Synthetic Biology, 2018, 3(1): ysx008

doi: 10.1093/synbio/ysx008
Research Article

A set of experimentally validated, mutually orthogonal primers for combinatorially specifying genetic components

Subu K. Subramanian[1,*], William P. Russ[1], and Rama Ranganathan[1,2,3,*]

[1]Green Center for Systems Biology, UT Southwestern Medical Center, Dallas, TX, USA, [2]Department of Pharmacology, UT Southwestern Medical Center, Dallas, TX, USA and [3]Department of Biophysics, UT Southwestern Medical Center, Dallas, TX, USA

*Corresponding authors: E-mail: subramaniansk@gmail.com; E-mail: rama.ranganathan@utsouthwestern.edu

Abstract

The design and synthesis of novel genes and deoxyribonucleic acid (DNA) sequences is a central technique in synthetic biology. Current methods of high throughput gene synthesis use pooled oligonucleotides obtained from custom-designed DNA microarray chips, and rely on orthogonal (non-interacting) polymerase chain reaction primers to specifically de-multiplex, by amplification, the precise subset of oligonucleotides necessary to assemble a full length gene. The availability of a large validated set of mutually orthogonal primers is therefore a crucial reagent for high-throughput gene synthesis. Here, we present a set of 166 20-nucleotide primers that are experimentally verified to be non-interacting, capable of specifying 13 695 unique genes. These primers represent a valuable resource to the synthetic biology community for specifying genetic components that can be assembled through a scalable and modular architecture.

Key words: orthogonal primers; DNA assembly; modular genetic engineering; genetic circuit

1. Introduction

Advances in high-throughput gene synthesis (1) and multiplex automated genome engineering (MAGE) (2, 3) technologies have made it possible to assemble gene libraries and genomes using custom oligonucleotide pools as starting reagents. A typical gene assembly workflow begins with gene-specific primer pairs that are used to selectively amplify the precise subset of oligonucleotides constituting a single gene, and ends with their assembly to build the full-length gene (1) (Figure 1A). This workflow is implemented in parallel for the simultaneous assembly of thousands of genes, with each gene specified by a unique combination of gene-specific primers. The set of gene-specific primers is thus crucial for deoxyribonucleic acid (DNA) assembly.

The gene-specific primers have the following design constraints: (i) their total number should be small in order to minimize synthesis costs, while still permitting specific amplification of thousands of genes, (ii) they should be 'well-behaved' polymerase chain reaction (PCR) primers (that is, should not form dimers and hairpins) and (iii) primer pairs must be mutually orthogonal or non-interacting, such that a given primer pair does not amplify fragments specified by another pair, which could potentially interfere with the subsequent assembly steps. Here, we report a library of 166 experimentally validated primers that satisfy these criteria, with the capacity to uniquely specify 13 695 genes.

2. Materials and methods

2.1 Algorithm parameters to design orthogonal primers

Filter 1 retains sequences ending in G/C, with composition of A <45%, GC between 40 and 60%, C between 20 and 30%,

Submitted: 23 June 2017; **Received (in revised form):** 17 November 2017; **Accepted:** 17 November 2017

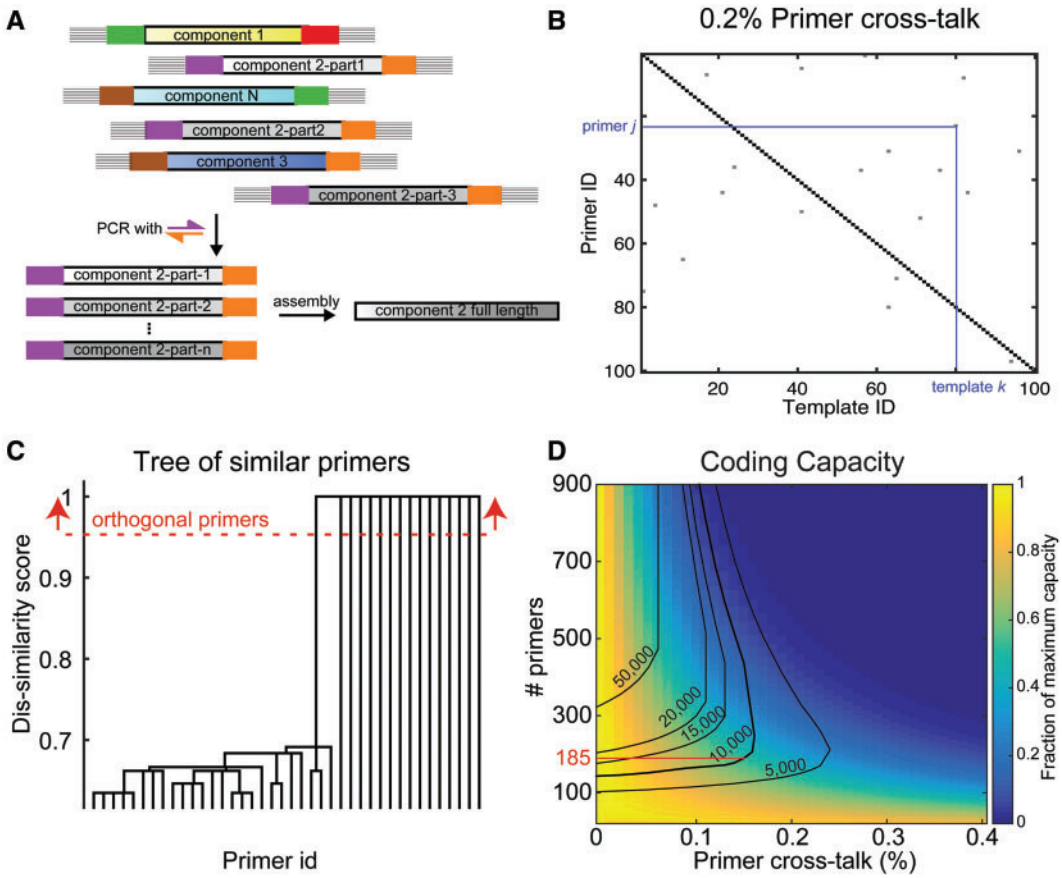

Figure 1. The need for a verified orthogonal primer set, illustrated for gene construction (**A**): Parts required to construct each of the N genes are specified by a unique pair of primers from the set of mutually orthogonal primers. All components of gene 2 are specifically amplified by the purple-orange primer pair, followed by (for instance) removal of the priming region using a type IIS restriction enzyme, and subsequent PCR overlap assembly to generate the full-length gene. (**B**) Schematic of the primer interaction matrix. The interaction between primer *j* and priming sequence for primer *k* is represented in pixel (*j*, *k*). The schematic contains 0.2% of the off-diagonal space for 100 primers marked randomly as interactions. (**C**) Representative interactions in B depicted in tree form, with branches connecting primers with similar interaction profiles. Orthogonal primers score above a dissimilarity threshold of 0.95. (**D**) The fraction of genetic components specifiable by the set of primers, as a function of inter-primer cross-talk. Contours connect points with equal coding capacity (denoted above the contour). With 185 primers, the coding capacity is ~17 000 components with no cross-talk. Capacity drops to ~10 000 with 0.16% primer cross-talk.

annealing temperature for primers in PCR (T_m) $\geq 58\,^{\circ}$C, and do not form hairpins/dimers involving 5 or more nucleotides (4). Sequences containing restriction enzyme recognition motifs (BamH1, EcoR1, HindIII, and BsrD1) were removed. Other parameters for Filter 1 and Filter 2 (network elimination) were kept unchanged from the original algorithm (4).

2.2 Validating primer interactions

The 185 PCRs, one reaction per gene specific primer, were performed with 30 rounds of amplification, 58 °C annealing temperature and 10 s extension time using Q5® Hot Start High-Fidelity DNA polymerase system. Buffer and primer concentration (0.5 uM each) were as recommended in the manual. All reactions had the same template (the oligonucleotide pools at 7 pg/ul) and reverse primer (CTTCTCCTTTACTAGTGAATTC). The amplicons were then independently taken through a final PCR to added Illumina Truseq adaptors (Figure 2B, binding at gray boxes) and sequenced on two sequencing runs with the 300 cycle MiSeq Reagent kit.

2.3 Calculating the cross interaction matrix

The number of occurrences of each of 185 12-mer barcodes in each of the amplicons was calculated using custom shell scripts and normalized such that the relative counts of *j*th barcode for the *j*th amplicon (self-interaction) is 1000. The earlier network elimination algorithm was used to identify the orthogonal primer set.

3. Results

A relatively small number of primers can specify a large number of genes if used combinatorially. For example, n primers can encode $\binom{n}{2}$ pairwise combinations (Number of ways to choose 2 components from a total of n components. Mathematically, it is $0.5{*}n{*}(n-1)$), rather than only $n/2$ pairs if each primer is used only once. The risk of the combinatorial approach, however, is that the effects of cross-talk between primers are exaggerated since each primer is used with every other in combination. Unintended cross-talk would reduce the coding capacity—the number of genes that can be uniquely specified by a primer set. While existing primer design algorithms attempt to minimize cross-talk, they have not been experimentally validated for orthogonality in PCR. We performed simulations to illustrate the impact of primer cross-talk on the coding capacity. The simulation randomly assigns cross-talk between primers, graphically represented in a symmetric primer interaction matrix

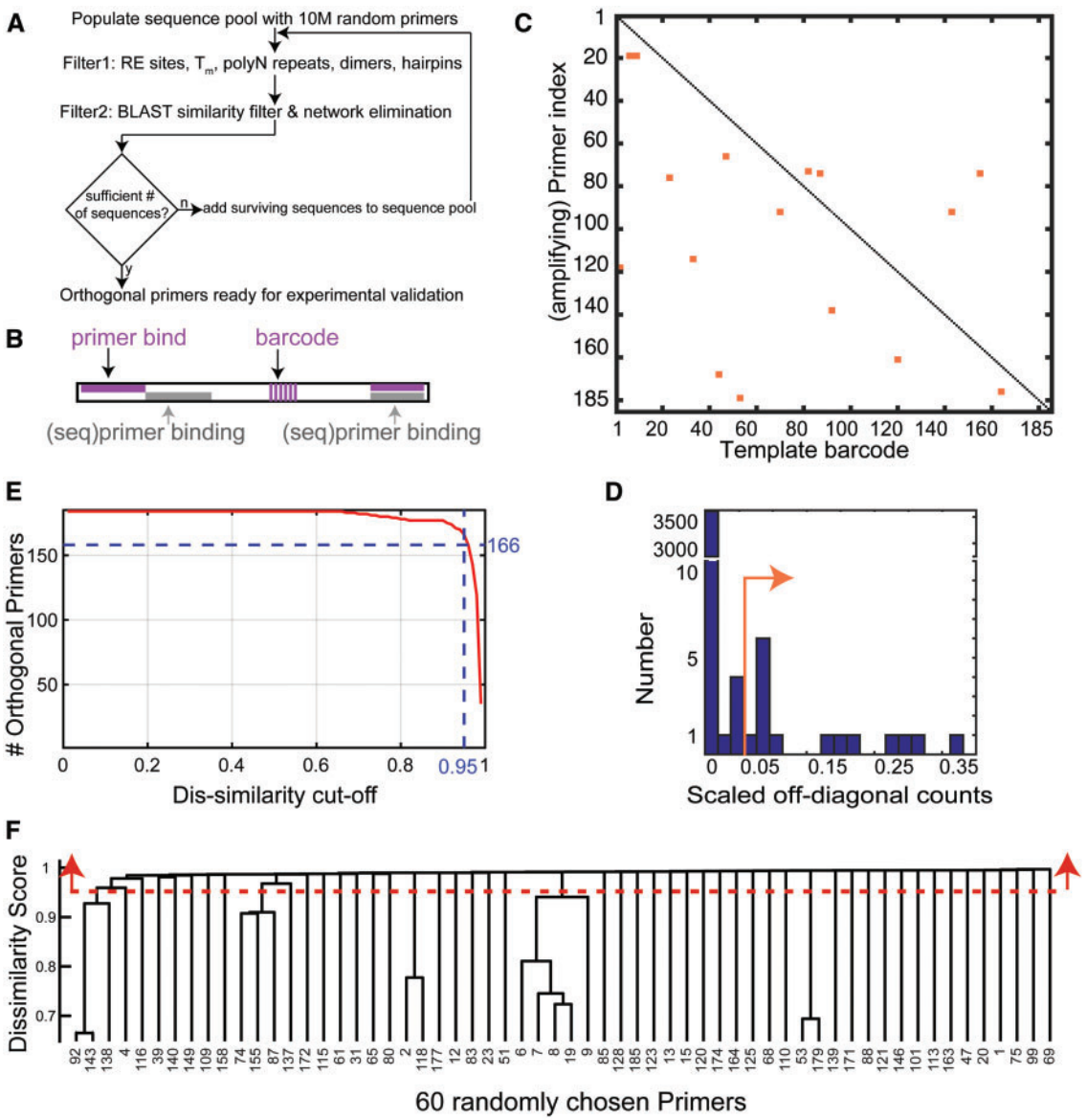

Figure 2. Design and validation of orthogonal primers. (**A**) Flowchart to computationally design non-interacting primers, (**B**) Template oligonucleotide schematic. A unique barcode is associated with each orthogonal primer. The barcode is amplified by PCR with the orthogonal primer and a universal reverse primer (purple boxes). In a subsequent PCR, sequencing primers (which anneal to gray boxes) prepare the barcodes for sequencing. (**C**) Experimentally determined interaction matrix. Self-priming of primers (principal diagonal) is set to 1. Significant (>0.05) off-diagonal elements denoting interactions are highlighted as orange pixels. (**D**) Distribution of cross-talk, off-diagonal elements in D. (**E**) Number of primers in the orthogonal set, as a function of dis-similarity cut-off. We have chosen a cut-off value = 0.95 for discussion (**F**) Representation of the raw data in C, used to identify the set of orthogonal primers.

(Figure 1B). Element (j, k) in the interaction matrix denotes the extent of amplification of primer binding site k when amplified by primer j. Row j thus denotes the amplicon profile of primer j—that is, the relative frequencies of amplifying all primer binding sites for primer j. Orthogonal primer profiles can be identified by first constructing a similarity tree (Figure 1C) in which primers with similar amplicon profiles are clustered together, and using a high dissimilarity score (0.95) as a threshold.

Note that the interaction matrix can be asymmetric. Although the binding profiles of primers are expected to be symmetric (if primer j can anneal to primer k, primer k anneals to primer j with identical energetics), primer extension by DNA polymerase occurs only if the 3′ end is double stranded. Therefore, the 3′ end of primer j annealing to the middle of primer k will show a $j \rightarrow k$ interaction but not a $k \rightarrow j$ interaction.

The simulations reveal two main findings. First, coding capacity depends sensitively on the percentage of cross-talk in the primer set; for example, in a 100 primer interaction matrix, just ten significant off-diagonal pixels—0.2% cross talk—excludes five primers from the orthogonal set, decreasing coding capacity by 10%. Second, for a given number of specifiable genes (contours, Figure 1D), the allowable cross-talk reaches a maximum as a function of the number of primers used. Based on these observations, we decided to design and validate a set of 185 primers. Even with maximum allowable cross-talk, this primer set should have a coding capacity of at least 10 000 genes, a value sufficient for the current capacity of high-throughput gene construction projects (Custom Array - Oligo Pools, http://www.customarrayinc.com/oligos_main.htm).

3.1 Primer design

Our algorithm for primer design is based on the DeLOB algorithm (4) that has been used to design orthogonal oligonucleotides for microarray hybridization. We modified the algorithm to generate 20-nucleotide primers, each with an annealing temperature of 58 °C (Figure 2A). The algorithm starts by generating 10 million random 20-mers followed by two sequential filters to remove suboptimal primers. The first filter selects sequences with 'good primer characteristics' (see Section 2) including favorable annealing temperatures and low propensities to form dimers and hairpins. The second filter—network elimination—uses BLAST (5) based pairwise sequence similarity scores to exclude similar sequences. We experimentally tested the top 185 dissimilar sequences (Supplementary Table S1) for orthogonality.

3.2 Experimental validation

We designed and purchased a pool of template oligonucleotides (Custom Array - Oligo Pools, http://www.customarrayinc.com/oligos_main.htm), each containing a primer binding site for one of the 185 gene-specific primers, an associated unique 12 nucleotide barcode, a common reverse priming site, and flanking adaptors for Illumina sequencing. Using the oligonucleotide pool as template, we performed individual PCR reactions with each of the 185 gene-specific primers and sequenced the amplicon (~28 000 reads each) by high-throughput sequencing to identify which of the 185 unique barcodes are amplified by each gene-specific primer. A binarized matrix of normalized interaction profile for each primer is shown in Figure 2C (for unnormalized counts, Supplementary Table S2) and the distribution of cross-talk pixels in Figure 2D. The interaction space (off-diagonal elements of the interaction matrix) is sparse, with only a few primers amplifying templates corresponding to other primers. To identify orthogonal primers, we calculated a dissimilarity score and generated the corresponding interaction tree (Figure 2F, methods in Supplementary Information). A dissimilarity threshold is used to identify orthogonal primers—primers above the threshold are directly assigned to the set of orthogonal primers. Among the primers that were observed to interact (below the dissimilarity threshold), we randomly chose one from each interacting clique to append to the orthogonal primer set. The total number of orthogonal primers depends on the dissimilarity threshold (Figure 2E). Using a threshold of 0.95 results in a set of 166 orthogonal primers (Supplementary Table S1), modifying the dissimilarity threshold (MATLAB scripts in Supplementary Information) will change the number of orthogonal primers (Figure 2E).

4. Discussion

Here, we report a framework to design and validate a relatively small orthogonal PCR primer library to specify a large number of genetic components. Our goal was to design a primer set to encode a component library size of at least 10 000. We designed and tested 185 primers (theoretical capacity of 17 020 components) for cross-talk and observed cross-talk in 0.11% of possible combinations. From this, we identified 166 mutually orthogonal primers, with a coding capacity of 13 695 components (Supplementary Tables S1 and S2). This is the first report of a validated primer set with a coding capacity >100 genes, a resource that should be broadly useful in high-throughput gene synthesis and multiplexed screening of DNA libraries (6).

Synthetic biology is the design and construction of biological systems to perform novel and useful functions. The synthetic biology community has contributed a large library of genetic parts (7), including standardized libraries like BioBricks (8) and YeastFab (9), encompassing a range of applications including biosensors (10) and programmable genetic circuits (11, 12). The next step in engineering biological systems is to use existing parts to compose complex and novel functions. The orthogonal primers reported here will enable experiments to specify gene or DNA fragments for high-throughput gene construction and to uniquely address genetic components (like biosensors and genetic circuits) in a genetic 'breadboard' background (13). The use of these primers can also be extended to orthogonal and modular CRISPR mediated gene regulation (14).

Supplementary data

Supplementary Data are available at SYNBIO Online.

Acknowledgments

We thank members of the Ranganathan Laboratory for discussions and critical reading of the manuscript.

Funding

Robert A. Welch Foundation [I-1366], the Lyda Hill Endowment for Systems Biology, the Green Center for Systems Biology and the National Institutes of Health through the NIH Director's Transformative Research Award [RO1-GM123456 to R.R.].

Conflict of interest statement. None declared.

References

1. Kosuri,S., Eroshenko,N., Leproust,E.M., Super,M., Way,J., Li,J.B. and Church,G.M. (2010) Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat. Biotechnol.*, 28, 1295–1299.
2. Wang,H.H., Isaacs,F.J., Carr,P.A., Sun,Z.Z., Xu,G., Forest,C.R. and Church,G.M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, 460, 894–898.
3. —— and Church,G.M. (2011) Multiplexed genome engineering and genotyping methods: applications for synthetic biology and metabolic engineering. *Methods Enzymol.*, 498, 409–426.
4. Xu,Q., Schlabach,M.R., Hannon,G.J. and Elledge,S.J. (2009) Design of 240, 000 orthogonal 25mer DNA barcode probes. *Proc. Natl. Acad. Sci. U S A*, 106, 2289–2294.
5. Altschul,S.F., Gish,W., Miller,W.T., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, 215, 403–410.
6. Stiffler,M.A., Subramanian,S.K., Salinas,V.H. and Ranganathan,R. (2016) A protocol for functional assessment of whole-protein saturation mutagenesis libraries utilizing high-throughput sequencing. *J. Vis. Exp.* doi:10.3791/54119.
7. Voigt,C.A. (2006) Genetic parts to program bacteria. *Curr. Opin. Biotechnol.*, 17, 548–557.
8. Smolke,C.D. (2009) Building outside of the box: iGEM and the BioBricks Foundation. *Nat. Biotechnol.*, 27, 1099–1102.
9. Guo,Y., Dong,J., Zhou,T., Auxillos,J., Li,T., Zhang,W., Wang,L., Shen,Y., Luo,Y., Zheng,Y. et al. (2015) YeastFab: the design

and construction of standard biological parts for metabolic engineering in Saccharomyces cerevisiae. *Nucleic Acids Res.*, 43, e88.
10. Prindle,A., Samayoa,P., Razinkov,I., Danino,T., Tsimring,L.S. and Hasty,J. (2012) A sensing array of radically coupled genetic "biopixels". *Nature*, 481, 39–44.
11. Padirac,A., Fujii,T. and Rondelez,Y. (2012) PNAS Plus: bottom-up construction of in vitro switchable memories. *Proc. Natl. Acad. Sci. U S A*, 109, 4–6.
12. Gupta,S., Bram,E.E. and Weiss,R. (2013) Genetically programmable pathogen sense and destroy. *ACS Synth. Biol.*, 2, 715–723.
13. Wei,X., Syed,A., Mao,P., Han,J. and Song,Y-A. (2016) Creating sub-50 nm nanofluidic junctions in PDMS microfluidic chip via self-assembly process of colloidal particles. *J. Vis. Exp.* doi: 10.3791/54145.
14. Didovyk,A., Borek,B., Hasty,J. and Tsimring,L. (2016) Orthogonal modular gene repression in escherichia coli using engineered CRISPR/Cas9. *ACS Synth. Biol.*, 5, 81–88.