Bioinformatics, 32(10), 2016, 1565–1567
doi: 10.1093/bioinformatics/btw047
Advance Access Publication Date: 23 January 2016
Applications Note

OXFORD

Sequence analysis

DesignSignatures: a tool for designing primers that yields amplicons with distinct signatures

Erik S. Wright and Kalin H. Vetsigian*

Department of Bacteriology, University of Wisconsin-Madison, Madison, WI 53715, USA

*To whom correspondence should be addressed.

Associate Editor: Alfonso Valencia

Received on 4 September 2015; revised on 12 December 2015; accepted on 20 January 2016

Abstract

Summary: For numerous experimental applications, PCR primers must be designed to efficiently amplify a set of homologous DNA sequences while giving rise to amplicons with maximally diverse signatures. We developed DesignSignatures to automate the process of designing primers for high-resolution melting (HRM), fragment length polymorphism (FLP) and sequencing experiments. The program also finds the best restriction enzyme to further diversify HRM or FLP signatures. This enables efficient comparison across many experimental designs in order to maximize signature diversity.

Availability and implementation: DesignSignatures is accessible as a web tool at www.DECIPHER.cee.wisc.edu, or as part of the DECIPHER open source software package for R available from BioConductor.

Contact: kalin@discovery.wisc.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Primer design is a fundamental step to any application of polymerase chain reaction (PCR), including high-resolution melting (HRM) analysis, fragment length polymorphism (FLP) analysis and amplicon sequencing. Although these techniques differ considerably, a common design objective is to construct primers that can efficiently amplify many different template variants while maximizing the difference between the resulting amplicon 'signatures'—melt curve, length or sequence depending on the experimental method. While many primer design tools are available (Noguera *et al.*, 2014), there currently does not exist a design tool that brings together different experimental techniques sharing the objective of maximizing signature diversity. Moreover, no automatic method is available to design primers for HRM analysis. Instead, semi-manual design approaches are typically used with the aid of software for melt curve prediction (Dwight *et al.*, 2011). Although this approach is reasonable for discriminating single nucleotide polymorphisms, it is unlikely to yield optimal results when the goal is to differentiate more complex variants. Furthermore, when variants are impossible to distinguish with FLP or HRM, digesting amplicons with a restriction enzyme can greatly diversify their signatures (Akey *et al.*, 2001). In such cases, an automated solution is highly desirable for exploring the space of possible primer and restriction enzyme pairings. By accurately predicting results *in silico*, design choices can be efficiently compared to maximize resolving power and minimize costs.

2 Methods

DesignSignatures designs primers in three or four steps:

1. Designing forward and reverse primers that will efficiently amplify as many input sequences (i.e. alleles) as possible.
2. Determining the set of PCR products for each combination of forward and reverse primers.
3. Scoring each candidate primer pair based on the diversity of its resulting amplicon signatures in sequencing, FLP or HRM.
4. Optionally, choosing the best restriction enzyme to further maximize signature diversity in FLP or HRM experiments.

The user provides a set of unaligned input DNA sequences (see the Supplementary Information (SI) text for usage information). The sequences (i.e. alleles) may be grouped if multiple templates will be present in the same PCR reaction, such as in the case of duplicate

genes. In design step 1, candidate 3′-ends are chosen that have the highest 8-mer frequency, since the 3′-subsequence must generally match all alleles for efficient amplification (Wright *et al.*, 2014). The most frequent 3′-ends are elongated into full-length primers that achieve reasonably high hybridization efficiency (>80% by default) at the user's specified experimental conditions (Noguera *et al.*, 2014). If desired, ambiguity is incorporated into the primer sequences to encompass more variants. The candidate primers matching the most input DNA sequences are used in subsequent steps.

In step 2, pairs of forward and reverse primers are used to 'amplify' the DNA *in silico*. In order to rapidly amplify many groups using a large number of primer pairs (≥500), we configured a fast method for predicting mismatched hybridization efficiency (Yilmaz *et al.*, 2012) to use DNA/DNA thermodynamic parameters. Regions of the DNA that are predicted to amplify with at least moderate efficiency (>50% by default) are include in the set of PCR products. In step 3, a signature is calculated for each set of amplicons, weighted by their amplification efficiency and/or length in accordance with the user's application. For sequencing, the signature is the 5-mer histogram of each group (i.e. allele), which is commonly used for classification. For FLP, the signature is the length(s) of amplicons belonging to each group. For HRM, the signature is the melt curve calculated using a linear-time algorithm (Tøstesen et al., 2003) with unified nearest neighbor parameters (SantaLucia, 1998).

Amplicon signatures are scored based on their average pairwise divergence (L^P-norm, where $p = 1$ by default) across all groups of input sequences. This approach rewards divergent signatures, and effectively penalizes similar signatures. The highest scoring primer pairs are then returned to the user, unless it is specified that a restriction enzyme will be used after PCR amplification. In step 4, restriction enzymes are used to digest the amplicons *in silico* into a set of shorter products. The set of all restriction enzymes available from New England BioLabs is provided for enzyme selection. The same scoring methodology is applied to the resulting DNA fragments according to the user's application. Finally, the top scoring combinations of primers and enzyme are returned to the user. The web tool also outputs visualizations of the predicted signatures and their pairwise distances.

3 Results

We first tested our algorithm for primer design by comparing its results to those of commonly used primers with the same objective (refer to the SI text for experimental methods). Although we would not expect the algorithm to output exactly the same primers, we would expect it to target the amplification of similar regions. As input, we used a set of 1601 ribosomal RNA (rRNA) operons extracted from 550 publically available bacterial genomes. Sequences were grouped by their genus of origin because many genomes contain more than one rRNA operon. We began by designing primers for amplicon sequencing with a product length between 350 and 500 base pairs (Fig. 1a). The top scoring primers overlapped with U515F and E939R, which are commonly used in 16S-based studies of bacterial diversity (Baker *et al.*, 2003). Next we designed primers for FLP analysis using the same input set. This yielded primers surrounding the variable length internal transcribed spacer (Fig. 1a), which is the same region used for automated ribosomal intergenic spacer analysis (ARISA), an FLP-based method (Jones *et al.*, 2007).

Since an equivalently widespread application does not yet exist for HRM analysis, we focused on designing primers targeting a

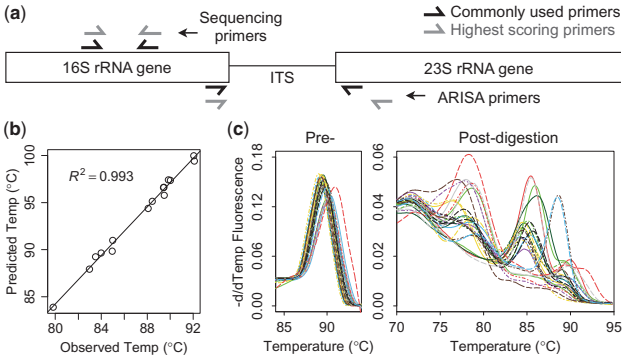

Fig. 1. (**a**) Primers designed by DesignSignatures were very similar to those commonly used for sequencing and ARISA targeting the rRNA operon. (**b**) Predictions of the melting peak for 15 distinct PCR products correlated strongly with those determined experimentally. (**c**) Primers designed to target the *rpoB* gene for HRM analysis resulted in substantially more diverse melt curve signatures after digestion with a restriction enzyme

variable region of the RNA Polymerase Subunit β (*rpoB*) gene, which is commonly used as a phylogenetic marker for bacteria belonging to the genus *Streptomyces*. We first verified our implementation of the algorithm for melt curve prediction using a set of 15 amplicons with diverse melt temperatures (T_m). Observed and predicted melt peaks were strongly correlated with an $R^2 = 0.993$, indicating that T_m prediction was very precise (Fig. 1b). However, T_m values required linear transformation to achieve high accuracy, as is typically observed due to the effects of salt concentration and intercalating dyes (Rasmussen *et al.*, 2007). Nevertheless, repeatable predictive offsets effectively cancel out since HRM analysis occurs on a relative basis.

To test the HRM algorithm, we designed primers for distinguishing the 26 *rpoB* sequences belonging to the genus *Streptomyces* available from GenBank. The results indicated that the top scoring primers would not generate sufficiently distinct melting signatures for the purposes of typing new strains. In contrast, digestion of the amplicons with a restriction enzyme (CviKI-1) was predicted to separate most strains. We verified these predictions experimentally by using the top scoring primers to amplify the DNA of 27 new *Streptomyces* isolates with different *rpoB* sequences, followed by digestion of the PCR products. As the outputs had indicated, the melt curve signatures after digestion were considerably more diverse than pre-digestion (Fig. 1c). This confirmed that DesignSignatures can assist in challenging experimental designs by suggesting primers, predicting their amplicons' signatures and indicating when it may be necessary to use a restriction enzyme or choose another target gene to achieve more diverse signatures.

Funding

We acknowledge support from the Simons Foundation Award 342039, the National Science Foundation Grant DEB 1457518 and the National Institute of Food and Agriculture, US Department of Agriculture, Hatch project 1006261.

Conflict of Interest: none declared.

References

Akey,J.M. *et al.* (2001) Melting curve analysis of SNPs (McSNP): a gel-free and inexpensive approach for SNP genotyping. *BioTechniques*, **30**, 358–362.

Baker,G.C. *et al*. (2003) Review and re-analysis of domain-specific 16S primers. *J. Microbiol. Methods*, **55**, 541–555.

Dwight,Z. *et al*. (2011) uMELT: prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. *Bioinformatics*, **27**, 1019–1020.

Jones,S.E. *et al*. (2007) Comparison of primer sets for use in automated ribosomal intergenic spacer analysis of aquatic bacterial communities: an ecological perspective. *Appl. Environ. Microbiol*., **73**, 659–662.

Noguera,D.R. *et al*. (2014) Mathematical tools to optimize the design of oligonucleotide probes and primers. *Appl Microbiol Biotechnol*., **98**, 9595–9608.

Rasmussen,J.P. *et al*. (2007) Use of DNA melting simulation software for in silico diagnostic assay design: targeting regions with complex melting curves and confirmation by real-time PCR using intercalating dyes. *BMC Bioinformatics*, **8**, 107.

SantaLucia,J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA*, **95**, 1460–1465.

Tøstesen,E. *et al*. (2003) Speed-up of DNA melting algorithm with complete nearest neighbor properties. *Biopolymers*, **70**, 364–376.

Wright,E.S. *et al*. (2014) Exploiting extension bias in polymerase chain reaction to improve primer specificity in ensembles of nearly identical DNA templates. *Environ. Microbiol*., **16**, 1354–1365.

Yilmaz,L.S. *et al*. (2012) Modeling formamide denaturation of probe-target hybrids for improved microarray probe design in microbial diagnostics. *PLoS One*, **7**, e43862.