Journal of Pathology Informatics 17 (2025) 100442

Contents lists available at ScienceDirect

Journal of Pathology Informatics

journal homepage: www.elsevier.com/locate/jpi

ARMSprimer3: An open-source primer design Python program for amplification refractory mutation system PCR (ARMS-PCR)

Jingwen Guo [a], Jeremy Grojean [b], Huazhang Guo [b,*]

[a] Department of Computer Science, Princeton University, 35 Older Street, Princeton, NJ 08540, United States of America
[b] Department of Pathology, Saint Louis University, 1402 S. Grand Blvd, St. Louis, MO 63104, United States of America

ARTICLE INFO

Keywords:
Single nucleotide polymorphisms (SNPs)
Amplification refractory mutation system PCR (ARMS-PCR)
Python
PCR primer design

ABSTRACT

Single nucleotide polymorphisms (SNPs) are DNA sequence variations of a single base pair. They are the underlying mechanism of most human genetic variation and etiology of many heritable human diseases. SNPs can be reliably detected by amplification refractory mutation system PCR (ARMS-PCR). ARMS-PCR is based on allele-specific PCR primers that only amplify DNA samples with the target allele and do not amplify DNA samples without the target allele. In addition to the allele-specific mismatch at the 3′ end, ARMS-PCR introduces additional deliberate mismatches near the 3′ end of the allele-specific primers to further destabilize the non-specific binding and priming on non-target alleles. This modification increases the specificity for SNP detection, but also increases the complexity of PCR primer design. We developed ARMSprimer3, a Python program to automate the ARMS-PCR primer design process. The validity of ARMSprimer3 was confirmed by successfully using it to develop diagnostic tests in our clinical molecular diagnostic lab. ARMSprimer3 is open-source software and can be freely downloaded from https://github.com/PCRPrimerDesign/ARMSprimer3.

Introduction

Single nucleotide polymorphisms (SNPs) are DNA sequence variations of a single base pair (bp), with each nucleotide variation giving rise to an allele. They are the underlying mechanism of most human genetic variation and etiology of many heritable human diseases. When they are associated with disease, the major allele (the more frequent allele) is also called wild type allele, and the minor allele (the less frequent allele) is also called the mutation allele. SNPs can be classified into common SNPs and rare SNPs, based on the minor allele frequency. Common SNPs have minor allele frequencies of no less than 1%, whereas rare SNPs have minor allele frequencies of under 1%. SNPs can be reliably detected by many molecular methods, such as sequencing, digital PCR, primer extension, and amplification refractory mutation system PCR (ARMS-PCR). Although next-generation sequencing is becoming popular for massive parallel SNP detection, PCR is still a more cost-effective technology for SNP detection when massive parallel detection is not the goal.

ARMS-PCR is based on allele-specific PCR primers that amplify DNA samples with the target allele and do not amplify DNA samples without the target allele.[1,2] In addition to the allele specific mismatch at the 3′ end, ARMS-PCR introduces additional deliberate mismatches near the 3′ end of the allele-specific primers to further destabilize the non-specific binding and priming on non-target alleles.[3] This modification increases the specificity of SNP detection, but also increases the complexity of the PCR primer design process. Although a free web-hosted computer program, Primer1,[4] was developed to ease the burden for a special ARMS-PCR, tetra-primer ARMS-PCR, there are currently no computer programs to facilitate general ARMS-PCR primer design.

In our molecular diagnostic labs, our old workflow relied on manually designing ARMS-PCR primers. It was a tedious process and required extensive knowledge of molecular biology, PCR primer design, and multiple bioinformatics tools. It took an experienced molecular biologist about 2 days to design one set of ARMS-PCR primers. The manual process also tended to introduce human error. Therefore, our goal was to develop an automatic computer software to facilitate general ARMS-PCR primer design.

Software design and implementation

Software design

The design goals of ARMSprimer3 were to: (1) be a general and flexible ARMS-PCR primer design software for broad applications; (2) be immune to primer binding interferences by neighboring common SNPs; (3) require minimal end user input through maximal automation; (4) be easy to modify and maintain. ARMSprimer3 achieved the design goals by: (1) implementing the ARMS specific additional mutation mechanism in the core

* Corresponding author at: Department of Pathology, Saint Louis University School of Medicine, 1402 S. Grand Blvd., St. Louis, MO 63104, USA.
E-mail address: huazhang.guo@ssmhealth.com (H. Guo).

http://dx.doi.org/10.1016/j.jpi.2025.100442
Received 13 October 2024; Received in revised form 25 March 2025; Accepted 13 April 2025
Available online 15 April 2025

program[5]; (2) avoiding designing primers on neighboring interfering common SNP loci; (3) integrating with the flexible and reliable free primer design software (primer3_core[6,7]) and UCSC genome browser online databases[8]; (4) being open source and implemented in Python 3 for its accessibility and popularity among bioinformaticians. In the following text and figures, Factor V Leiden mutation (SNP ID: rs6025) is used to illustrate the software usage and outputs.

ARMSprimer3 input

ARMSprimer3 is a Python program running from the UNIX command line (see Fig. 1A). A target human SNP ID is the only mandatory input from the end user. The provided default primer design parameters, such as preferred primer/product length and primer melting temperature (T_m), are tailored for SYBER green-based real-time ARMS-PCR. However, they can be optionally modified to tailor to other variants of ARMS-PCR, such as florescent probe-based real-time ARMS-PCR and tetra-primer ARMS-PCR. The parameters used by ARMSprimer3 are identical to primer3_core and details can be found in the documentation of primer3_core.

ARMSprimer3 workflow (see Fig. 2)

ARMSprimer3 downloads the detailed information of the target SNP from the UCSC Genome Browser MySQL database. From the same database, it then downloads a 1001-bp genomic sequence centered on the target SNP, with 500-bp upstream and 500-bp downstream of the target SNP (Fig. 1A). ARMSprimer3 also downloads detailed information of all neighboring common SNPs located in the downloaded genomic sequence (Fig. 1B). The corresponding location of the neighboring common SNPs are masked with "n" characters in the genomic sequence (yellow arrows in Fig. 1C, D, E, and F). The masking prevents designing primers on the neighboring common interfering SNP loci, thereby ensuring maximal primer binding strength and accuracy of the predicted primer T_m.

Based on the masked genomic sequence, ARMSprimer3 generates four variants of the sequence (Fig. 3): (1) wild type allele of the target SNP with additional mutations at −2 positions; (2) wild type allele with additional mutations at −3 positions (comparing capitalized nucleotides indicated by pink arrows in Fig. 1C and E); (3) mutation allele with additional mutations at −2 positions; and (4) mutation allele with additional mutations at −3 positions (comparing capitalized nucleotides indicated by pink arrows in Fig. 1D and F). Then, ARMSprimer3 drives primer3_core program to design ARMS primers around the target SNP locus based on all four templates. For each template, ARMSprimer3 tries two ways to design ARMS-PCR primers: (1) force the 3′ end of the left primer to be at the target SNP position (Fig. 1E and F); (2) force the 3′ end of the right primer to be at the target SNP position. The two ways are referred to as force-left and force-right, respectively.

ARMSprimer3 output

In addition to primer files containing designed ARMS-PCR primers, ARMSprimer3 saves additional intermediate files: (1) genomic sequence file, (2) a list of all SNPs in the genomic sequence, (3) masked genomic sequence where the loci of common interfering SNPs are replaced with character "n", (4) masked wild type or mutation sequences with additional mutations at −2 or −3 locations, (5) parameters used by primer3_core, and (6) ARMSprimer3 screen output. These intermediate files can be used for debugging and other purposes.

Results

We developed an open-source Python program, ARMSprimer3, to automate the ARMS-PCR primer design process. ARMSprimer3 requires a UNIX operating system, Internet connection, and three free computer programs to be in the executable path, i.e. *primer3_core* (primer3.org), UCSC (University of California, Santa Cruz) *twoBitToFa*, and MySQL client *mysql* (mysql.com). ARMSprimer3 takes a human SNP ID as the mandatory input and a primer-design parameter file as an optional input. It automatically retrieves all necessary information and designs ARMS-PCR primers around the target SNP.

Before we developed ARMSprimer3, we manually designed ARMS-PCR primers for factor V Leiden mutation. Our molecular biologist spent 1 h downloading the reference genomic sequence (1001-bp) containing the wild type target allele and then created a mutation sequence containing the mutation target allele. After that, he spent 7 h downloading all related common SNPs and manually annotating their locus with an "n" in each sequence template. He spent another 3 h creating additional mutations in each sequence template at the −2 and −3 positions by following ARMS-PCR rules.[5] Finally, he spent 2 h running primer3_core to design primers. A total of 13 h was spent by a person with expertise in molecular biology and bioinformatics to design ARMS-PCR primers for a single target SNP. During the tedious sequence editing processes to do "n" masking and additional mutating at −2 and −3 positions, at least five errors were introduced into the sequence templates. All errors were corrected by manual double and triple checks. In contrast, it typically takes ARMSprimer3 20–50 s (depending on the speed of the Internet connection and the response time of the UCSD Genome Browser database) to design primers for one target SNP, with no errors as per a manual check. Because ARMSprimer3 automates the manual process with identical steps, the primers designed are identical, but time savings are significant.

The validity of ARMSprimer3 was confirmed by successfully developing four diagnostic tests using ARMSprimer3 in the molecular diagnostic laboratory at Saint Louis University. The PCR primer sequences designed are as follows. For factor V Leiden mutation (rs6025), see Fig. 1E and F (pink boxes). For prothrombin G20210A mutation (rs1799963), the common forward primer is "ttgtgtttctaaaactatggttccca", the wild type reverse primer is "gcactgggagcattgaggAtc", and the mutation type reverse primer is "gcactgggagcattgaggAtT". For hereditary hemochromatosis-related C282Y (rs1800562) mutation, the common reverse primer is "ccagatcacaatgaggggct", the wild type forward primer is "tggggaagagcagagatatacTtg", and the mutation type forward primer is "tggggaagagcagagatatacTtA". For hereditary hemochromatosis-related H63D mutation (rs1799945), the common reverse primer is "tggaaacccatggagttcgg", the wild type forward primer is "cagctgttcgtgttctatgGtc", and the mutation type forward primer is "cagctgttcgtgttctatgGtG". By using the default primer-design parameter file, all primers designed have T_m values close to 60 °C, and all PCR products are shorter than 150-bp. This enabled us to use a standard two-step SYBER green real-time PCR cycling condition for all four diagnostic tests, i.e., 95 °C for 15 s, followed by 60 °C for 60 s, for a total of 50 cycles. For each test sample, we ran two PCR reactions: one reaction with the common primer and the wild type primer to amplify the wild type allele, and the other reaction with the common primer and the mutation primer to amplify the mutation allele. Therefore, after PCR amplification, each test sample produced two Ct (Cycle Threshold) values: one for the wild type allele, and the other for the mutation allele. The allele status of the test sample was determined by comparing the two Ct values. All four diagnostic tests achieved 100 % accuracy by testing 30 reference DNA samples (30 DNA samples with known genotype for each genetic locus) from reference laboratories. All four diagnostic tests required only a minimum of 5 ng of input DNA. After the validation data were approved by the Collage of American Pathologists (CAP), we began routinely testing patient samples, and the tests were monitored by CAP proficiency testing programs twice a year. The four diagnostic tests achieved 100 % accuracy for all CAP proficiency test programs.

By default, ARMSprimer3 designs primers by downloading SNPs and genomic DNA sequences from the UCSC genome database of *Homo sapiens*. To design primers for other species, the database source can be modified to point to the UCSC genome databases of other species.

Fig. 1. A sample UNIX command line input to run ARMSprimer3 and selected outputs of ARMSprimer3. Factor V Leiden mutation (SNP ID: rs6025) is used to illustrate the software usage and outputs. Screenshot (A) shows the UNIX command line input to run ARMSprimer3 and the screen output of ARMSprimer3, including the retrieved detailed information of the target SNP and its genomic sequence around the target SNP (500 base pairs on each side of the target SNP, total of 1001 base pairs). The central five nucleotides are capitalized with the target SNP wild type allele in the center (pink arrows). ARMSprimer3 retrieves detailed information of all common SNPs located in the downloaded genomic sequence (B). The common SNPs in the wild type (C) and mutation (D) genomic sequences are masked with "n" characters (yellow arrows) to prevent designing primers on them. To introduce additional deliberate mismatches into the PCR primers, additional mutations are generated at the −3 position on both sides of the target SNP (pink arrows in yellow boxes, also see Fig. 3) on the masked wild type (E) and mutation (F) genomic sequences. The ARMS-PCR primers (pink boxes and "/<" characters in yellow boxes) are designed based on the "n" masked and additionally mutated DNA sequence templates (yellow boxes). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Discussion

We developed ARMSprimer3 to replace the time-consuming manual primer design process. In contrast to our previous 2-day manual process, ARMSprimer3 automatically designs multiple sets of ARMS-PCR primers within seconds. It can be run by lab personnel without expertise in molecular biology and PCR primer design or familiarity with multiple bioinformatics tools.

ARMSprimer3 is not the only computer software that designs primers for ARMS-PCR. Primer1[4] is a web-based software that designs primers for tetra-primer ARMS-PCR. To use Primer1, users need to manually download the genomic DNA sequence including the target SNP, annotate the position of the target SNP, annotate the wild type and mutation allele nucleotides of the target SNP, and upload them through the Primer1 web-interface. In contrast, ARMSprimer3 automates the whole process. The other limitation of Primer1 is that it can only design primers for tetra-primer ARMS-PCR. In contrast, ARMSprimer3 can design primers for most variations of ARMS-PCR: (1) it defaults to design primers for SYBR Green real-time ARMS-PCR; (2) by changing the parameter file, it can design primers for probe-based real-time ARMS-PCR or electrophoresis-based tetra-primer ARMS-PCR.

One limitation is that ARMSprimer3 sometimes may not be able to find appropriate primer sets with the default setting. This might be due to either the PCR design parameters being too stringent or the common interfering SNPs being too dense in the region near the target SNP. The solutions to this problem are to either relax the primer design parameters or to turn the common interfering SNP mask off.

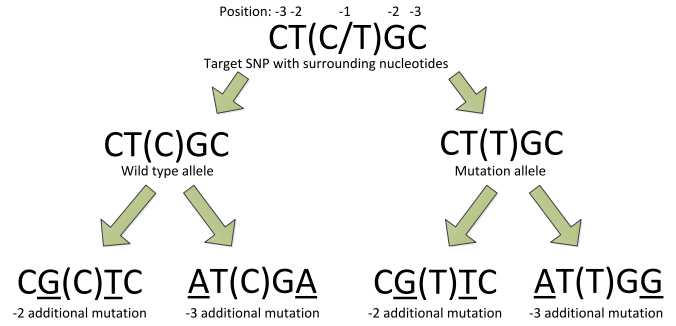

Fig. 3. Diagram showing additional mutations generated by ARMSprimer3 to further destabilize the non-specific binding and priming on non-target alleles. Factor V Leiden mutation (SNP ID: rs6025) is used to illustrate the positions and nucleotides of the additional mutations. Notes: (1) The target locus is in parentheses. (2) Additional −2 or −3 mutations are underlined. (3) There are two −2 mutations or two −3 mutations on either side of the target locus. (4) Also see Fig. 1 for more sequence context of Factor V Leiden mutation.

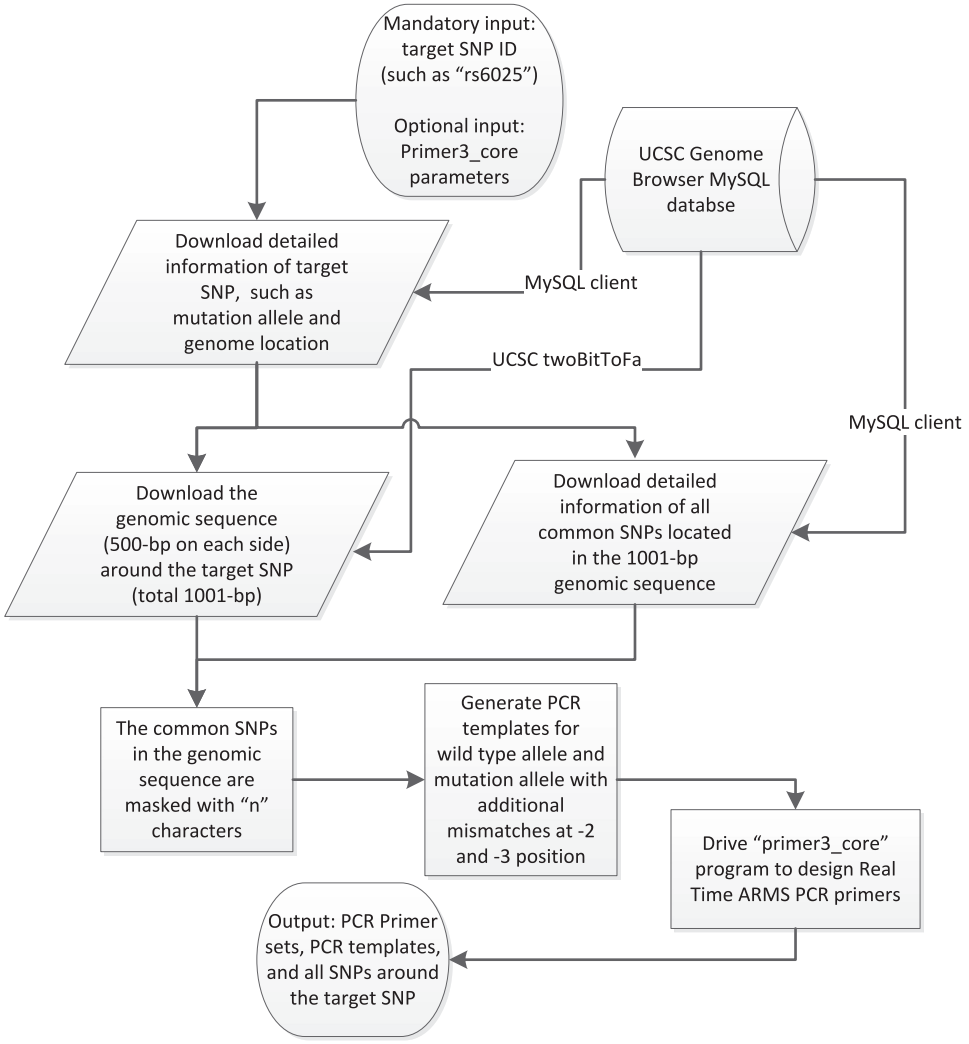

Fig. 2. Diagram of the ARMSprimer3 workflow.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

References

1. Newton CR, Graham A, Heptinstall LE, et al. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res Apr 11, 1989;17(7):2503–2516. https://doi.org/10.1093/nar/17.7.2503.
2. Wu DY, Ugozzoli L, Pal BK, Wallace RB. Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. Proc Natl Acad Sci USA Apr 1989;86(8):2757–2760. https://doi.org/10.1073/pnas.86.8.2757.
3. Ferrie RM, Schwarz MJ, Robertson NH, et al. Development, multiplexing, and application of ARMS tests for common mutations in the CFTR gene. Am J Hum Genet Aug 1992;51(2):251–262.
4. Ye S, Dhillon S, Ke X, Collins AR, Day IN. An efficient procedure for genotyping single nucleotide polymorphisms. Nucleic Acids Res Sep 1, 2001;29(17):E88. https://doi.org/10.1093/nar/29.17.e88.
5. Little S. Amplification-refractory mutation system (ARMS) analysis of point mutations. Curr Protoc Hum Genet May 2001. https://doi.org/10.1002/0471142905.hg0908s07. Chapter 9:Unit 9 8.
6. Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. Bioinformatics May 15, 2007;23(10):1289–1291. https://doi.org/10.1093/bioinformatics/btm091.
7. Untergasser A, Cutcutache I, Koressaar T, et al. Primer3–new capabilities and interfaces. Nucleic Acids Res Aug 2012;40(15), e115. https://doi.org/10.1093/nar/gks596.
8. Nassar LR, Barber GP, Benet-Pages A, et al. The UCSC genome browser database: 2023 update. Nucleic Acids Res Jan 6, 2023;51(D1):D1188–D1195. https://doi.org/10.1093/nar/gkac1072.