environmental microbiology

Environmental Microbiology (2014) **16**(5), 1354–1365 doi:10.1111/1462-2920.12259

Exploiting extension bias in polymerase chain reaction to improve primer specificity in ensembles of nearly identical DNA templates

Erik S. Wright,[1,4] L. Safak Yilmaz,[3] Sri Ram,[4] Jeremy M. Gasser,[2] Gregory W. Harrington[2] and Daniel R. Noguera[2,5]*

Departments of [1]Bacteriology and [2]Civil and Environmental Engineering, University of Wisconsin – Madison, Madison, WI, USA.

[3]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA.

[4]Systems Biology Theme, Wisconsin Institute for Discovery, University of Wisconsin – Madison, Madison, WI, USA.

[5]Environmental Chemistry and Technology Program, University of Wisconsin – Madison, Madison, WI, USA.

Summary

We describe a semi-empirical framework that combines thermodynamic models of primer hybridization with experimentally determined elongation biases introduced by 3′-end mismatches for improving polymerase chain reaction (PCR)-based sequence discrimination. The framework enables rational and automatic design of primers for optimal targeting of one or more sequences in ensembles of nearly identical DNA templates. In situations where optimal targeting is not feasible, the framework accurately predicts non-target sequences that are difficult to distinguish with PCR alone. Based on the synergistic effects of disparate sources of PCR bias, we used our framework to robustly distinguish between two alleles that differ by a single base pair. To demonstrate the applicability to environmental microbiology, we designed primers specific to all recognized archaeal and bacterial genera in the Ribosomal Database Project, and have made these primers available online. We applied these primers experimentally to obtain genus-specific amplification of 16S rRNA genes representing minor constituents of an environmental DNA sample. Our results demonstrate that inherent PCR biases can be reliably employed in an automatic fashion to maximize sequence discrimination and accurately identify potential cross-amplifications. We have made our framework accessible online as a programme for designing primers targeting one group of sequences in a set with many other sequences (http://DECIPHER.cee.wisc.edu).

Received 5 July, 2013; revised 6 August, 2013; accepted 20 August, 2013. *For correspondence. E-mail noguera@engr.wisc.edu; Tel. (+608) 263 7783; Fax (+608) 262 5199.

Introduction

Polymerase chain reaction (PCR) is one of the most widely used technologies in the life sciences, and recent advances in digital (Baker, 2012) and microfluidic (White *et al.*, 2011) PCR promise to continue this trend. The exponential nature of PCR implies that small biases in amplification efficiency can be quantitatively translated into substantial differences in amplicon concentrations. While significant effort has been given to recognizing the effects of these biases (von Wintzingerode *et al.*, 1997; Pinto and Raskin, 2012), the potential for exploiting inherent PCR biases for sequence discrimination remains largely untapped. This sequence discrimination problem is relevant in many scenarios including genotyping and metagenomic profiling where it is desirable to target a specific subset of sequences among a pool of closely related templates. For example, it is often desirable to verify or quantify the abundance of a specific organism of interest in a sample that was previously characterized using next-generation sequencing. Although sequence discrimination scenarios are frequently encountered in research, to the best of our knowledge, the problem has not been studied systematically and quantitatively.

Because replication is initiated at the 3′-end of primers, recent efforts have focused on exploring the influence of 3′-terminal mismatches on amplification (Kwok *et al.*, 1990; Huang *et al.*, 1992; Day *et al.*, 1999; Ayyadevara *et al.*, 2000; Wu *et al.*, 2009; Stadhouders *et al.*, 2010). While it is generally true that such mismatches enable discrimination and primer design approaches incorporating terminal mismatches have been developed (Cha *et al.*, 1992; Li *et al.*, 2004; Fu *et al.*, 2008), precise criteria for their efficacy and broad validity to generalized discrimination problems have not been established. One challenge to such an approach is posed by the need to

consider diverse sources of bias in PCR including both hybridization efficiency and polymerase elongation efficiency. While hybridization efficiency is reasonably well described by equilibrium thermodynamic models, the elongation efficiency is a kinetic process that is more difficult to incorporate into primer design.

Another challenge to reliably employing 3′-end mismatches for increased specificity in PCR is unifying conflicting reports on the contribution of various 3′-end mismatches to amplification efficiency. For example, different studies (Huang *et al*., 1992; Christopherson *et al*., 1997; Day *et al*., 1999; Ayyadevara *et al*., 2000) provided widely disparate estimations of the effect of an A/C (primer/template) terminal mismatch, varying from comparable to the perfect A/T match (Day *et al*., 1999) to nearly complete inhibition of amplification (Ayyadevara *et al*., 2000). Likewise, different investigators have found diametrically opposite results with terminal C/A mismatches ranging from little or no amplification to minor cycle delays (Ayyadevara *et al*., 2000; Wu *et al*., 2009; Stadhouders *et al*., 2010). These contradictory results underscore the challenges with adopting a systematic approach to the sequence discrimination problem.

In this study, we develop a semi-empirical framework for sequence discrimination that combines experimentally determined elongation efficiency of primers with mismatches at or near their 3′-terminus with thermodynamic model predictions of primer/template hybridization efficiency. Careful decomposition of the bias into these two sources enabled us to design primers with improved specificity towards desired targets while resolving some of the contradictions in the literature mentioned above. To illustrate the utility of our framework, we design primers for sequence discrimination between two alleles differing at a single nucleotide (one target sequence, one base pair sequence difference) as well as selective amplification of 16S rRNA genes from minor constituents in an environmental sample (multiple target sequences, complex sequence differences). Our findings demonstrate that PCR bias can be systematically used for designing primers to maximize specificity and sensitivity to a target group of DNA sequences.

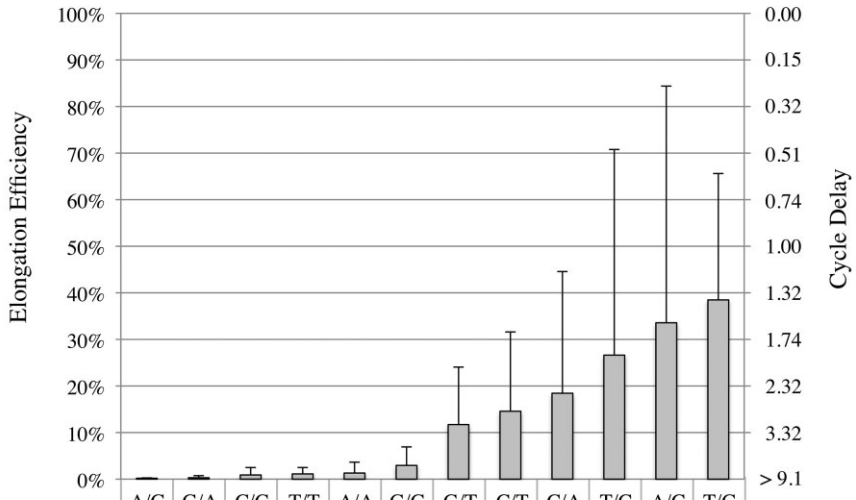
Results

Effect of 3′-terminal mismatch is dependent on neighbouring nucleotide

To systematically approach the sequence discrimination problem, we developed a quantitative framework for predicting amplification efficiency given a pair of primers and a DNA template. Here amplification efficiency is defined as the fraction of original DNA templates, which may have mismatches to the primer, that are replicated by the DNA polymerase in each cycle of PCR. We posited that the amplification efficiency depended on two main factors: the efficiency with which the primers hybridized to the template and the efficiency with which bound primers were elongated. The former was computed using standard thermodynamic models (Mathews *et al*., 1999), and the latter was modelled on quantitative PCR (qPCR) measurements of amplification using 171 primers mismatched to their respective DNA template. For each primer/template pair, we measured the fraction of the original mismatched templates that were elongated in each PCR cycle when annealing was performed at temperatures low enough that even mismatched primers would be fully hybridized, and used this as our metric for elongation efficiency (see Supplementary Methods).

We first looked at the set of 78 primer/template pairs with mismatches at the primer's 3′-terminal nucleotide that covered all 12 possible mismatch types (Table S1). Cycle delays caused by 3′-terminal mismatches ranged from an average of a single cycle up to nine cycles depending on the mismatch type (Fig. 1). The ranked order of elongation efficiencies generally agreed with prior observations. For instance, mismatches that have been reported as causing a severe impact on strand elongation in most studies (A/G, G/A, C/C and A/A) were also among the most disruptive mismatches in our experiments, while mismatches that have been generally reported to have weaker negative effects on PCR amplification (G/T, C/T, C/A, T/C, A/C and T/G) had the highest elongation efficiencies in our study. Interestingly, the four mismatch types (A/C, G/T, C/A, T/G) that upon amplification would result in a nucleotide transition (i.e. a mutation from a purine to a purine or from a pyrimidine to a pyrimidine) in the amplified product were among the six mismatch types with the highest elongation efficiency. This observation is consistent with base transitions being observed more frequently than base transversions in mutation events (Tindall and Kunkel, 1988). Symmetric mismatches (e.g. T/G versus G/T) showed a high degree of similarity (Fig. 1), with less than two cycles of separation on average, in corroboration with previous studies (Stadhouders *et al*., 2010).

Further investigation into the variability of elongation efficiencies revealed the effect of the nucleotides adjacent to the terminal mismatch (Fig. 2). Thymine or guanine bases in the penultimate (2nd) primer position greatly impeded elongation of almost all terminal mismatch types, and were therefore the most obstructive bases to have neighbouring a terminal mismatch ($P < 0.001$). Moreover, for each type of neighbouring nucleotide, the efficiency of elongation depended on the mismatch type. For example, the elongation efficiency of a primer containing a C/T terminal mismatch ranged from nearly zero for a T or G nearest neighbour, to moderate efficiency (> 20%) for a C or A nearest neighbour. The antepenultimate (3rd) base



	A/G	G/A	C/C	T/T	A/A	G/G	G/T	C/T	C/A	T/C	A/C	T/G
Ayyadevara *et al*. (2000)	+	+	+	+	+	+	++	+	+	+	+	+
Day *et al*. (1999)	+	++	++	++	+	++	+++	++	++	++	+++	++
Huang *et al*. (1992)	–	–	–	+	+	–	++	+++	++	++	++	++
Kwok *et al*. (1990)	+	+	+	+++	++	+++	+++	+++	+++	+++	+++	+++
Li *et al*. (2004)	+	+	+	+++	+	+	+++	+++	+++	+++	+++	+++
Stadhouders *et al*. (2010)	+	+	+	++	+	+	+++	++	+++	++	+++	+++
Wu *et al*. (2009)		–			–				–			

Fig. 1. Comparison of measured elongation efficiencies to previously published results. Mean elongation efficiencies of all possible 3′-terminal mismatches (primer/template), and qualitative comparison to previously published results (–, no amplification; +, poor; ++, moderate; and +++, high amplification). Error bars indicate standard deviation between mean elongation efficiencies of different primers tested with the same 3′-terminal mismatch. Cycle delay is shown as the delay in amplification of a mismatched template relative to a perfect match template of equivalent concentration.

also had an effect on elongation efficiency, with an adenine base on the primer causing less amplification delay ($P < 0.005$). However, a lack of every combination of antepenultimate base and mismatch type prevented a full analysis of this position's effect (Table S1). Analysis of the primer's preantepenultimate (4th) nucleotide did not indicate any effect on elongation efficiency. Likewise, there was no statistical significance of the template nucleotide after the primer's terminus (i.e. the first nucleotide added).

Mismatches, insertions and deletions near the 3′-terminus generally have less effect on elongation efficiency than mismatches at the 3′-terminus

Next, we proceeded to determine the positional effects of mismatches on elongation efficiency. We tested a set of 46 primers (Table S2) with single mismatches between the 2nd and 7th position from the 3′-end (Fig. 3). In most cases, terminal mismatches caused greater delays than non-terminal mismatches. In accordance with the findings

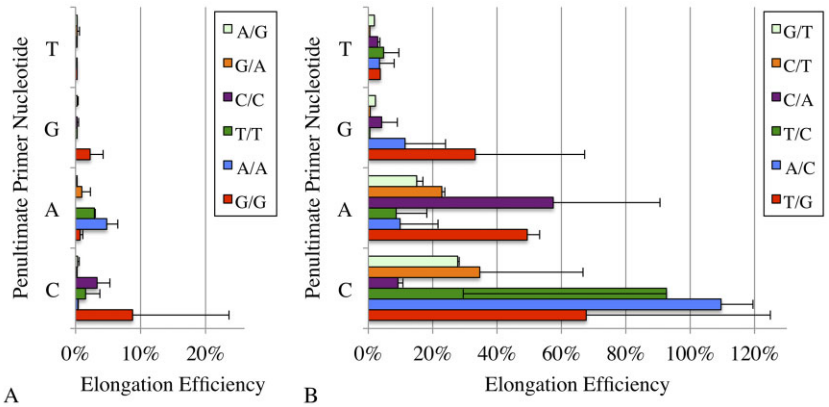


Fig. 2. Dependence of terminal mismatch elongation efficiency on neighbouring nucleotide. Variation in elongation efficiencies by mismatch type (primer/template) according to penultimate primer nucleotide for (A) the most disruptive and (B) least disruptive mismatch types. Error bars show standard deviations between different primer sequences and templates tested.

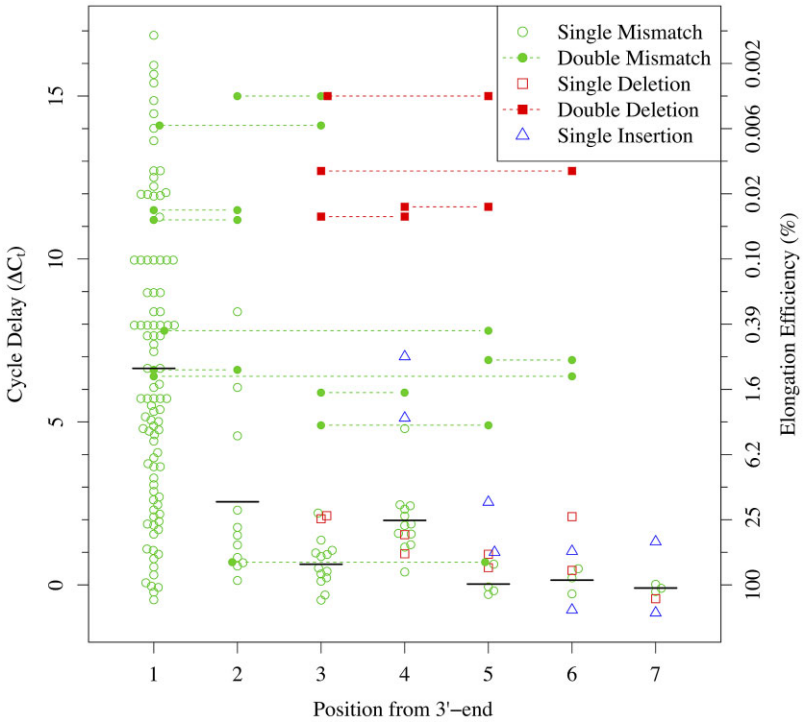


Fig. 3. Elongation efficiency's dependence on position for single and double mismatches, insertions and deletions. Measured cycle delay for each individual primer sequence according to mismatch position from 3′-terminus of primer. Position 1 is defined as the primer's 3′-end; position 2 is the penultimate primer position, and so forth. For insertions and deletions, the position marks where the nucleotides were added or deleted, respectively, in the perfect match primer sequence. A horizontal solid black line shows the mean value for single mismatches in each position. Double mismatches and double deletions are depicted with a dashed line connecting the two affected positions on the same primer. Since hybridization efficiency was maximized by using a low annealing temperature, elongation efficiency (right axis) is directly interconvertible to cycle delay (left axis).

of Stadhouders and colleagues (2010), delays were observed up to the 4th position, and largely disappeared for mismatches placed in the 5th to 7th position. Mismatches located in the 4th position caused a greater delay than mismatches located in the 3rd position ($P < 0.0002$). The positional dependency corresponds closely to the footprint of *Taq* DNA polymerase, which has been estimated at five nucleotides in length (Johnson and Beese, 2004). However, it is unknown why a mismatch in the 4th position was more detrimental to elongation than in the 3rd position. Possibly due to the relatively small sample size in comparison to the number of potential permutations, we were unable to find a significant correlation with other factors such as mismatch type or nearest neighbour base-pairings.

Unlike single mismatches, the effect of double mismatches was apparent beyond the 4th position (Fig. 3). The cycle delays of double mismatches were generally greater than the product of their individual mismatches' elongation efficiencies. The observed amplification of double mismatched primers in some experiments differ with those of Stadhouders and colleagues (2010) in which no double mismatches were found to elongate. However, because the experiments by Stadhouders and colleagues were performed closer to melt temperature whereas our experiments were performed at lower temperatures (Supplementary Methods), their results may reflect both decreased elongation efficiency and a large drop in hybridization efficiency caused by internal mismatches.

Next, we tested whether insertions and deletions (indels) near the 3′-end would induce multiple mismatches by shifting the entire neighbouring 3′-subsequence and therefore have a greater effect than double mismatches (Table S2). We compared the elongation efficiencies of 21 primers with nucleotides either added to or removed from positions 3–7. Interestingly, indels in most positions were elongated with a delay similar to that of a single mismatch. Nevertheless, double deletions in the 3rd to 6th positions all exhibited delays greater than 11 cycles.

Elongation time and type of polymerase affect the elongation efficiency of terminal mismatches

It has been established that mismatch incorporation into an elongating DNA strand is dependent on the kinetics of mismatch incorporation with respect to the duration of time that the polymerase remains active at the 3′-hydroxyl group (Huang *et al.*, 1992). In order to determine the degree to which elongation efficiency increases with additional elongation time, we created duplicate experiments with 14 primers representing a variety of terminal mismatches (Table S1). Elongation time was either 10 s or 180 s instead of the typical 30 s used in other experiments, and we tested templates with short PCR products that could easily be elongated within 10 s. On average, the delay in amplification decreased by 2.7 cycles when more time was given for elongation ($P < 0.0001$). The

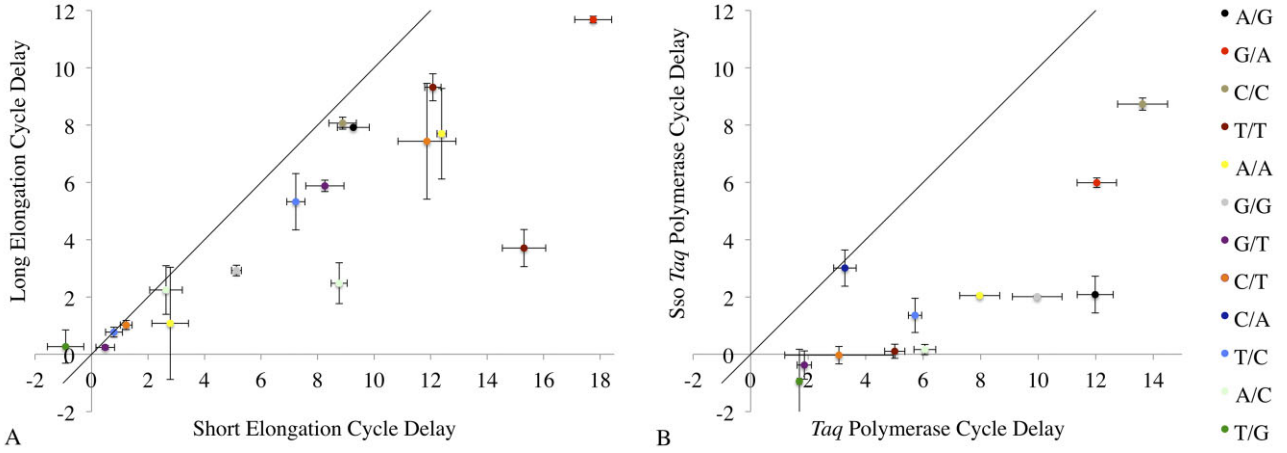


Fig. 4. Effect of increased extension time or polymerase processivity on mismatched cycle delays. Comparison of cycle delays for primers with 3′-terminal mismatches relative to their perfect match counterpart.
A. Extension time was either short (10 s) or long (180 s).
B. The DNA polymerase used was either Sso-*Taq* or standard *Taq* while other reagent conditions were held constant. Error bars show the standard deviation between three replicate wells. The solid line follows the line of identity. The legend on the right describes colouring of each mismatch type (primer/template).

effect of elongation time was more substantial for the mismatched primers that resulted in the largest cycle delays (Fig. 4A).

The type of DNA polymerase is another factor that is commonly accepted to influence amplification efficiency. Unlike *Taq*, polymerases with proofreading activity are capable of removing terminal mismatches to initiate elongation from a truncated primer (Yuryev, 2007), and are therefore rarely used in qPCR. In contrast, Sso-*Taq* was created by covalent addition of a dsDNA binding protein, which increases the processivity of *Taq* DNA polymerase without compromising catalytic activity or altering *Taq*'s lack of proofreading (Wang *et al*., 2004). To test the hypothesis that Sso-*Taq* would more efficiently extend 3′-terminal mismatches, we tested a set of 12 primers (Table S1) using both standard *Taq* polymerase and the Sso-*Taq* polymerase while all other experimental and reagent conditions were held constant. The use of Sso-*Taq* decreased cycle delays by an average of 4.8 cycles (Fig. 4B) relative to standard *Taq* ($P < 0.0002$).

Rational primer design improves specificity in the discrimination of a single nucleotide difference using qPCR

Having established the parameters influencing elongation efficiency, we incorporated them into an empirical model (Supplementary Methods) and completed our amplification efficiency prediction framework. We then sought to use the framework to design primers that can distinguish between two templates differing at a single nucleotide polymorphism (SNP). We selected the wild-type and R172K alleles of the human *IDH2* gene cloned into plasmids as our test templates, and compared four different primer design strategies (Table S3) for genotyping these two alleles. The R172K mutation is a rare variant of the *IDH2* gene that has been implicated in human gliomas (Yan *et al*., 2009).

In the first strategy, we used the framework to design primers with the SNP in the centre of either a forward or reverse primer where it was predicted to have the greatest negative effect on hybridization efficiency, but would not compromise elongation efficiency. We amplified an equal concentration of each allele using a mismatched primer together with a corresponding perfect match forward or reverse primer. We used an annealing temperature of 64°C so that the target template would have a high hybridization efficiency (predicted as > 80%) and the non-target template would have low hybridization efficiency (predicted as < 10%). For the four primers tested, this strategy resulted in modest 1.4–3.4 cycle delays in amplification of the non-target allele relative to the target allele (Table S3). For comparison, a 10% relative efficiency is equivalent to a 3.3 cycle delay, which is insufficient for adequate discrimination by qPCR.

In the second strategy, we used the framework to design primers with a mismatch at the 3′-terminus of the forward or reverse primer, resulting in a G/T, T/G, A/C or C/A mismatch with a C or A nearest neighbour depending on the allele and strand (positive or negative) that the primer targeted. These mismatch types and corresponding penultimate nucleotides were previously measured with high elongation efficiencies (Fig. 2), and therefore, this SNP represented the greatest sequence discrimination problem at this scale. With these terminally mismatched primers, we observed delays of 2.2, 7.6, 4.7 and

5.9 cycles, for the G/T, C/A, A/C and T/G mismatches respectively, showing in most cases an improvement in specificity over the first strategy. For comparison, a 1% relative efficiency would be equivalent to a 6.6 cycle delay, or about a 100-fold enrichment of the target template relative to the non-target template if both were present at equal concentration.

In order to develop a robust SNP discrimination method, we sought to construct primers that could provide a delay of greater than 10 cycles, corresponding to over 1000-fold ($2^{10} = 1024$) enrichment of the targeted allele. We therefore tested a third primer design strategy in which we purposely placed an additional mismatch to both the wild type and SNP alleles at the 6th position from the 3′-end. This interior mismatch was predicted to slightly decrease hybridization efficiency and not affect elongation efficiency of the target allele. However, according to our observations (Fig. 3), we hypothesized that this additional mismatch would greatly decrease elongation efficiency of the non-target allele as it was now designed to have a disruptive double mismatch near the 3′-end. Indeed, adding the mismatch in the 6th position resulted in delays of 11.6–12.7 cycles for each non-target allele relative to the target allele, proving that the induced mismatch in the 6th position was the most effective design strategy tested. It is worth noting that an induced mismatch will affect hybridization efficiency to the target, and the annealing temperature may need to be optimized with a standard temperature gradient experiment. Although it is preferable not to design a primer with mismatches to the target template, in cases where adequate specificity cannot be achieved with terminal mismatches alone (strategy #2), the induced mismatch may offer a means of obtaining the desired specificity.

An approach where the induced mismatch is placed in the penultimate primer position, termed TaqMAMA, has been suggested previously in literature (Li *et al.*, 2004). For the fourth strategy, we designed primers with tandem mismatches at the 3′-end that have previously been described as most detrimental to non-target amplification. This approach resulted in cycle delays between 9.0 and 13.7 cycles, which was comparable to placing the induced mismatch at the 6th position. However, in one case, the penultimate mismatch also delayed amplification of the target allele by over four cycles, and this could not be alleviated by lowering the annealing temperature from 64°C to 60°C. This delay was expected based on our studies of non-terminal mismatches (Fig. 3), which showed that single mismatches in the 2nd position often substantially lowered elongation efficiency. In contrast, mismatches in the 6th position did not substantially delay amplification of the target allele once the annealing temperature was lowered to compensate for the decrease in hybridization efficiency caused by the single mismatch.

Owing to the short amplicon size of about 50 base pairs, we repeated the experiment by lowering the time given for annealing from 30 s to 15 s and the time of the elongation step from 30 s to 5 s. This resulted in increased cycle delays in most cases as was anticipated based on earlier experiments (Fig. 4A). Taken together, these results indicate that a combination of reduced elongation time and induced mismatch are sufficient to attain more than a 10 cycle delay based on a single nucleotide difference, which is equivalent to less than 0.1% amplification efficiency of the non-target allele. Therefore, placing the SNP at the 3′-terminus and inducing a second mismatch at the 6th position resulted in more than two orders of magnitude improvement in fold specificity over simply placing the mismatch in the centre of the primer.

Incorporation of elongation efficiency into high throughput design improves primer specificity

Having used our framework for SNP discrimination, we next asked whether we could expand the scope of our approach to specifically target a subset of templates in a pool of many nearly identical ones, and whether incorporating elongation efficiency into primer design improved performance in complex sequence discrimination scenarios. We used our framework to develop a programme named Design Primers, which selects the most specific primers for amplifying a specified set of targets in the presence of an ensemble of different non-target DNA sequences. This programme is described in detail in Supplementary Methods and has been made available as part of the DECIPHER package for R (Wright, 2012) and also online (http://DECIPHER.cee.wisc.edu). The algorithm was designed to compare all possible combinations of potential forward and reverse primers and select the primer set that produced the smallest potential amplification of non-target groups.

We applied the programme to the design of genus-specific primers for each of 1834 bacterial and 109 archaeal genera comprising a total of 1 696 150 16S rRNA sequences in the Ribosomal Database Project (RDP) (Cole *et al.*, 2009). As input parameters, we chose to design primers of length 17–26 nucleotides with up to four permutations required to achieve at least 90% coverage of sequences belonging to the target genus. Primers' lengths were adjusted to achieve high (> 80%) efficiency with the target group at an annealing temperature of 64°C and appropriate reagent concentrations. The best scoring set of forward and reverse primers were chosen in which resulting amplicon size would be between 300 and 1200 nucleotides. We defined potential cross-amplifications as any non-targets with a predicted delay of less than 10 cycles for the same initial concentration of template (≥ 0.1% relative efficiency). To estimate

an upper bound to the number of genera for which it is possible to prevent cross-hybridization, we first performed a hypothetical simulation that assumed any mismatch was able to completely block amplification. We found that cross-amplification of one or more other genera could not be prevented for 21% of genera, illustrating the inherent challenges underlying genera-level discrimination.

For comparison, the algorithm was run once using estimated elongation efficiencies (strategy #2), and a second time assuming 100% elongation efficiencies for all mismatched primers, which would correspond to discrimination using hybridization efficiency alone (strategy #1). Making use of elongation efficiencies, the algorithm was able to find primer sets with no predicted cross-amplifications for 47% of genera and primer sets for 70% of genera with five or fewer potential non-target cross-amplifications. When elongation efficiency of terminal mismatches was not considered, the algorithm found primer sets for only 34% of genera with zero non-target amplifications and 58% of genera with five or fewer non-target amplifications, thus demonstrating the value of using elongation efficiencies in the design of group-specific primers and in extending the scope of PCR-based methods for complex sequence discrimination tasks. In total, employing elongation efficiency predictions improved the specificity scores of 59% of the designed primer sets over the use of hybridization efficiency alone.

Next, the algorithm was modified to induce a mismatch in each primer at the 6th position from the 3′-end (strategy #3). For each candidate primer, the mismatch type was chosen that had the highest predicted hybridization efficiency to the target allele, which was required to be at least 10%. The set of all potential primers was then scored for specificity in the same manner as described above. With an induced mismatch, the algorithm was able to find primer sets with no predicted cross-amplifications for 66% of genera, and five or fewer potential cross-amplifications for 85% of genera. This represented a 19% improvement over the use of elongation efficiencies without an induced mismatch, and brought the results much closer to the theoretical upper limit. In either case, it was not possible to find specific primer sets for a number of genera, which was anticipated given the highly conserved characteristics of the 16S rRNA gene. The set of designed primers for each genus is available online (http://DECIPHER.cee.wisc.edu).

The designed genus-specific primers successfully enrich for their respective target genus in the presence of numerous non-target sequences

In order to experimentally validate the effectiveness of the designed primers, we used four genus-specific primer sets to amplify DNA extracted from a water sample collected from the Sacramento Delta in California. Initially, we used universal 16S primers and 454 pyrosequencing to determine the microbial community composition available for targeting with genus-specific primers (Fig. 5A). We classified the resulting sequences using the RDP classifier (Lan *et al*., 2012) to determine the microbiota present in the sample. Using the primer sets we designed previously with elongation efficiency (strategy #2), we targeted four genera representing minor fractions of the three most abundant phyla present in the water sample, which were *Bacteroidetes* (60%), *Proteobacteria* (21%) and *Actinobacteria* (10%). The genera chosen for targeting were the actinobacteria *Arthrobacter*, the bacteroidetes *Emticicia* and *Ohtaekwangia*, and the proteobacteria *Escherichia*, representing 1, 20, 16 and 1 out of the 11 379 sequences obtained using universal primers respectively.

To determine the efficacy of the primers targeting these four genera (Table S4), we amplified the environmental sample and then pyrosequenced the resulting amplicons. All four genus-specific primer sets overwhelmingly enriched for their target genus (Fig. 5). For example, all of the 2130 sequences obtained using *Emticicia* primers were classified as *Emticicia*, and 99.3% of the 1744 sequences obtained using *Arthrobacter* primers were classified as *Arthrobacter* (Fig. 5B). In the case of *Escherichia*, all of the 2883 sequences could be classified into the family *Enterobacteriaceae*. However, it was not possible to reliably refine the classification to the genus-level using the short amplicon sequences as has been previously described in literature (Mizrahi-Man *et al*., 2013). In total, 27.0% of the 2883 sequences most closely matched reference sequences belonging to *Escherichia*, and an additional 69.8% matched multiple reference sequences in the family *Enterobacteriaceae* including the genus *Escherichia* (Fig. 5C). It should be noted that in our original set of 11 379 sequences obtained with universal primers, we only detected the *Escherichia* genus in the family *Enterobacteriaceae*, suggesting that most of these unclassifiable sequences indeed belonged to *Escherichia*.

The 1095 sequences obtained with *Ohtaekwangia* primers were mostly classified (66.5%) as belonging to the target genus. The remaining sequences also belonged to the *Bacteroidetes* phylum, with most of them classified as part of the order *Sphingobacteriales*, which includes the *Chitinophagaceae* family (Fig. 5D). The primer design algorithm predicted that the *Ohtaekwangia* primers would cross-amplify with sequences belonging to genera from this family, and this cross-amplification was further expected because the sequences from *Chitinophagaceae* obtained using universal primers were 200 times more abundant than *Ohtaekwangia* sequences.

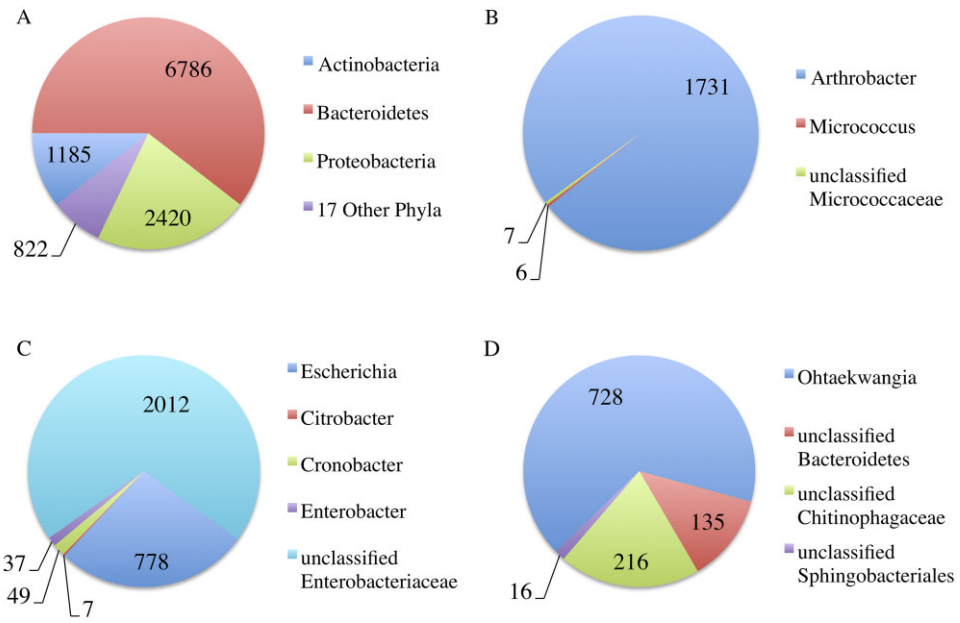


Fig. 5. Classification of sequences obtained with universal or genus specific primers. Relative number of sequences obtained from a water sample using different primers targeting the 16S rRNA gene. First, the sample was sequenced with universal primers (A), and four genera were selected that each represented a minor fraction of the three most common phyla present in the sample. Primers targeting *Arthrobacter* resulted in sequences classified predominantly as *Arthrobacter* (B). Primers targeting the genus *Escherichia* resulted in all *Enterobacteriaceae* sequences (C), but genus-level classification for most sequences was not possible due to the short amplicon size. *Ohtaekwangia* primers also amplified the much more abundant order *Sphingobacteriales* (D), as predicted by the primer design program. The fourth genus tested, *Emticicia*, resulted in all sequences belonging to the target genera, and is therefore not depicted.

Because the 16S region amplified by the *Ohtaekwangia* primers was encompassed by the universal primer sequences, we were able to determine the *Chitinophagaceae* sequences from the universal set that most closely matched the *Ohtaekwangia* primers. These proximal sequences belonged to the genus *Sediminibacterium*, which was predicted as a potential cross-amplification with 2.9% relative efficiency by the primer design programme, resulting from one terminal mismatch in the forward primer and one internal mismatch in the reverse primer. This illustrated the algorithm's ability to identify potential cross-amplifications and also supported the design choice of a cross-amplification cut-off of 0.1% relative efficiency.

Finally, we compared the primers designed for *Ohtaekwangia* using only hybridization efficiency (strategy #1) with those using both elongation efficiency and an induced mismatch at the 6th position from the 3′-end (strategy #3). In order to directly compare the three strategies without sequencing bias, we first amplified the environmental sample and then digested the resulting PCR products with a restriction enzyme that cuts at a target site present in *Ohtaekwangia* sequences but not in *Sediminibacterium* sequences (Fig. S1). Quantification of the resulting bands indicated that *Ohtaekwangia* comprised about 78% of the PCR products obtained from design strategy #2, which is consistent with the fraction found in sequencing. The PCR products obtained using strategy #1 were only 2% *Ohtaekwangia*, while the fraction obtained using strategy #3 was 93%. Taken together the results validated the programme's ability to design genus-specific primers, the use of elongation efficiency to improve specificity and the use of our framework for complex sequence discrimination tasks. Further improvement in specificity using these three strategies may be obtained by optimizing the PCR annealing temperature if a pure target DNA template is available.

Discussion

We have developed a semi-empirical framework for quantifying PCR bias by combining models of hybridization efficiency with experimental determination of factors influencing elongation efficiency. In doing so, we identified and quantified the context dependent extension of terminal mismatches, as well as the synergistic effect of multiple mismatches on elongation efficiency. This framework enabled a primer design approach exploiting *Taq*'s inherent biases to improve specificity in PCR. Additionally, the framework allowed us to resolve some of the inconsistencies among previous investigations of terminal mismatch effects (Huang *et al.*, 1992; Christopherson *et al.*, 1997;

Day *et al*., 1999; Ayyadevara *et al*., 2000; Wu *et al*., 2009; Stadhouders *et al*., 2010). For example, accounting for the penultimate nucleotide resolves several of the discrepancies between the results of Stadhouders and colleagues (2010) and Huang and collagues (1992). These two studies differed on the delay imposed by A/C, C/C, G/G, G/T and T/G terminal mismatches, but this can be explained by the penultimate base being a cytosine in the primers used by Stadhouders and collagues, which we found to result in smaller cycle delays than the adenine used in the Huang and colleagues primers for each of these mismatch types (Fig. 2).

The primary application of our framework is in the rational design of primers for use in targeting specific groups within collections of nearly identical templates, which is a common case in environmental microbiology studies when differentiating between taxonomic units defined with a phylogenetic marker gene (Ledeker and De Long, 2013). Non-target mismatched sequences can be amplified at a low rate in the initial cycles resulting in perfect match templates for subsequent cycles. Thus, the targeted sequences are enriched only in the early cycles, and the extent of enrichment increases exponentially with the cycle delay suffered by non-targeted templates. Unlike lowered hybridization efficiency, decreased elongation efficiency does not modify the shape (i.e. slope) of the amplification curve (Fig. S2), and therefore, it is not possible to tell using qPCR whether a 3′-end mismatch amplified based on the shape of the amplification curve alone. For these reasons, when designing target-specific primers using our framework, we sought to maximize predicted cycle delay by classifying any non-target sequences with a predicted delay of less than 10 cycles as potential cross-amplifications ($\geq 0.1\%$ relative to maximal efficiency). Note that this threshold corresponds to at least a 1000-fold ($2^{10} = 1024$) enrichment of target sequences over discriminated non-targets. This criterion is conservative by design because it not only gives a reasonable margin of safety for large-scale applications such as metagenomic profiling, but also provides a buffer against large standard deviations in cycle delay observed for some mismatches in the experimental calibration data incorporated in our model (Table S5).

We have demonstrated that PCR bias can be systematically exploited for targeting sequences of interest. To our knowledge, this is the first time measured extension parameters have been integrated into a primer design algorithm to further increase specificity over hybridization efficiency alone. The framework we developed has three key advantages over previously described primer design approaches. Firstly, by accounting for the roles of sequence context, elongation time and polymerase type in determining elongation efficiency of terminal mismatches, the framework provides a flexible means for making rational design choices that maximize sequence discrimination. Secondly, partitioning amplification efficiency into hybridization and elongation efficiencies enabled us to determine the positional effect of mismatches, insertions and deletions on elongation efficiency. Based on this information, we showed that inducing a mismatch to the target template at the 6th position was preferable to the 2nd position where it might affect elongation efficiency of the target amplicon.

Finally, we showed that our framework could be applied in an automated fashion to larger complex sequence discrimination scenarios consisting of a multitude of closely related DNA templates. We used our framework to show that extension biases can be exploited to improve the specificity of genus-specific primers targeting the 16S rRNA gene, a task that cannot be performed effectively in most cases with hybridization efficiency alone. The primer design programme based on our framework has been integrated into the DECIPHER package (Wright, 2012), which is a part of the BioConductor project for the R Statistical Programming Language (R Development Core Team, 2012). Given a set of DNA sequences, the programme can be used to design the most specific primers to any user-defined target group with minimal predicted amplification efficiency to non-target groups (Fig. S3). The genus-specific primers that we designed in this study and a web tool that can be used for standard primer design tasks are available online (http://DECIPHER.cee.wisc.edu).

Experimental procedures

Primers and template DNA

The 13 DNA templates used in this study to determine mismatch elongation efficiencies were taken from 16S clone libraries previously sequenced (Noguera *et al*., 2008) and deposited in the public GenBank repository with accession numbers shown in Table S6. These templates were chosen because their full-length 16S rRNA gene sequence was known, and each could be classified into a unique genus. Template DNA resided in a pCR 4-TOPO cloning vector (Invitrogen, Grand Island, NY, USA). Each clone's 16S rRNA gene product was isolated by first amplifying with M13 primers, purifying with a QIAquick spin column (Qiagen, Valencia, CA, USA), and then amplifying with the bacterial universal primers 27FY+M (Frank *et al*., 2008) and 1406R (Baker *et al*., 2003). This was followed by re-purification and dilution to an approximate concentration of 10^6 copies per microlitre as calculated from the known length of the amplicon and the amount of DNA determined by a NanoDrop 2000 spectrophotometer (Thermo Scientific, Madison, WI, USA).

The IDH2-WT and IDH2-R172K alleles were inserted into a pcDNA3.1 cloning vector (Invitrogen) and transformed into *Escherichia coli* using ampicillin resistance as a selection marker. Both vectors were extracted and purified using the

Wizard Plus SV system (Promega, Madison, WI, USA). The environmental sample used in this study was collected from the South Bay Aqueduct near San Jose, CA. Approximately 400 ml of water was filtered through a 22 μm nitrocellulose filter (Whatman International Ltd, Maidstone, England), and DNA was extracted using the FastDNA Spin Kit for Soil (MP Biomedicals, Solon, OH, USA). After amplification using a unique set of primers (Table S4), samples were diluted and re-amplified with one primer having a unique bar-coded tail sequence in order to prevent bias generated by the bar code sequence (Berry *et al*., 2011). The resulting PCR products were sequenced with a GS Junior (Roche, Basel, Switzerland) and deposited in GenBank under accession number SRP020345.

Forward and reverse perfect match primers (Table S6) were modified to create a set of 93 primers having a 3′-terminal mismatch to their respective DNA template (Table S1). All of these mismatched primers were predicted to have hybridization efficiencies greater than 99.9% at the annealing temperature of 50°C. We also designed a set of 78 primers with non-terminal mismatches, insertions or deletions within seven nucleotides of their 3′-end (Table S2). Because some of these non-terminal mismatches were predicted to lower hybridization efficiency more than terminal mismatches, we extended the length of some mismatched primers by one or two nucleotides (at the 3′-end) such that all mismatched primers maintained a predicted hybridization efficiency greater than 99.9% at the 50°C annealing temperature used in experiments. This primer extension technique was originally used by Ayyadevara and colleagues (2000) to incorporate additional mismatch types while maintaining the same primer site.

qPCR conditions

The reagents used in each qPCR reaction consisted of 0.8 μl of forward and reverse primers at a concentration of 10 μM, 2.4 μl of autoclaved DNA-free water purified using a Photronix RGW-5 (Vangard International, Neptune, NJ, USA), 2 μl of bovine serum albumin (Fisher Scientific, Fair Lawn, NJ, USA) at a concentration of 0.1 g l^{-1}, 4 μl of template DNA and 10 μl of iQ SYBR Green Supermix (Bio-Rad, Hercules, CA, USA). The iQ SYBR Green Supermix contained 100 mM KCl, 40 mM tris-HCl buffer, 1.6 mM dNTPs, 6 mM $MgCl_2$, 0.2 μM SYBR Green I and iTaq DNA polymerase. Experiments with Sso-*Taq* polymerase were performed using Bio-Rad's SsoAdvanced SYBR Green Supermix, which contains the same concentration of each reagent except with SsoAdvanced *Taq* in place of iTaq DNA polymerase. Controls containing water in place of template were run individually for each primer set. All primers were synthesized by the UW-Biotechnology Center.

A Roche LightCycler qPCR machine was utilized for mismatch experiments, and a Bio-Rad iCycler was used for temperature gradient experiments because the LightCycler did not have temperature gradient capability. It is worth noting that the method described herein is applicable for PCR or qPCR using all standard thermal cyclers. The qPCR thermal cycling temperature profile consisted of at least 50 cycles of: denaturation for 30 s at 95°C, annealing for 30 s at 50°C and elongation for 30 s at 72°C. Although 30 cycles were typically sufficient for complete amplification of targets, the thermal cycling was continued to determine the threshold cycle of non-targets that amplified up to 17 cycles later. Finally, a melt curve analysis was conducted starting from 55°C and using 0.1°C increments every 10 s. The melt peak was used to verify that each PCR product size was close to the expected length and that the amplified DNA was not the result of primer-dimer artefacts. Analysis of the amplification curves was performed with the programme qpcR (Ritz and Spiess, 2008). Cycle delays were compared using the Mann–Whitney rank-sum test for unpaired data, or Wilcoxon signed-rank test for paired data.

Analysis of sequences obtained using genus-specific primers

DNA was amplified using universal or genus-specific primers (Table S4). The resulting amplicons were sequenced using a GS Junior and processed with the R programming language (R Development Core Team, 2012) as follows: sequences matching the unique primer and barcode were identified, trimmed of primer sequences and low-quality base calls. Sequences with more than one ambiguity or fewer than 201 base pairs were filtered from the set of amplicons. Potential chimeras were identified with DECIPHER's Find Chimeras tool (Wright *et al*., 2012) and removed. The remaining sequences were aligned using SSU-ALIGN (Nawrocki, 2009). To mitigate the possibility of obtaining ambiguous classifications from short sequences (Liu *et al*., 2008), we classified the sequences into the group of their most closely related (almost full-length, ≥ 1200 nucleotides) sequence present in the RDP database (Cole *et al*., 2009) based on a neighbour-joining tree. When multiple reference sequences were equal distance from the query sequence, it was considered unclassified at the taxonomic level shared by the reference sequences.

PCR products obtained with *Ohtaekwangia* primers were digested for 16 h using the restriction enzyme HinfI (Thermo Scientific). This enzyme cuts at the restriction site GANTC, which is located near the centre of *Ohtaekwangia* amplicons but does not exist in *Sediminibacterium* sequences. Equal amounts of PCR product (10 μl) before and after digestion were run on a 1.5% agarose gel for 1 h at 8 volts cm^{-1} alongside a 100 base pair ladder (Promega). The gel was then stained for 1 h with ethidium bromide (6 μl/50 ml) and then destained in water for 5 min. The gel was imaged with a Bio-Rad Gel Doc XR (Fig. S1) and then analysed using ImageJ (Schneider *et al*., 2012) as follows: After rolling ball background subtraction, gel lanes were defined and average pixel intensity was plotted across the lane. The fraction of target was determined from the ratio of areas under the curves at the locations of the digested and undigested amplicons. For example, the ladder's 500 base pair band is designed to be threefold more intense than the band at 400 base pairs, and this ratio was found to be 2.83-fold using the areas under the intensity curve.

Acknowledgements

This work was partially supported by the Water Research Foundation (WaterRF project 4291). We thank Alex Bastian

for conducting many of the preliminary experiments that were not shown herein, John Denu and Wei Yu for providing the IDH2-WT and IDH2-R172K alleles, Jackie Strait for preparing the 16S rRNA gene templates, Colin Fitzgerald for assisting with sequencing, and the Center for High Throughput Computing for providing computing resources. The authors declare that they have no competing interests.

References

Ayyadevara, S., Thaden, J.J., and Reis, R.J.S. (2000) Discrimination of primer 3′-nucleotide mismatch by Taq DNA polymerase during polymerase chain reaction. *Anal Biochem* **284:** 11–18.

Baker, G.C., Smith, J.J., and Cowan, D.A. (2003) Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* **55:** 541–555.

Baker, M. (2012) Digital PCR hits its stride. *Nat Methods* **9:** 541–544.

Berry, D., Ben Mahfoudh, K., Wagner, M., and Loy, A. (2011) Barcoded Primers used in multiplex amplicon pyrosequencing bias amplification. *Appl Environ Microbiol* **77:** 7846–7849.

Cha, R.S., Zarbl, H., Keohavong, P., and Thilly, W.G. (1992) Mismatch amplification mutation assay (MAMA): application to the c-H-ras gene. *PCR Methods Appl* **2:** 14–20.

Christopherson, C., Sninsky, J., and Kwok, S. (1997) The effects of internal primer-template mismatches on RT-PCR: HIV-1 model studies. *Nucleic Acids Res* **25:** 654–658.

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., *et al*. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37:** D141–D145.

Day, J.P., Bergstrom, D., Hammer, R.P., and Barany, F. (1999) Nucleotide analogs facilitate base conversion with 3′ mismatch primers. *Nucleic Acids Res* **27:** 1810–1818.

Frank, J.A., Reich, C.I., Sharma, S., Weisbaum, J.S., Wilson, B.A., and Olsen, G.J. (2008) Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl Environ Microbiol* **74:** 2461–2470.

Fu, Q., Ruegger, P., Bent, E., Chrobak, M., and Borneman, J. (2008) PRISE (PRImer SElector): software for designing sequence- selective PCR primers. *J Microbiol Methods* **72:** 263–267.

Huang, M.M., Arnheim, N., and Goodman, M.F. (1992) Extension of Base mispairs by taq dna-polymerase – implications for single nucleotide discrimination in PCR. *Nucleic Acids Res* **20:** 4567–4573.

Johnson, S.J., and Beese, L.S. (2004) Structures of mismatch replication errors observed in a DNA polymerase. *Cell* **116:** 803–816.

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

Lan, Y.M., Wang, Q., Cole, J.R., and Rosen, G.L. (2012) Using the RDP classifier to predict taxonomic novelty and reduce the search space for finding novel organisms. *PLoS ONE* **7:** e32491.

Ledeker, B.M., and De Long, S.K. (2013) The effect of multiple primer-template mismatches on quantitative PCR accuracy and development of a multi-primer set assay for accurate quantification of pcrA gene sequence variants. *J Microbiol Methods* **94:** 224–231.

Li, B.H., Kadura, I., Fu, D.J., and Watson, D.E. (2004) Genotyping with TaqMAMA. *Genomics* **83:** 311–320.

Liu, Z.Z., DeSantis, T.Z., Andersen, G.L., and Knight, R. (2008) Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res* **36:** e120.

Mathews, D., Burkard, M., Freier, S., Wyatt, J., and Turner, D. (1999) Predicting oligonucleotide affinity to nucleic acid targets. *RNA* **5:** 1458–1469.

Mizrahi-Man, O., Davenport, E.R., and Gilad, Y. (2013) Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PLoS ONE* **8:** e53608.

Nawrocki, E.P. (2009) Structural RNA homology search and alignment using covariance models. PhD Thesis. Saint Louis, MO, USA: School of Medicine, Washington University.

Noguera, D.R., Yilmaz, L.S., Harrington, G.W., and Goel, R.K. (2008) *Identification of Heterotrophic Bacteria That Colonize Chloraminated Drinking Water Distribution Systems*. Denver, CO, USA: Awwa Research Foundation.

Pinto, A.J., and Raskin, L. (2012) PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS ONE* **7:** e43093.

R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*. http://www.R-project.org Vienna, Austria: R Foundation for Statistical Computing.

Ritz, C., and Spiess, A.N. (2008) qpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. *Bioinformatics* **24:** 1549–1551.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9:** 671–675.

Stadhouders, R., Pas, S.D., Anber, J., Voermans, J., Mes, T.H.M., and Schutten, M. (2010) The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5′ nuclease assay. *J Mol Diagn* **12:** 109–117.

Tindall, K.R., and Kunkel, T.A. (1988) Fidelity of DNA-synthesis by the Thermus-aquaticus DNA-polymerase. *Biochemistry* **27:** 6008–6013.

Wang, Y., Prosen, D.E., Mei, L., Sullivan, J.C., Finney, M., and Vander Horn, P.B. (2004) A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Res* **32:** 1197–1207.

White, A.K., VanInsberghe, M., Petriv, O.I., Hamidi, M., Sikorski, D., Marra, M.A., *et al*. (2011) High-throughput microfluidic single-cell RT-qPCR. *Proc Natl Acad Sci USA* **108:** 13999–14004.

von Wintzingerode, F., Gobel, U.B., and Stackebrandt, E. (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21:** 213–229.

Wright, E.S. (2012) *DECIPHER: Database Enabled Code for Ideal Probe Hybridization Employing R* [WWW document].

URL http://www.bioconductor.org/packages/2.12/bioc/html/DECIPHER.html.

Wright, E.S., Yilmaz, L.S., and Noguera, D.R. (2012) DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. *Appl Environ Microbiol* **78:** 717–725.

Wu, J.H., Hong, P.Y., and Liu, W.T. (2009) Quantitative effects of position and type of single mismatch on single base primer extension. *J Microbiol Methods* **77:** 267–275.

Yan, H., Parsons, D.W., Jin, G.L., McLendon, R., Rasheed, B.A., Yuan, W.S., *et al*. (2009) IDH1 and IDH2 mutations in gliomas. *N Engl J Med* **360:** 765–773.

Yuryev, A. (2007) *PCR Primer Design*. Totowa, NJ, USA: Humana Press.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Supplementary Methods. Detailed description of methods used for primer design.

Fig. S1. Comparison of three different design strategies for targeting *Ohtaekwangia* sequences. Gel runs of PCR products before and after digestion with the restriction enzyme HinfI, which cuts near the centre of *Ohtaekwangia* amplicons. Lane 1 contains a 100 base pair ladder. Lanes 2 and 3 contain PCR products obtained with *Ohtaekwangia* primers designed with hybridization efficiency alone (strategy #1) before and after digestion respectively. The PCR products in lanes 4 and 5 were obtained with primers designed using elongation efficiency (strategy #2), and lanes 6 and 7 with primers designed using elongation efficiency and an induced mismatch in the 6th position from the 3′-end. Intensity profiles from the top to bottom of lanes 3, 5 and 7 are shown in (B), (C) and (D) respectively. The digested (shorter) target amplicon (*Ohtaekwangia*) is coloured in green, whereas undigested non-target amplicons are coloured in red. Note that the reverse primer's target site in strategy #3 (lanes 6 and 7) is shifted towards the forward primer by 31 nucleotides relative to the reverse primer's target site used for strategy #2 (lanes 4 and 5). This difference in amplicon sizes (Table S2) explains the shorter digested and undigested product lengths in lanes 6 and 7 relative to lanes 4 and 5 respectively.

Fig. S2. Comparison of decreased hybridization efficiency with decreased elongation efficiency. Amplification curves for the *Mycobacterium* template. This figure illustrates (A) decreased hybridization efficiency and (B) decreased elongation efficiency.

A. Equal initial concentrations of template were amplified with an annealing gradient from 50°C to 75°C.

B. Equal initial concentrations of template were amplified using either perfect match forward (F) and reverse (R) primers (solid line), or one primer with a 3′-terminal mismatch (primer/template, dashed lines).

Fig. S3. Flowchart of describing how to design a primer step-by-step using the online Design Primers web tool.

Table S1. Terminal mismatched primers and observed elongation efficiencies.

Table S2. Efficiency of elongation of 3′ non-terminal mismatches.

Table S3. Results obtained with primers designed to discriminate alleles of the Human IDH2 gene using qPCR ($n = 3$).

Table S4. Primer sets used to validate the primer design methodology.

Table S5. Average efficiency of elongation of 3′ terminal mismatches separated by the penultimate primer nucleotide.

Table S6. DNA templates and perfect match primer sequences used to determine relative elongation efficiency of 3′ terminal mismatches.