

UCSF

UC San Francisco Previously Published Works

Title

primerjinn: a tool for rationally designing multiplex PCR primer sets for amplicon sequencing and performing in silico PCR.

Permalink

<https://escholarship.org/uc/item/9s17w1w6>

Journal

BMC Bioinformatics, 24(1)

Authors

Metcalfe, John

Limberis, Jason

Publication Date

2023-12-12

DOI

10.1186/s12859-023-05609-1

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

SOFTWARE

Open Access



primerJinn: a tool for rationally designing multiplex PCR primer sets for amplicon sequencing and performing in silico PCR

Jason D. Limberis^{1*} and John Z. Metcalfe¹

*Correspondence:
Jason.Limberis@ucsf.edu

¹ Division of Pulmonary and Critical Care Medicine, Trauma Centre, Zuckerberg San Francisco General Hospital, University of California, San Francisco, San Francisco, CA, USA

Abstract

Background: Multiplex PCR amplifies numerous targets in a single tube reaction and is essential in molecular biology and clinical diagnostics. One of its most important applications is in the targeted sequencing of pathogens. Despite this importance, few tools are available for designing multiplex primers.

Results: We developed primerJinn, a tool that designs a set of multiplex primers and allows for the in silico PCR evaluation of primer sets against numerous input genomes. We used primerJinn to create a multiplex PCR for the sequencing of drug resistance-conferring gene regions from *Mycobacterium tuberculosis*, which were then successfully sequenced.

Conclusions: primerJinn provides a user-friendly, efficient, and accurate method for designing multiplex PCR primers for targeted sequencing and performing in silico PCR. It can be used for various applications in molecular biology and bioinformatics research, including the design of assays for amplifying and sequencing drug-resistance-conferring regions in important pathogens.

Keywords: PCR, Primer design, Targeted sequencing, Multiplex PCR

Introduction

Multiplex PCR amplifies numerous targets in a single tube reaction and is essential in molecular biology and clinical diagnostics. One of its most important applications is in the targeted sequencing of pathogens. By using multiple primer pairs to amplify specific target regions in a single reaction, multiplex PCR allows for the simultaneous detection and identification of multiple pathogens or drug resistance-conferring regions in a single sample, making it a valuable tool for diagnostic and epidemiological studies. Despite this importance, few tools are available for designing multiplex PCR primers suitable for targeted sequencing, [1–7] some of which are no longer accessible, and none of which also perform in silico PCR. This is because designing primers for multiple targets simultaneously is challenging, requiring careful consideration of multiple factors such as primer



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

specificity, to reduce off target amplification; amplicon length, to ensure relatively even proportions of amplicons; and primer interactions, to limit primer dimer formation, under the specific reaction conditions for the, typically high-fidelity, polymerase needed for accurate sequencing, which require higher annealing temperatures. Yet, these parameters are critical for designing efficient primer sets for use on clinical samples, which may have low amounts of DNA and contain numerous different microbes.

We developed primerJinn, a tool that designs a set of multiplex PCR primers and allows for the *in silico* PCR to evaluate them against numerous input genomes. primerJinn uses primer3 [8] to create primers and a clustering method to select the best primer set based on the amplicon size, melting temperature, and primer interactions. The *in silico* PCR function uses BLAST [9] to identify primer pairs that amplify a DNA sequence of a user-specified maximum, at a given annealing temperature and salt concentration and provides detailed information about the primers and amplicons. Our tool also incorporates approximations for melting temperatures utilizing Q5 Hot Start High-Fidelity Polymerase buffers (NEB, USA), which differ significantly from most other polymerases. primerJinn provides an efficient and accurate method for designing multiplex PCR primers and performing *in silico* PCR and can be a valuable resource for researchers in the field of targeted sequencing for pathogens.

Implementation

Primer design

primerJinn is written in Python and uses primer3 to design primers surrounding each specific target range in an input design FASTA file (Fig. 1). By default, these primers amplify a region of 400–800 nucleotides, have an optimal length of 20 (range 10, 40) nucleotides, an optimal T_m of 65 °C (range 62 °C, 68 °C). These are optimal for amplification with a high-fidelity polymerase and subsequence sequencing using Illumina. Primers are checked for specificity against a mispriming library consisting of an optional input background FASTA file, and the portions of the input design FASTA file that are not used to design a particular primer pair. If no primer pair can be designed for a target, the input parameters are iteratively relaxed until a pair is found. Since high-fidelity polymerase buffers tend to increase the T_m of primers, we have included an approximation for the highest-fidelity polymerase, Q5 from NEB. We also use the NEB T_m calculator API to output the final T_m for the selected primer set. Following the design of one hundred (default value) primer pairs for each region, a matrix is constructed including T_m and amplicon size for each primer pair, and heterodimer formation probability (based on the Gibbs free energy) for each primer combination. This matrix is used to generate clusters. The cluster with the most regions covered is selected, and missing primers are added from the next closest cluster using an Euclidean metric and Ward linkage criterion. The output is written to an Excel file, and includes Illumina adapter tails if selected, allowing for the nested PCR barcoding of samples.

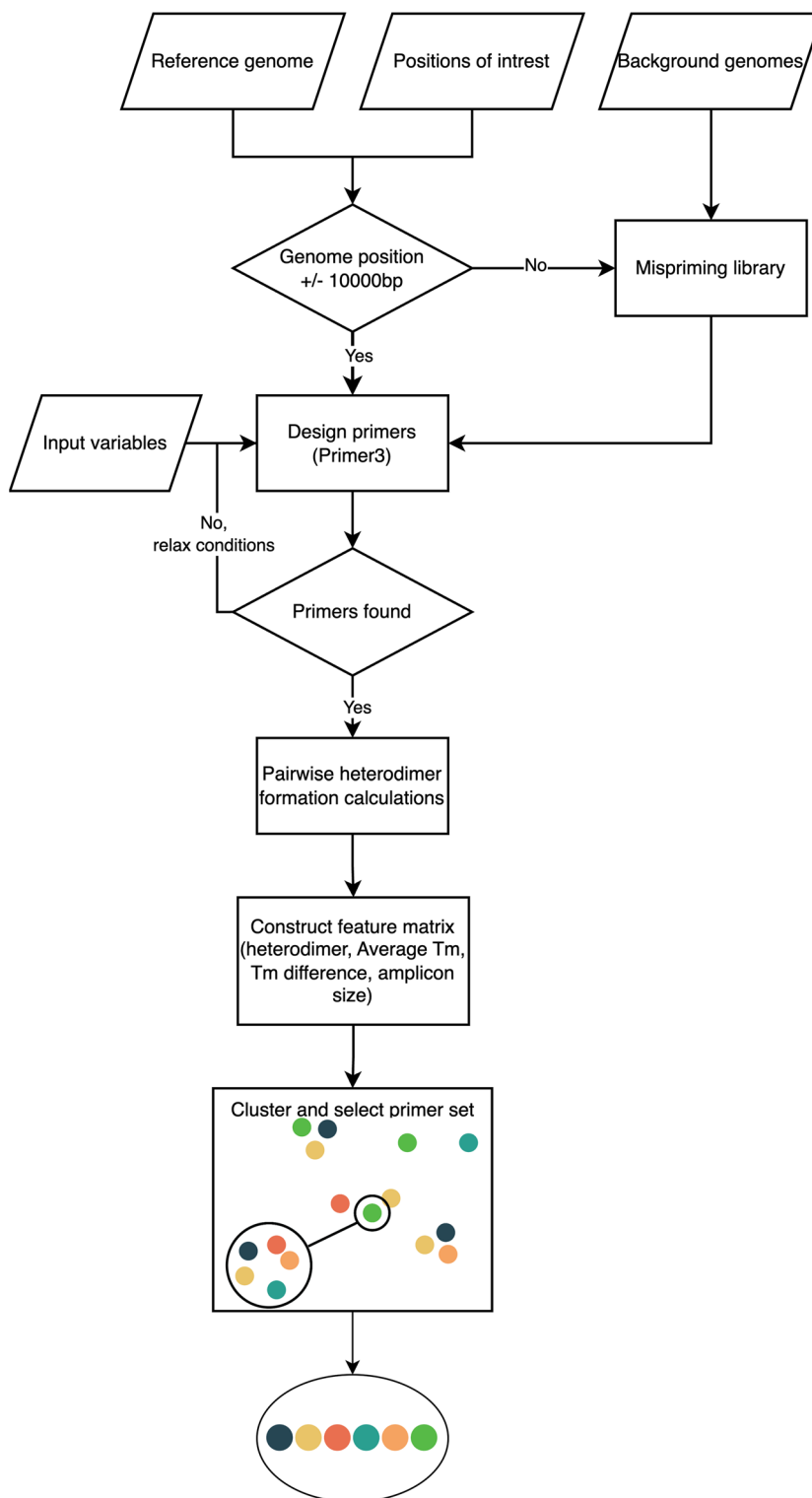


Fig. 1 primerJinn primer design workflow

In silico PCR

primerJinn also allows for the in silico PCR evaluation of primers. It takes a reference FASTA file and primer sequences as input and returns the binding position (located using blastn-short) and product length of any pair of primers that generate a product at, or below the input T_m (default is 70 °C) and the maximum amplicon size (default is 2000 nucleotides). Options include annealing temperature, salt concentration, maximum product length, and whether two or more bases at 5' end of the primer are required to bind. The output is written to an Excel file.

Results

To evaluate primerJinn we selected eight drug resistance-conferring gene regions from *Mycobacterium tuberculosis*, the etiological agent of tuberculosis. We passed the 4.4 Mb, high GC (~65%) genome FASTA (NC_000962.3) to primerJinn with the regions listed in Table 1. primerJinn output one primer set for each position (Table 2), with the mean primer T_m and amplicon size being 65 °C (range 64 °C, 67 °C) and 665 nucleotides (range 454, 791), respectively. We then used these primers as input for primerJinn in silico PCR function, which appropriately returned only the eight expected amplicons. Finally, we synthesized the 16 primers and performed singleplex and multiplex PCR on *M. tuberculosis* H37Rv genomic DNA using NEB Q5 HotStart DNA polymerase MasterMix for 35 cycles with denaturation, annealing, and extension for 10s at 98 °C, 20s at 65 °C and 30s at 72 °C. We ran electrophoresis gels and saw the expected bands (Fig. 2A), which showed no mispriming to the human genome (lane 10). We then sequenced the amplicon pool using 2 × 250 bp paired-end Illumina sequencing and aligned the reads to the target amplicons using BWA-men [10]. We had an 82% alignment to the target amplicons (Fig. 2B) at similar depths for each amplicon, with an unaligned remainder consisting of low-quality reads, singletons, and primer dimers. We also evaluated our Q5 T_m approximation settings against 10,000 random DNA sequences from 15 to 25 nucleotides long (1000 in each group). Our T_m approximations were a mean of –0.21 °C (SD 0.71 °C) below that of the NEB T_m.

Table 1 Genomic targets for primerJinn multiplex design of drug resistance-conferring regions on *M. tuberculosis* H37Rv (NC_000962.3)

Gene	Drug	Genomic position start	Genomic position end
<i>gyrB</i>	Fluoroquinolones	6578	7250
<i>fgd1</i>	Delamanid	490,900	491,416
<i>rpoB</i>	Rifampicin	761,007	761,277
<i>rv0678</i>	Bedaquiline	778,990	779,488
<i>fbtC</i>	Linezolid	1,303,831	1,303,911
<i>atpE</i>	Bedaquiline	1,461,045	1,461,291
<i>inhA</i>	Isoniazid	1,674,182	1,674,222
<i>katG</i>	Isoniazid	2,154,831	2,154,873
<i>pncA</i>	Pyrazinamide	2,288,681	2,289,242

Table 2 primerJinn output for eight drug resistance conferring gene regions of *M. tuberculosis* H37Rv (NC_000962.3)

Primer pair	Forward primer (5' ->3')	Reverse primer (5' ->3')	Forward Tm (°C)	Reverse Tm (°C)	Product Size (bp)	Name	Forward Primer Tm NEB (°C)	Reverse Primer Tm NEB (°C)
1	GAGGAACACCACTAG TACCG	CTCGATGACTTTACG GCCAT	65	64	675	Target 1,461,045– 1,461,291	64	64
2	CATGGGATATGGAGC GATCG	GGGGTCGTAGGAGAT CTTGA	66	66	791	Target 490,900– 491,416	65	65
3	CCGGTTGTCCATTCC GTTTA	CTGTACGTATTTGGG TTGCC	64	66	454	Target 1,303,831– 1,303,911	65	64
4	GGATGCGAGCTATAT CTCCG	AATACGCCGAGATGT GGATG	65	65	458	Target 1,674,182– 1,674,222	65	65
5	CAACAGTTCATCCCG GTTCG	GACGGATTTGTCGCT CACTA	65	67	759	Target 2,288,681– 2,289,242	66	64
6	GCCACCATCGAATAT CTGGT	GCTCCAGGAAAGGGAA TCATC	66	65	778	Target 761,007– 761,277	64	65
7	GCATACCGAACGTCA CAGAT	AGGTCACCTACAAA AACGG	66	65	665	Target 778,990– 779,488	65	65
8	GCTCTTAAGGCTGGC AATCT	CGGTCACACTTTCCGG TAAGA	65	66	577	Target 2,154,831– 2,154,873	65	64

Annealing temperatures (Tms) are in Celsius (Tms have been rounded for illustration purposes), and the product sizes are given in nucleotides

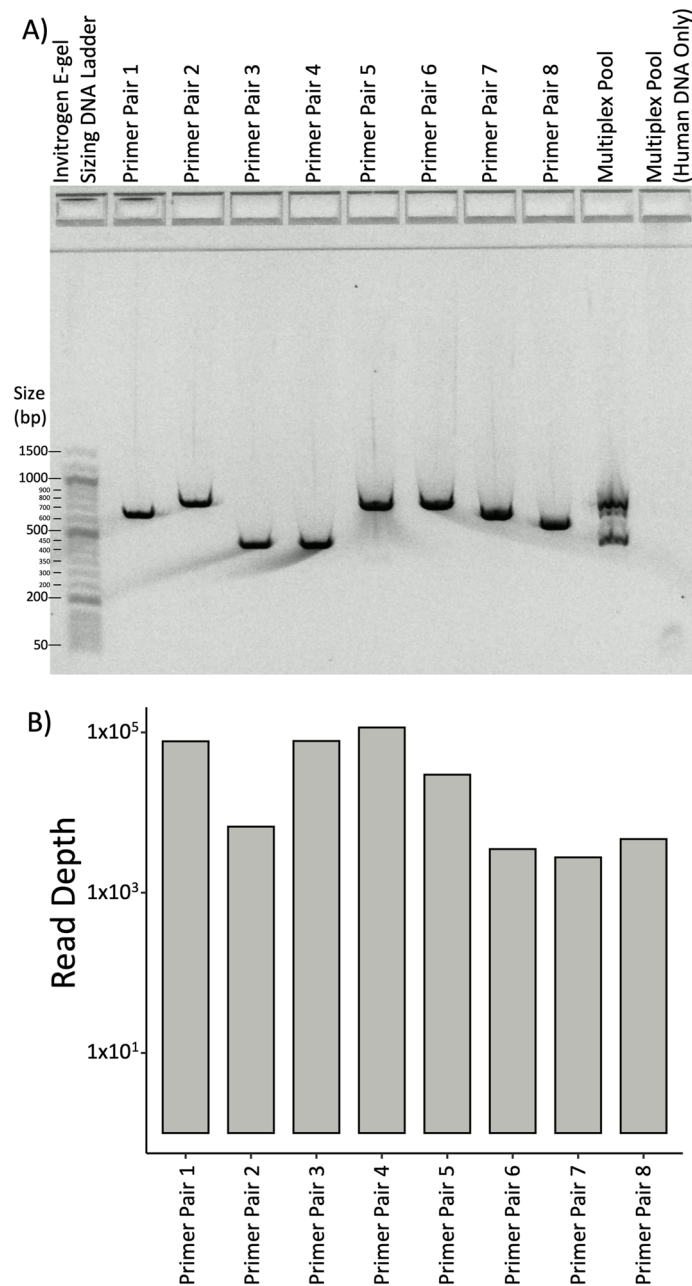


Fig. 2 **A** Electrophoresis gel showing expected amplicon sizes and specificity for each individual and pooled primer pairs and **B** coverage of the sequencing of the pooled multiplex PCR from lane 8

Discussion

primerJinn provides a user-friendly, efficient, and accurate method for designing multiplex PCR primers and performing in silico PCR. We demonstrate that primerJinn can design a multiplex PCR primer for a reaction using Q5 high-fidelity polymerase for an Illumina targeted sequencing and generate similar coverage per target. There are several limitations to primerJinn. First, primerJinn requires the user to input the regions of interest and cannot fix any overlapping regions in this list. Second, it is only designed for

designing primers for sequencing and cannot design primers producing different amplicon sizes for end-point PCR bases analysis, nor can it design primer probe pairs for qPCR or degenerate primers. Finally, it does not have a graphical user interface, which could be a barrier for some users. primerJinn can be used for various applications in molecular biology and bioinformatics research, including the design of assays for amplifying and sequencing drug-resistant conferring regions in important pathogens.

Conclusion

primerJinn provides a user-friendly, efficient, and accurate method for designing multiplex PCR primers for targeted sequencing and performing in silico PCR.

Availability and requirements

Project name: primerJinn. Project home page: <https://github.com/SemiQuant/PrimerJinn>. Operating system(s): Platform independent. Programming language: Python. Other requirements: BLAST+. License: GNU GPL.

Any restrictions to use by non-academics: License needed.

Acknowledgements

Not applicable.

Author contributions

JL and JM wrote and reviewed the manuscript. JL performed the experiments and prepared the figures.

Funding

NIAID grant award R01AI153213.

Availability of data and materials

Code and raw data are available at <https://github.com/SemiQuant/PrimerJinn>.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Received: 5 June 2023 Accepted: 8 December 2023

Published online: 12 December 2023

References

1. Kechin A, Borobova V, Boyarskikh U, Khrapov E, Subbotin S, Filipenko M. NGS-PrimerPlex: high-throughput primer design for multiplex polymerase chain reactions. *PLoS Comput Biol*. 2020;16(12):e1008468. <https://doi.org/10.1371/JOURNAL.PCBI.1008468>.
2. Yuan J, Yi J, Zhan M, et al. The web-based multiplex PCR primer design software Ultiplex and the associated experimental workflow: up to 100-plex multiplicity. *BMC Genom*. 2021;22(1):1–17. <https://doi.org/10.1186/S12864-021-08149-1/TABLES/4>.
3. Xie NG, Wang MX, Song P, et al. Designing highly multiplex PCR primer sets with simulated annealing design using dimer likelihood estimation (SADDLE). *Nature Commun*. 2022;13(1):1–10. <https://doi.org/10.1038/s41467-022-29500-4>.
4. Francis F, Dumas MD, Wisser RJ. ThermoAlign: a genome-aware primer design tool for tiled amplicon resequencing. *Sci Rep*. 2017;7(1):1–15. <https://doi.org/10.1038/srep44437>.
5. Brown SS, Chen YW, Wang M, Clipson A, Ochoa E, Du MQ. PrimerPooler: automated primer pooling to prepare library for targeted sequencing. *Biol Methods Protoc*. 2017. <https://doi.org/10.1093/BIOMETHODS/BPX006>.
6. Hammet F, Mahmood K, Green TR, et al. Hi-Plex2: a simple and robust approach to targeted sequencing-based genetic screening. *Biotechniques*. 2019;67(3):118–22. <https://doi.org/10.2144/BTN-2019-0026>.

7. Yuan J, Yi J, Zhan M, et al. The web-based multiplex PCR primer design software Ultiplex and the associated experimental workflow: up to 100-plex multiplicity. *BMC Genom.* 2021;22(1):1–17.
8. Untergasser A, Cutcutache I, Koressaar T, et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 2012;40(15):e115. <https://doi.org/10.1093/NAR/GKS596>.
9. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
10. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Published online March 16, 2013. <https://arxiv.org/abs/1303.3997v2>. Accessed 29 Aug 2023.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

