Improving Specificity in qPCR with Primer Design Utilizing *Taq*'s Terminal Mismatch Bias

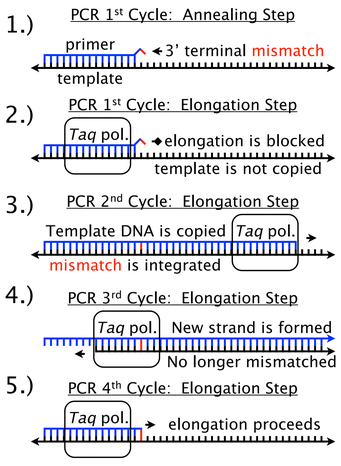
Erik S Wright , LS Yilmaz, S Ram, JM Gasser, GW Harrington, DR Noguera

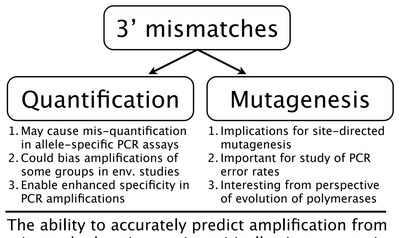
WISCONSIN UNIVERSITY OF WISCONSIN-MADISON

DEPARTMENT OF Civil and Environmental Engineering UNIVERSITY OF WISCONSIN-MADISON

3' Terminal Mismatches

Why do 3' mismatches matter in PCR?



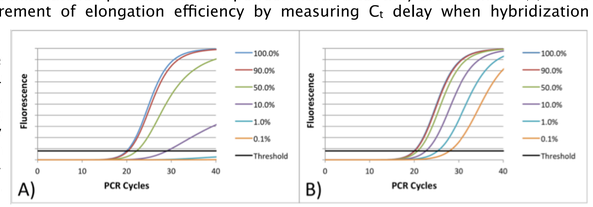


The ability to accurately predict amplification from mismatched primers is critically important in many applications of PCR. *Taq* DNA polymerase is sensitive to 3' mismatches because it lacks 3' to 5' exonuclease proofreading activity. Research has shown that some types of terminal mismatches are elongated at high efficiency, while other types greatly impede elongation. In this study, we determined that the amplification efficiency of terminal mismatches is also dependent on the penultimate base pairing. We utilized *Taq*'s bias to increase specificity in PCR, and applied our algorithm to design genus-specific primers targeting the 16S rDNA gene.

How was elongation efficiency measured?

Decreased hybridization efficiency results in a delayed amplification and a decrease in slope of the amplification curve (A). In contrast, decreased elongation efficiency only results in delayed amplification because mismatched primers become perfect match after they are extended (B). This allows direct measurement of elongation efficiency by measuring C_t delay when hybridization efficiency is 100%.

Hybridization can be maximized using a low annealing temp, and then measuring elongation efficiency of mismatches in comparison to non-mismatched primers.



	Cycle delay (ΔC_t)
GGCACTCCCATGGCGAC	11.8 ± 0.04 Rev.
GGCACGCCCATGGCGAC	9.4 ± 0.30 Rev.
CACCATTGGCAGGCACG	3.4 ± 0.06 Rev.
AGCCCATCACCATTGGCAG	2.2 ± 0.51 For.
5' ...AAGCCCATCACCATTGGCAAGCACGCCCATGGCGACC...	3' SNP allele
5' ...AAGCCCATCACCATTGGCAGGCACGCCCATGGCGACC...	3' wild type
AGCCCATCACCATTGGCAA	7.1 ± 0.02 For.
AGCCCATCACCATCGGCAA	12.5 ± 0.21 For.
CATTGGCAAGCACGCCC	2.8 ± 0.08 For.
AGCACGCCCATGGCGACC	8.5 ± 0.14 Rev.

Elongation efficiency of mismatches can be accurately measured by employing a primer mismatched to the target template. The perfect match and mismatched primers are amplified separately and the delay between them is converted into an elongation efficiency. Here, several mismatched primers are shown along with their corresponding cycle delays relative to the perfect match primer. In this case primers were design to distinguish between the wild type allele and a single nucleotide polymorphism (SNP). Using a variety of different design strategies, we were able to obtain 1000-fold specificity between the SNP and wild type alleles. This technique is clinically relevant for the inexpensive, accurate, and quick detection of DNA mutations.

Amplification in PCR obeys the exponential growth equation (Eq. 1). The observed efficiency can be measured using a standard dilution series experiment (Eq. 2). In contrast, elongation efficiency can be measured by the cycle delay imposed by a mismatched primer relative to perfect match (Eq. 3). Since there are two primers, the observed efficiency reflects the geometric mean of the two individual primer efficiencies (Eq. 4), as shown to the right.

qPCR Formulas

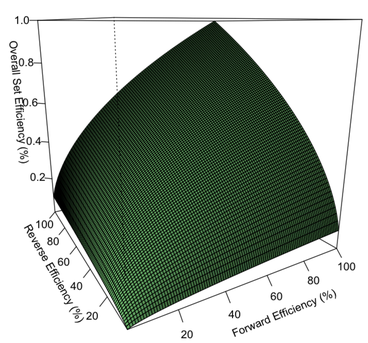
$$copies\ of\ template = initial\ copies * (1 + efficiency)^{n\ cycles} \quad (1)$$

$$Eff_{obs} = 10^{-1/slope} - 1 \quad (2)$$

$$Eff_{elong} = C_{t\_app}/C_t \quad (3)$$

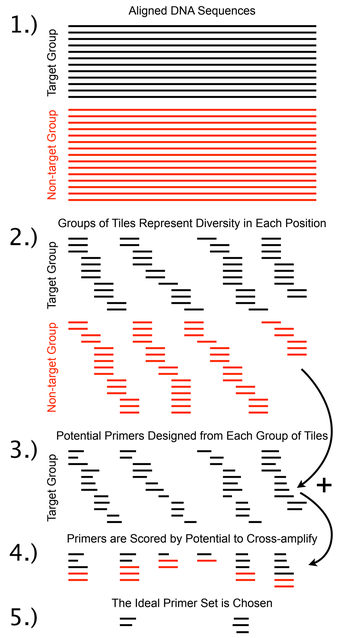
$$Eff_{obs} = \sqrt{Eff_{f\_primer} * Eff_{r\_primer}} \quad (4)$$

$$Eff_{primer} = Eff_{hyb} * Eff_{elong} \quad (5)$$



Primer Design Algorithm

How were these findings applied?



An algorithm was developed for choosing the optimal set of forward and reverse primers to amplify a target group while minimizing the potential to cross-amplify non-target groups.

First, a file of aligned sequences are classified into target and non-target groups. Second, a set of overlapping k-mers, termed "tiles," is formed to represent every possible target site in each group. Tiles are chosen to cover at least 90% of the sequences in the group.

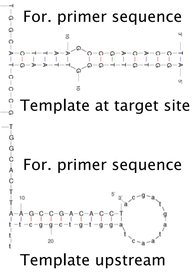
Third, a set of potential primers is designed from each set of tiles. The length of primers is optimized such that they meet a minimum hybridization efficiency at the annealing temperature. Primers are subject to several constraints: a maximum of four runs of a single base or di-nucleotide repeats, and a maximum number of permutations.

Forth, candidate primers are scored by their ability to result in a false positive hybridization. Here we use the mismatch parameters described previously to predict the amplification efficiency of non-target hybridizations. The efficiency of each mismatched primer is modeled as the multiplication of the elongation and hybridization efficiencies (Eq. 5).

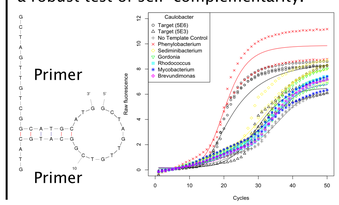
Fifth, the optimal set of forward and reverse primers is chosen to minimize false positive overlap. The overall amplification efficiency is the geometric mean of the forward and reverse efficiencies (Eq. 4.).

What challenges were uncovered?

A set of primers designed for *Dysgonomonas* was amplifying *Arcicella* with moderate efficiency. This was not predicted based on the sequence at the primer's target site. It turned out that the primer was binding over 200 bp upstream of the target site. In response, we added a comprehensive search to the algorithm to find alternative target sites located nearby.



Primers designed to amplify *Caulobacter* exhibited a strong case of primer-dimer formation. In response, we implemented a robust test of self-complementarity.



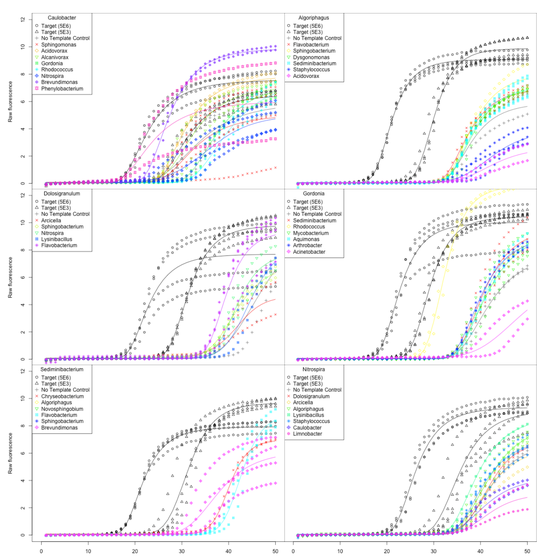
What was the accuracy of this approach?

A clone library was created from the 16S rDNA gene of organisms originating from a drinking water distribution system. The sequences of 40 DNA templates were classified according to genera.

Ten pure target DNA templates were amplified using their respective sets of genus-specific forward and reverse primers. Six of these are depicted here.

Non-target template DNA was amplified alongside the target DNA. In most cases targets amplified ahead of non-targets that were 1000-times more concentrated.

In cases where the non-target template amplified before the target template the algorithm was able to predict the potential for cross-amplification.



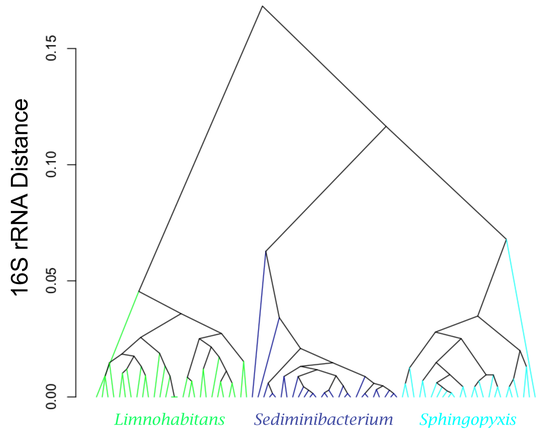
16S rRNA Applications

How was the design algorithm validated?

Using our algorithm we designed three cocktails of primers to specifically amplify the 16S rDNA of genera commonly found in fresh water ecosystems.

We extracted DNA from an environmental sample collected from the South Bay Aqueduct in Northern California that is sourced by the Sacramento-San Joaquin River Delta.

Target Genus	*Forward Primer(s) (5' to 3')*	*Reverse Primer(s) (5' to 3')*
Sediminibacterium	GAAGGATTAAGGTCCTCTGGATTGT	AGCGTCAATATATGCGTAGTAAGCTG AGCGTCAATATATGTGTAGCCTGC
Sphingopyxis	GACAGTACCGGGAGAATAAGCT GACAGTACCTGGAGAATAAGCTCC GACAGTACCTAGAGAATAAGCTCCG	CAGCACCTGTCACTGATCCA CAGCACCTGTCACTAATCCAGC CAGCACCTGTCACTTGTCCA
Curvibacter	CGTAGCGAAAGCTACGCTAATACCGC CGTAGCGAAAGTTACGCTAATACCG	CTTCGTTACTGAGTCAGTGAAGACC

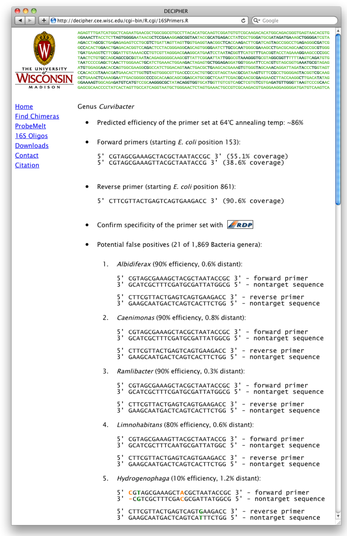


Extracted DNA was amplified separately using each primer cocktail. A clone library was formed from the resulting PCR products and sequenced using the Sanger method. Both *Sediminibacterium* and *Sphingopyxis* primers resulted in all cloned sequences belonging to their respective genus.

However, the sequences created using *Curvibacter* primers were all classified as *Limnohabitans*, a closely related organism that also perfectly matched the primer sequences. These results indicated that *Curvibacter* was present in much lower abundance than *Limnohabitans*, and lead to further investigation.

How could someone else use the results?

Find 16S primers online at: DECIPHER.cee.wisc.edu



- Here we created and validated an algorithm for designing primers targeting a specific group of sequences while minimizing affinity for non-target groups. Our algorithm employs a tiling approach and complex model of amplification that improves primer design for large groups of sequences separated into closely related groups.
- Designing primers for a conserved gene is difficult, and does not always yield the desired level of specificity (see *Curvibacter* above). However, in some cases the results can be valuable for practical applications (see *Staphylococcus* above). Our online tool makes the user aware of a primer set's shortcomings before using the primers experimentally.
- In cases where the desired specificity cannot be met using terminal mismatches alone, an induced mismatch may offer a means of further increasing primer specificity. The induced mismatch takes advantage of the non-linear nature of mismatched extension in that multiple mismatches are often extended much less efficiently than single mismatches.