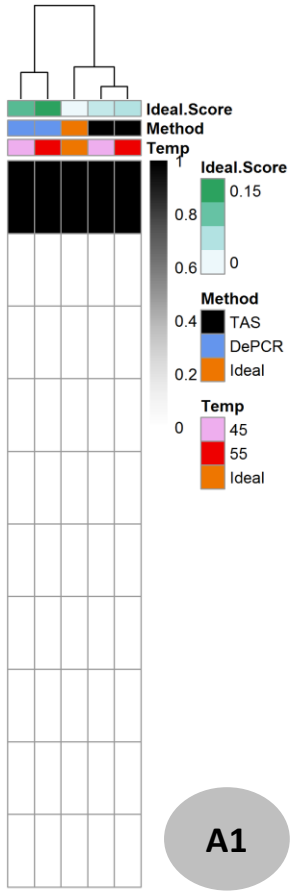
	DePCR45	DePCR55		DePCR45	DePCR55		DePCR45	DePCR55		DePCR45	DePCR55		DePCR45	DePCR55	
% Reads 0 MM	96.07	95.73		20.51	22.4	*	1.39	1.34		0.48	0.39		12.16	19.31	*
% Reads 1 MM	3.68	3.99		77.21	75.57	*	95.79	95.93		2.07	2.45	*	47.68	65.16	*
% Reads 2 MM	0.25	0.27		2.21	1.99	*	2.76	2.66		96.12	95.82	*	29.86	12.81	*
% Reads 3MM	0	0.01		0.06	0.04	*	0.06	0.07		1.34	1.34		10.31	2.72	*
% Reads with 3' MM				22.46	19.63	*	24.06	20.38	*				27.41	21.24	*
% Reads with Mid' MM				34.31	36.69	*	33.92	37.58	*				34.6	37.1	*
% Reads with 5' MM				43.23	43.68		42.03	42.04					37.99	41.66	*
Average Shannon	0.26	0.27		2.31	2.26	*	2.31	2.25	*	3.25	2.58	*	3.43	2.76	*

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27781v1 | CC BY 4.0 Open Access | rec: 4 Jun 2019, publ: 4 Jun 2019

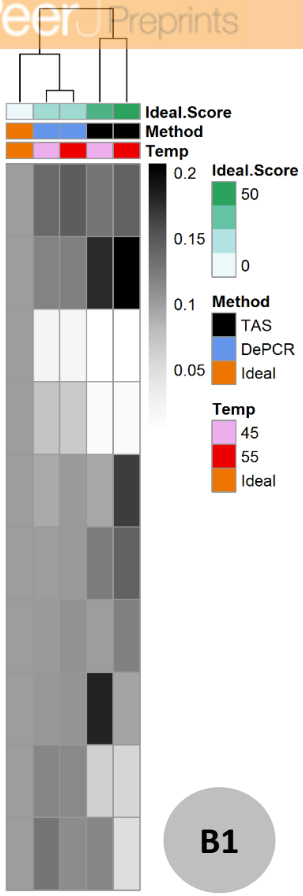
Figure 4(on next page)

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

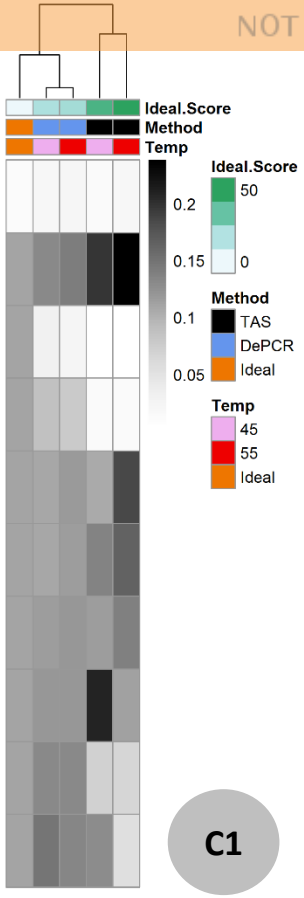
One-way clustered heatmaps of untransformed template profiling during amplification with a single primer (806F_v1) with a varying range of templates ("A1" = 1 template, "B1" = 10 templates, "C1" = 10 templates, "D1" = 4 templates, and "E1" = 4 templates) as described in text and **Table 1**. Samples (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.



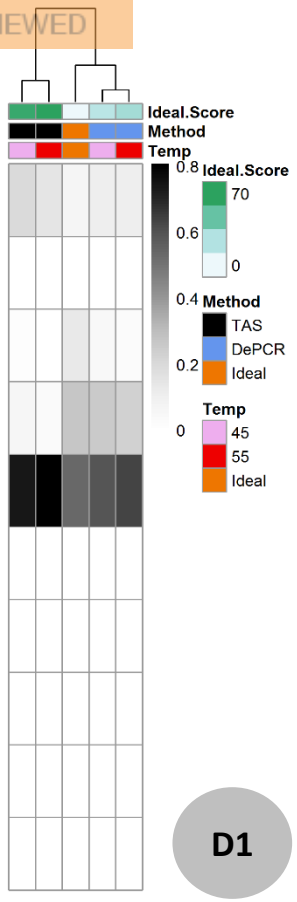
A1



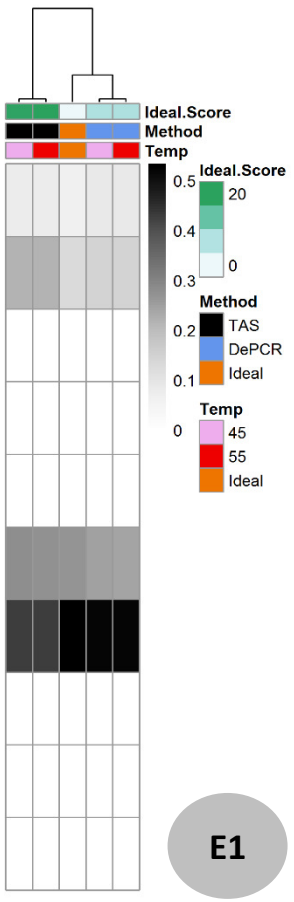
B1



C1



D1



E1

Average Ideal Score

TAS	0.06	*
DePCR	0.17	
TAS45	0.05	
TAS55	0.06	
DePCR45	0.15	
DePCR55	0.20	

TAS	50.80	*
DePCR	22.82	
TAS45	45.63	*
TAS55	55.97	
DePCR45	22.42	
DePCR55	23.23	

TAS	54.32	*
DePCR	22.45	
TAS45	48.72	*
TAS55	59.91	
DePCR45	21.04	*
DePCR55	23.87	

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

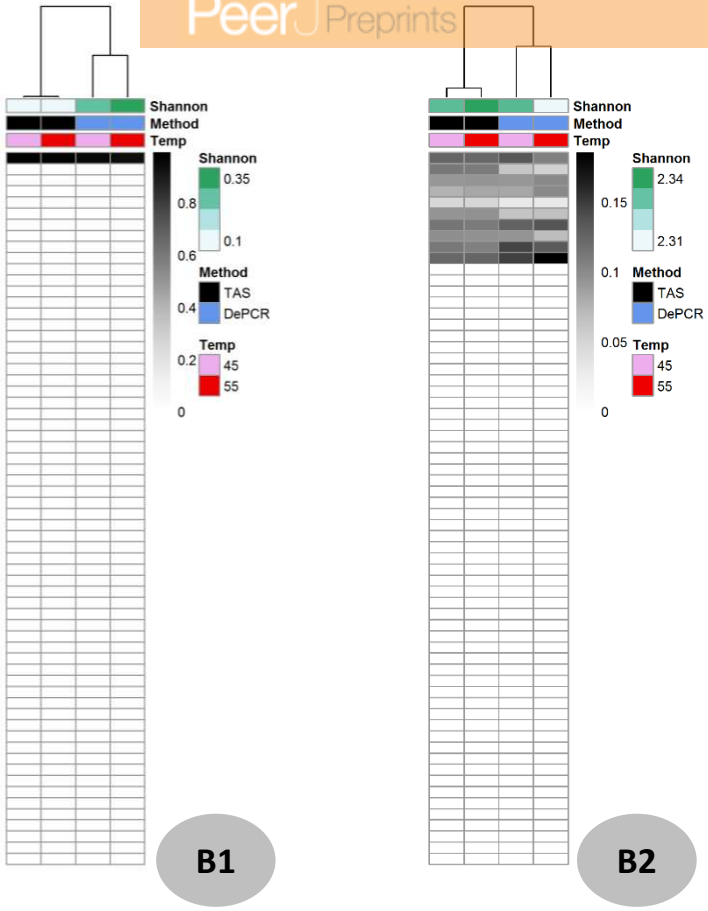
TAS	20.53	*
DePCR	7.30	
TAS45	20.28	
TAS55	20.78	
DePCR45	7.11	
DePCR55	7.47	

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27781v1 | CC BY 4.0 Open Access | rec: 4 Jun 2019, publ: 4 Jun 2019

Figure 5(on next page)

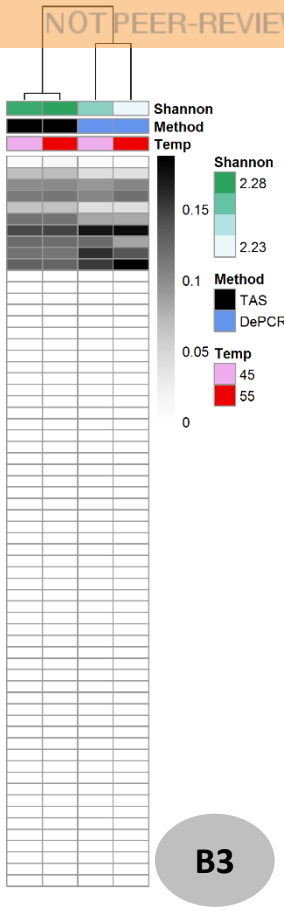
Effect of PCR methodology, annealing temperature, and primer pool on PUPs in experiments with ten templates.

In experiments B1-4, all ten synthetic DNA templates were added to amplification reactions at equimolar concentrations, while primer pools varied (**Table 1**). One-way clustered heatmaps of untransformed PUPs during amplification with 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.

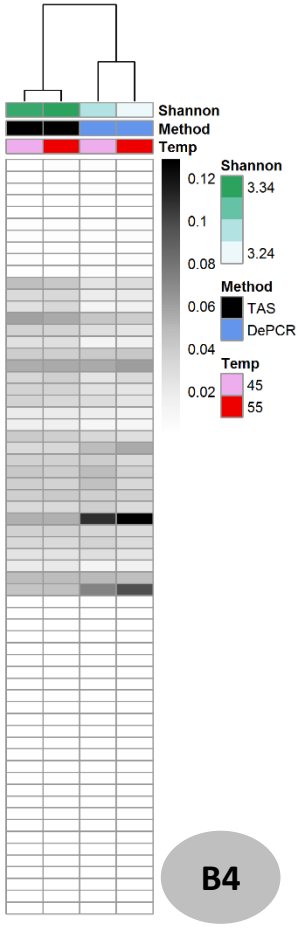


B1

B2



B3



B4

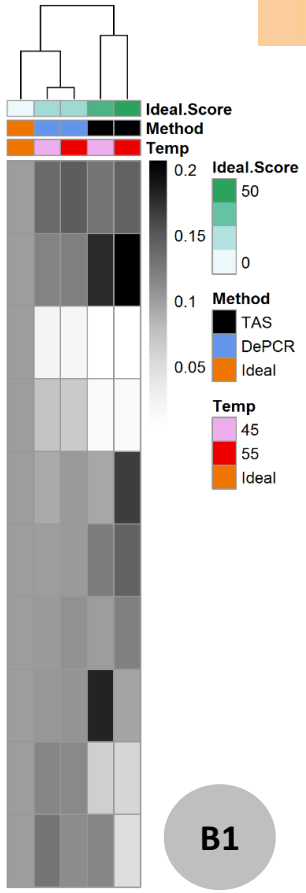
	DePCR45	DePCR55		DePCR45	DePCR55		DePCR45	DePCR55		DePCR45	DePCR55	
% Reads 0 MM	14.27	16.77	*	28.98	48.05	*	29.10	49.31	*	0.56	0.76	
% Reads 1 MM	84.21	81.72	*	46.54	43.24	*	42.17	40.86	*	49.53	74.13	*
% Reads 2 MM	1.50	1.48		24.26	8.63	*	28.46	9.73	*	45.13	22.37	*
% Reads 3MM	0.02	0.03		0.22	0.08	*	0.27	0.10	*	4.77	2.75	
% Reads with 3' MM	22.00	19.63	*	23.28	25.47	*	23.19	24.82	*	26.39	22.68	*
% Reads with Mid' MM	36.82	39.84	*	34.68	35.29	*	35.92	36.26		36.84	40.90	*
% Reads with 5' MM	41.18	40.52	*	42.04	39.24	*	40.89	38.91	*	36.77	36.42	
Average Shannon	0.27	0.36	*	**2.33**	**2.31**	*	**2.26**	**2.23**	*	**3.27**	**3.22**	*

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27781v1 | CC BY 4.0 Open Access | rec: 4 Jun 2019, publ: 4 Jun 2019

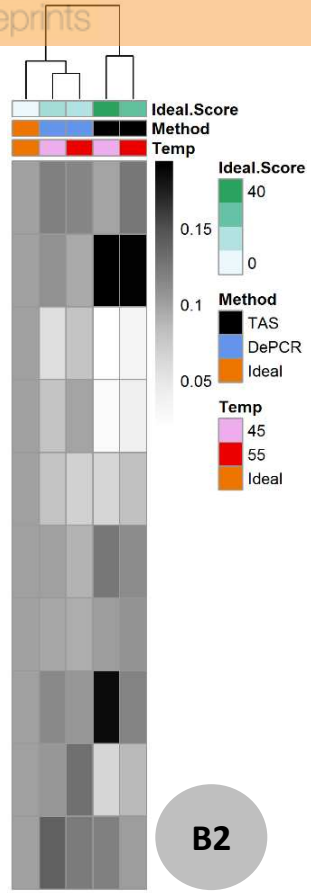
Figure 6(on next page)

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

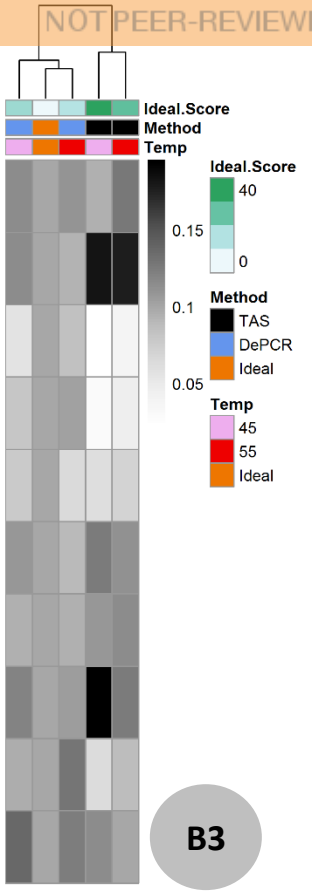
One-way clustered heatmaps of untransformed template utilization profiling during amplification of an equimolar pooling of all ten synthetic DNA templates and varying primer pools ("B1" = 1 primer, "B2" = 10 primers, "B3" = 9 primers, and "B4" = 27 primers) as described in text and **Table 1**. Samples (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.



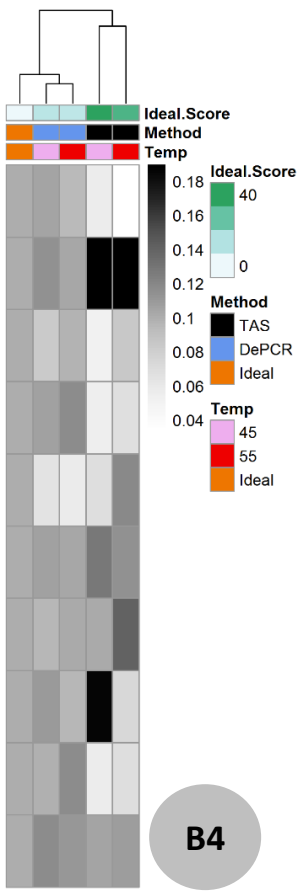
B1



B2



B3



B4

Average Ideal Score

TAS	50.80	*
DePCR	22.82	
TAS45	45.63	*
TAS55	55.97	
DePCR45	22.42	
DePCR55	23.23	

TAS	38.60	*
DePCR	16.55	
TAS45	45.48	*
TAS55	31.73	
DePCR45	18.00	*
DePCR55	15.10	

TAS	38.38	*
DePCR	16.69	
TAS45	45.24	*
TAS55	31.52	
DePCR45	19.30	*
DePCR55	14.41	

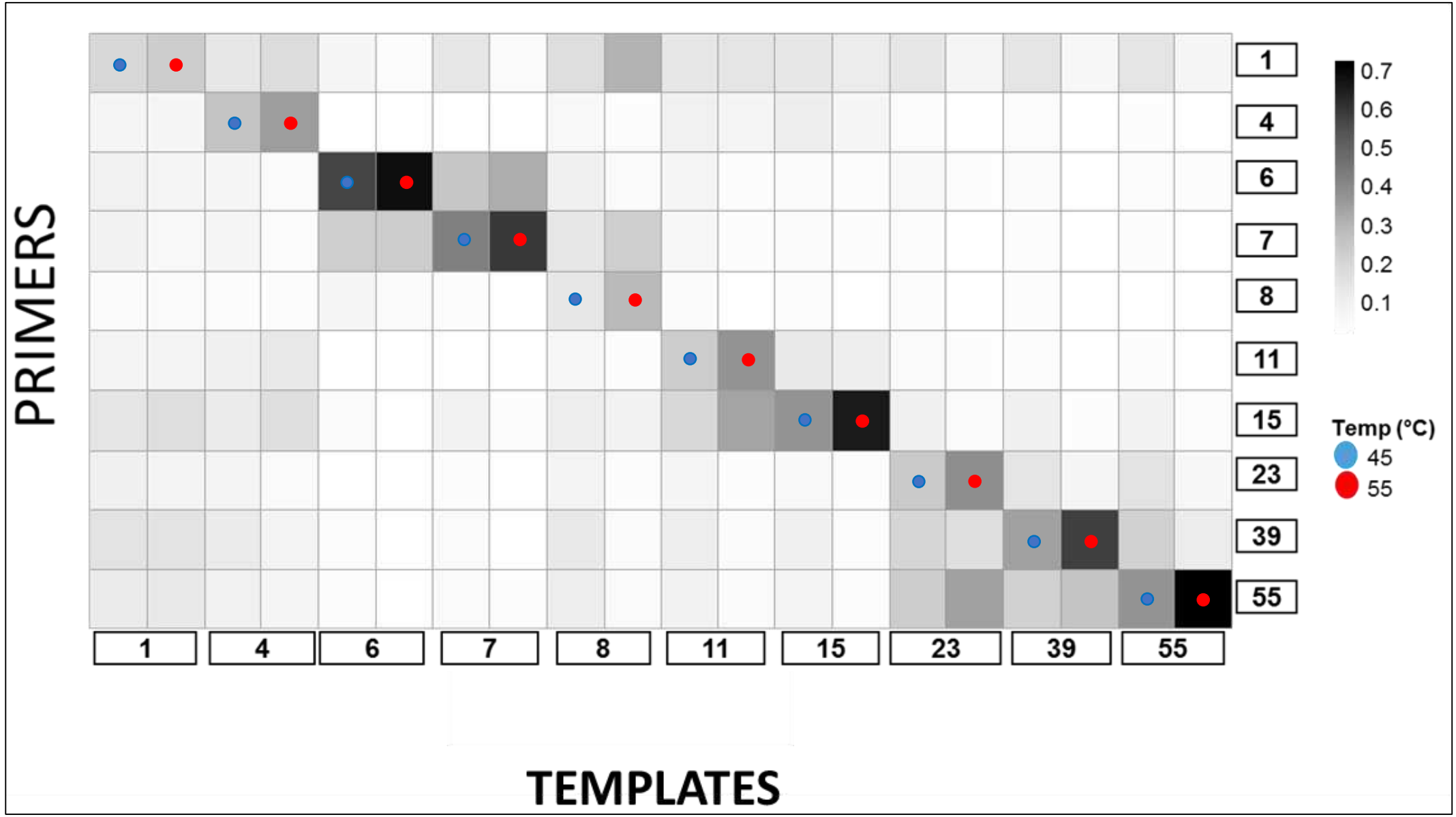
TAS	38.13	*
DePCR	11.53	
TAS45	42.14	*
TAS55	34.13	
DePCR45	11.82	
DePCR55	11.28	

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27781v1 | CC BY 4.0 Open Access | rec: 4 Jun 2019, publ: 4 Jun 2019

Figure 7 (on next page)

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 6.** Primer utilization is shown for annealing temperatures 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.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27781v1 | CC BY 4.0 Open Access | rec: 4 Jun 2019, publ: 4 Jun 2019

Table 1(on next page)

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

Detailed explanation of 16 experiments conducted in this study using synthetic DNA templates and oligonucleotides.

Experiment 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.

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.