Associate PCR-RFLP Assay Design With SNPs Based on Genetic Algorithm in Appropriate Parameters Estimation

Li-Yeh Chuang, Yu-Huei Cheng, *Member, IEEE*, Cheng-Huei Yang, and Cheng-Hong Yang*, *Senior Member, IEEE*

***Abstract*—Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) is a commonly used laboratory technique and useful in small-scale basic research studies of complex genetic diseases that are associated with single nucleotide polymorphisms (SNPs). Before PCR-RFLP assay for SNP genotyping can be performed, a feasible primer pair observes numerous constraints and an available restriction enzyme for discriminating a target SNP, are required. The computation of feasible PCR-RFLP primers and find available restriction enzymes simultaneously aim at a target SNP is a challenging problem. Here, we propose an available method which combines the updated core of SNP-RFLPing with a genetic algorithm to reliably mine available restriction enzymes and search for feasible PCR-RFLP primers. We have in silico simulated the method in the SLC6A4 gene under different parameter settings and provided an appropriate parameter setting. The wet laboratory validation showed that it indeed usable in providing the available restriction enzymes and designing feasible primers that fit the common primer constraints. We have provided an easy and kindly interface to assist the researchers designing their PCR-RFLP assay for SNP genotyping. The program is implemented in JAVA and is freely available at http://bio.kuas.edu.tw/ganpd/.**

***Index Terms*—Genetic algorithm (GA), genotyping, polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP), primer design, single nucleotide polymorphisms (SNPs).**

I. Introduction

SINGLE nucleotide polymorphism (SNP) genotyping plays an important role in population genetics and evolutionary studies [1], pharmacogenetic analysis [2], malignancy studies [3], [4], preventive medicine [5], [6], personalized medicine [7] and forensics [8]. Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) is a simple laboratory technique implemented to investigate the causes of genetic variations and mutations; it is especially useful in small-scale research studies of complex genetic diseases [9]. The relatively simple, inexpensive and accurate characteristics make PCR-RFLP be a popular method [9], [10]. In PCR-RFLP assay applying to SNPs, specific restriction enzymes are essential to discriminate SNP genotype. Before performing PCR-RFLP, a feasible primer pair observes numerous constraints are also required.

To date, a few systems provide PCR-RFLP assay design, such as V-MitoSNP, identifies the available restriction enzymes from REBASE [11] and designs PCR-RFLP primer sets for all mt-SNPs [12]. However, only several simple constraints used make it lack in stringent quality for PCR-RFLP assay. Besides, it also only provides PCR-RFLP assay for mitochondrial SNPs. SNP Cutter employs a pre-selected or customizable list of restriction enzymes and uses Primer3 [13] to look for PCR-RFLP primer sets. It is a helpful tool for PCR-RFLP, but it requires a specific format (or NCBI dbSNP reference ID) as input and does not design good PCR-RFLP primers since the Primer3 program has not been updated. Prim-SNPing is an improved software tool with PCR-RFLP assay design function for cost-effective SNP genotyping [14]. Nevertheless, the incorporated window-sliding strategy makes it always search inefficiently. For the above reasons, the development of an improved method for PCR-RFLP assay design is greatly mandated.

PCR-RFLP assay design is a very challenging task since numerous primer constraints must be conformed to [13], [15]–[17] and available restriction enzymes need to provided simultaneously aim at a target SNP. Numerous primer constraints typically include the primer length, primer length difference, PCR product size, GC proportion, melting temperature (T_{m}), melting temperature difference ($T_{\mathrm{m-diff}}$), GC clamp, dimers (including cross-dimers and self-dimers), hairpin structure, and specificity. Available restriction enzymes are pivotal to discriminate the genotype of the target SNP. Many primer design approaches have been proposed, e.g., dynamic programming [18], genetic algorithm (GA) [17], memetic algorithm (MA) [19], [20], particle swarm optimization (PSO) [21], parthenogenetic algorithm MG-PGA [22], greedy algorithm [23], heuristic algorithm [24], and others. Nevertheless, most of these methods do not focus on the design of PCR-RFLP assay for SNP genotyping. Recently, mismatch PCR-RFLP based on GA is proposed [25]. In the past, we introduced a simple genetic algorithm (GA) to design PCR-RFLP assay and publish a brief version in an international conference [26]. The intent of the conference is communication and feedback among researchers. We proposed the

Manuscript received September 16, 2012; revised March 10, 2013; accepted April 08, 2013. Date of current version May 29, 2013. This work is partly supported by the National Science Council (NSC) in Taiwan under grant NSC99-2622-E-151-019-CC3, NSC99-2221-E-151-056-, and NSC101-2221-E-464-001-. *Asterisk indicates corresponding author.*

L.-Y. Chuang is with the Department of Chemical Engineering & Institute of Biotechnology and Chemical Engineering, I-Shou University, Kaohsiung 84008, Taiwan (e-mail: chuang@isu.edu.tw).

Y.-H. Cheng is with the Department of Digital Content Design and Management, Toko University, Chiayi 61363, Taiwan (e-mail: yuhuei.cheng@gmail.com).

C.-H. Yang is with the Department of Electronic Communication Engineering, National Kaohsiung Marine University, Kaohsiung 81157, Taiwan (e-mail: chyang@mail.nkmu.edu.tw).

*C.-H. Yang is with the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung 80778, Taiwan (e-mail: chyang@cc.kuas.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNB.2013.2258469

1536-1241/$31.00 © 2013 IEEE

Authorized licensed use limited to: b-on: Instituto Politecnico de Tomar. Downloaded on April 19,2024 at 18:43:55 UTC from IEEE Xplore. Restrictions apply.

simpler method there and discussed it with peers at the conference. Some ideas and problems were collected at this conference. These inspired us to improve the proposed method. In order to increase the quality and scientific value of the study, we conducted new research different from the original conference paper. In this study, we propose an available method which combines the updated core of SNP-RFLPing [27]–[29] with a genetic algorithm (GA) to reliably mine available restriction enzymes and search for feasible PCR-RFLP primers. The different parameter settings in the proposed method are *in silico* evaluated and the appropriate parameter settings are found to design the PCR-RFLP assay. A wet laboratory validation showed its usability. The user-friendly interface is provided and freely available at http://bio.kuas.edu.tw/ganpd/.

II. Materials and Methods

A. Problem Description and Definition

Let T_D be a DNA template sequence that is made up by nucleotide codes of the DNA with an identified SNP. T_D is defined as follows:

$$T_D = \{B_i | i \text{ is the index of DNA sequence}, \exists! B_i \in \text{SNP}\} \quad (1)$$

where B_i represents the regular nucleotides "A," "T," "C," "G," the SNP IUPAC code (M, R, W, S, Y, K, V, H, D, B, or N) or the dNTPs format ([dNTP1/dNTP2]). The symbol $\exists!$ represents the existence and uniqueness. We focus only on true SNPs as described in dbSNP [30] of NCBI as target SNPs, i.e., deletion/insertion polymorphisms (DIPs) and multi-nucleotide polymorphisms (MNPs) are not included.

The PCR-RFLP assay design problem now consists of designing a pair of sub-sequences of corresponding constraints from T_D and finding at least one restriction enzyme to distinguish the genotype of the target SNP. One sub-sequence is called the forward primer (P_f) and the other is called the reverse primer (P_r). The forward primer and the reverse primer are defined as follows:

$$P_f = \{B_i | \forall B_i \in \{\text{`A', `T', `C', `G'}\}, F_s \le i \le F_e, i \text{ is the index of } T_D\} \quad (2)$$

$$P_r = \{\overline{B_i} | \forall B_i \in \{\text{`A', `T', `C', `G'}\}, R_s \le i \le R_e, i \text{ is the index of } T_D\} \quad (3)$$

where F_s and F_e denote the start index and the end index of P_f in T_D; R_s and R_e denote the start index and the end index of P_r in T_D. P_f and P_r are called a primer pair. $\{\overline{B_i}\}$ is the anti-sense sequence of B_i. For the sequence B_i = "GTCTACGTCGAAC," for example, the complement sequence is "CAGATGCAGCTTG," since the complement nucleotide of "A" is "T" and the complement nucleotide of "C" is "G." The anti-sense sequence $\{\overline{B_i}\}$ is the reverse of the complement sequence, i.e., $\{\bar{B}\} =$ "GTTCGACGTAGAC."

A simple vector (called "individual" in a GA) consisting of F_s, F_l, P_l and R_l (shown as Fig. 1) can determine a PCR-RFLP assay. We define this vector as:

$$P_v = (F_s, F_l, P_l, R_l) \quad (4)$$

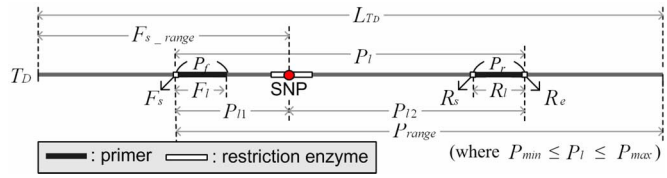

Fig. 1. Parameters for PCR-RFLP assay design. The following symbols are used: F_s : Start position of the forward primer; R_e : End position of the reverse primer; F_l : Length of the forward primer; P_l : PCR product length between F_s and R_e; R_l : Length of the reverse primer; R_s : Start position of the reverse primer; P_{l1} : PCR product length between F_s and SNP; P_{l2} : PCR product length between SNP and R_e; F_{s_range} : Random range of F_s; $P_{\min}$: Minimum PCR product length; $P_{\max}$: Maximum PCR product length ($P_{\min}$ and $P_{\max}$ are provided by user preset values); P_{range} : Length from F_s to the DNA template end, and L_{TD} : Length of DNA template.

With P_v, we can calculate the start position of the reverse primer by:

$$R_s = F_s + P_l - R_l \quad (5)$$

Consequently, the forward primer and the reverse primer can be obtained from P_v. P_v is the encoding prototype in the PCR-RFLP assay design problem. In later sections, P_v is used to perform evolutionary computations of the GA.

B. The Proposed Method

The proposed method contains six separate processes, a flowchart of which is shown in Fig. 2. These processes are 1) mining of SNP-RFLP restriction enzymes, 2) judgment on availability of restriction enzymes, 3) creation of a random initial population, 4) evaluation of the fitness function, 5) judgment of termination criteria, and 6) selection, crossover, mutation, and replacement operations, respectively. These processes are described below.

1) Mining of SNP-RFLP Restriction Enzymes: First, the available restriction enzymes for discriminating a target SNP must be mined. The proposed method uses the updated core of SNP-RFLPing [27]–[29], which is a time-saving application for mining restriction enzymes for RFLP assays. It was developed by our team in order to obtain available restriction enzymes.

2) Judgment on Availability of Restriction Enzymes: A target SNP must be recognized by the restriction enzymes obtained through the mining process. If no restriction enzyme is available to distinguish the alleles of the target SNP, the natural PCR-RFLP primer design is insignificant for SNP genotyping, and the proposed method is terminated. Otherwise, the proposed method proceeds with the following processes.

3) Creation of a Random Initial Population: A fixed number of individuals (P_v) are randomly generated for an initial population without duplicates. F_s is randomly generated between 1 and $(L_{\text{TD}} - P_{\min} + 1)$. This range is labeled F_{s_range}. F_l is randomly generated between the minimum and the maximum primer length. The minimum primer length is set to 16 nt and the maximum primer length is set to 28 nt [17]. In order to limit the PCR product length, the proposed method randomly generates P_l between $P_{\min}$ and $P_{\max}$ ($P_{\min}$ and $P_{\max}$ are provided by user preset values), i.e., P_{range}. R_l is randomly generated in the same way as F_l.

Authorized licensed use limited to: b-on: Instituto Politecnico de Tomar. Downloaded on April 19,2024 at 18:43:55 UTC from IEEE Xplore. Restrictions apply.

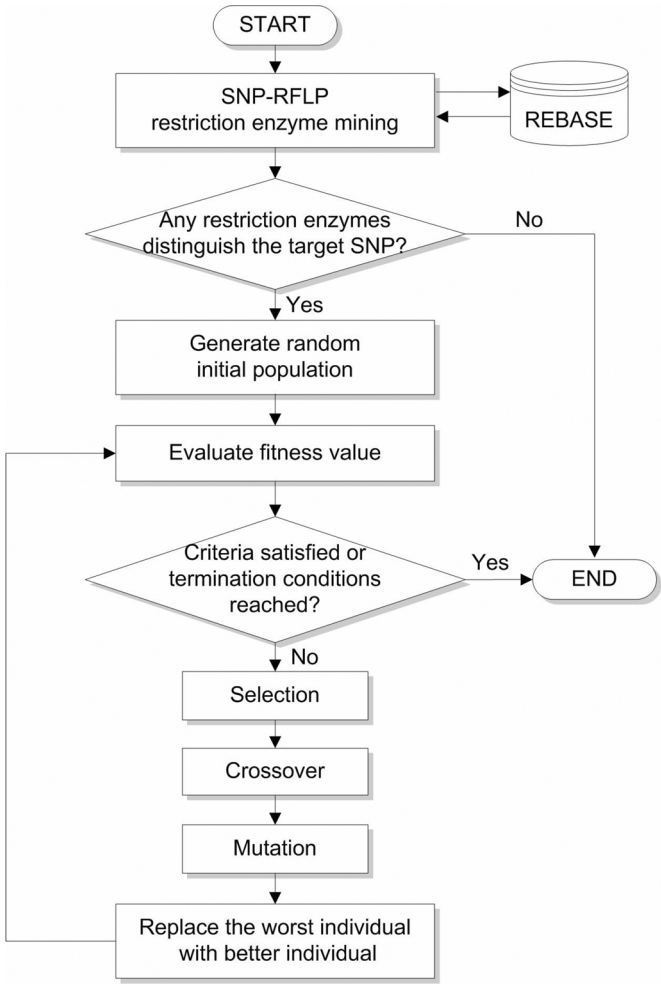

Fig. 2. Flowchart of PCR-RFLP assay design. At first, the restriction enzymes for the target SNP are mined. Then a judgment of whether the restriction enzymes distinguish the target SNP is done. If no restriction enzymes are available, the algorithm stops; else the algorithm proceeds with the following processes. A random initial population is generated, and then the fitness values of all individuals in the population are calculated with the fitness.

4) Evaluation of the Fitness Function: The fitness value of each individual is evaluated in turn by a fitness function. In order to make a primer pair satisfies the design constraints, primer constraints are used to evaluate the fitness value. The fitness value is minimized (i.e., a fitness value equal to 0 indicates the best natural primer pair). The proposed fitness function (6) in the method consists of nine primer constraint functions.

$$\begin{aligned}&\mathrm{Fitness}(P_v)\\&\quad=3\times(\mathrm{Len}_{\mathrm{diff}}(P_v)+\mathrm{GC}_{\mathrm{proportion}}(P_v)\\&\qquad+\mathrm{GC}_{\mathrm{clamp}}(P_v))\\&\qquad+10\times(Tm(P_v)+Tm_{\mathrm{diff}}(P_v)\\&\qquad+\mathrm{dimer}(P_v)+\mathrm{hairpin}(P_v))\\&\qquad+50\times\ \mathrm{specificity}\ (P_v)\\&\qquad+60\times\ \mathrm{product}(P_v)\end{aligned}\tag{6}$$

In the fitness function, weights are used to discriminate the significance of each primer constraint function. We use four different weight values to represent different degrees of importance for these functions, namely 3, 10, 50, and 60. These weight values were chosen based on our experience and the experimental requirements of the PCR experiments. They can be adjusted by users to satisfy different experimental requirements. A larger weight value represents a more importance constraint function. These functions are described below.

a) $\mathrm{Len}_{\mathrm{diff}}(P_v)$*:* In a PCR experiment, a length difference of no more than 3 bps between the forward primer and the reverse primer is considered optimal. $\mathrm{Len}_{\mathrm{diff}}(P_v)$ is used to check whether the length difference of a primer pair exceeds 3 bps. We use 0 and 1 to represent the two different conditions.

$$\begin{aligned}&\mathrm{Len}_{\mathrm{diff}}(P_v)\\&\quad=\begin{cases}0, & \text{if the absolute value of } (F_l-R_l)\le 3\\1, & \text{other conditions}\end{cases}\end{aligned}\tag{7}$$

b) $Tm(P_v)$*:* The $Tm(P)$ function used here is calculated by the Wallace formula [31] (8), which is simple and considered suitable for approximate neighbor thermodynamic calculations [32]. The $Tm(P_v)$ function is used to check whether the melting temperature Tm of a primer pair lies between 45 °C and 62 °C (9).

$$\begin{aligned}&Tm(P)\\&\quad=(\#\mathrm{G}+\#\mathrm{C})\times 4+(\#\mathrm{A}+\#\mathrm{T})\times 2\end{aligned}\tag{8}$$

$$\begin{aligned}&Tm(Pv)\\&\quad=\begin{cases}0, & \text{if } 45\le Tm(P_f)\le 62 \text{ and } 45\le Tm(P_{\mathrm{r}})\le 62\\1, & \text{if } 45\le Tm(P_f)\le 62 \text{ or } 45\le Tm(P_{\mathrm{r}})\le 62\\2, & \text{other conditions}\end{cases}\end{aligned}\tag{9}$$

c) $T_{\mathrm{m_{diff}}}(P_v)$*:* The $T_{\mathrm{m_{diff}}}(P_v)$ function checks whether the difference of the melting temperature exceeds 5 °C; the criterion is described by (10) using the above $Tm(P)$ function. The difference of the melting temperature is smaller, the better PCR performance. The reason is in the same tube, the melting temperature is identical.

$$\begin{aligned}&T_{\mathrm{m_{diff}}}(P_v)\\&=\begin{cases}0, & \text{if the absolute value of } (Tm(P_f)-Tm(P_{\mathrm{r}}))\le 5\\1, & \text{other conditions}\end{cases}\end{aligned}\tag{10}$$

d) $\mathrm{GC}_{\mathrm{proportion}}(P_v)$*:* The GC proportion in a primer is denoted as GC% (P), a value that indicates the ratio of the nucleotides "G" and "C" that appear in the primer. An appropriate GC ratio in a primer lies in the range of 40–60%. The $\mathrm{GC}_{\mathrm{proportion}}(P_v)$ function is used to check whether the GC% (P) of the forward and reverse primers is between 40% and 60% or not. GC% (P) and GC $_{\mathrm{proportion}}(P_v)$ are defined as:

$$\mathrm{GC}\%(P)=\frac{G_{\mathrm{number}}(P)+C_{\mathrm{number}}(P)}{|P|}\tag{11}$$

where $G_{\mathrm{number}}(P)$ represents the number of nucleotides "G" and $C_{\mathrm{number}}(P)$ represents the number of nucleotides "C."

$$\mathrm{GC}_{\mathrm{proportion}}(P_v)=\begin{cases}0, \text{if } 40\le\ \mathrm{GC}\%(P_f)\le 60\\\qquad\text{and } 40\le\ \mathrm{GC}\%(P_{\mathrm{r}})\le 60\\1,\ \text{if } 40\le\ \mathrm{GC}\%(P_f)\le 60\\\qquad\text{or } 40\le \mathrm{GC}\%(P_{\mathrm{r}})\le 60\\2, \text{other conditions}\end{cases}\tag{12}$$

Authorized licensed use limited to: b-on: Instituto Politecnico de Tomar. Downloaded on April 19,2024 at 18:43:55 UTC from IEEE Xplore. Restrictions apply.

e) $\mathrm{GC}_{\mathrm{clamp}}(P_v)$*:* The function $\mathrm{GC}_{\mathrm{clamp}}(P_v)$ is used to check whether the $3'$ terminal end of a primer is either nucleotide "G" or "C"; it is defined as follows:

$$\mathrm{GC}_{\mathrm{clamp}}(P_v) = \begin{cases} 0, \text{ if } P_f 3' \text{ end is `G' or `C'} \\ \quad \text{and } P_{\mathrm{r}} 3' \text{ end is `G' or `C'} \\ 1, \text{if } P_f \ 3' \text{ end is `G' or `C'} \\ \quad \text{or } P_{\mathrm{r}} \ 3' \text{ end is `G' or `C'} \\ 2, \text{ if neither } P_f 3' \\ \quad \text{nor } P_{\mathrm{r}} \ 3' \text{ is `G' or `C'} \end{cases} \tag{13}$$

f) dimer (P_v)*:* If annealing between two primers takes place, a dimer is formed. The occurrence of dimers during a PCR experiment is problematic. Consequently, this condition should be avoided. Possible dimers include cross-dimers and self-dimers. A cross-dimer is formed when P_f and P_r anneal to each other, and a self-dimer is formed when P_f and P_f, or P_r and P_r anneal to each other. The function *dimer*(P_v) is used to check whether the forward primer and the reverse primer anneal to each other or to themselves.

$$\mathrm{dimer}\ (P_v) = \begin{cases} 0, \text{ if } P_f \text{ and } P_{\mathrm{r}} \text{ do not} \\ \quad \text{form a primer dimer} \\ 1, \text{ if either } P_f \text{ or } P_{\mathrm{r}} \\ \quad \text{forms a cross} - \text{dimer or a self} - \text{dimer} \\ 2, \text{ other conditions} \end{cases} \tag{14}$$

g) hairpin (P_v)*:* A primer may anneal to itself and thus forms a hairpin. The occurrence of this condition influences the results of a PCR experiment. The function *hairpin*(P_v) is used to check for this condition in a primer pair. The *hairpin*(P_v) function can be written as:

$$\mathrm{hairpin}\ (P_v) = \begin{cases} 0, & \text{if neither } P_f \text{ nor } P_{\mathrm{r}} \text{ forms a hairpin} \\ 1, & \text{if either } P_f \text{ or } P_{\mathrm{r}} \text{ forms a hairpin} \\ 2, & \text{if both } P_f \text{and } P_{\mathrm{r}} \text{ form a hairpin} \end{cases} \tag{15}$$

h) specificity (P_v)*:* A PCR experiment might fail if the primer is not site-specific and recurs in the DNA sequence. The *specificity*(P_v) function is defined as the number of recurring instances of P_f and P_r in T_D, and thus determines whether the primer pair is repeated in the template DNA sequence or not.

$$\mathrm{specificity}(P_v) = \text{number of } P_f \text{ and } P_r \text{ recurrences in } T_D \tag{16}$$

i) product (P_v)*:* In order to ensure that the digested allelic fragments can be easily distinguished by electrophoresis, we used a $1:2:3$ ratio as a standard to estimate the PCR-RFLP product size with a tolerance range of 25 bps and a minimum product size of more than 100 bps.

$$\mathrm{product}(P_v) = \begin{cases} 0, & \text{if product size ratio} \\ & \text{corresponds to the criteria} \\ 1, & \text{otherwise} \end{cases} \tag{17}$$

5) Judgment of Termination Criteria: Two termination criteria are used in the proposed method: i) an individual fitness value reaches 0, and ii) a preset number of iterations (generations) is reached. The number of iterations set in our simulated *in silicon* experiments is discussed in the following "Parameter settings" section.

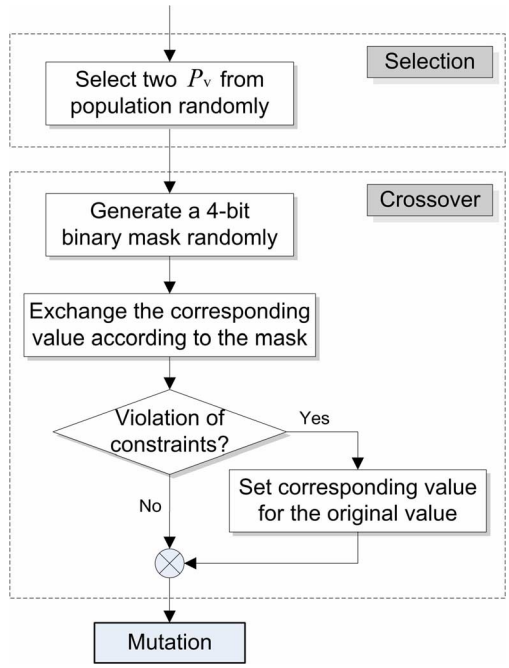

Fig. 3. Selection and crossover flowchart for PCR-RFLP assay design. Initially, the selection operation uses random selection to select two individuals from a population. A 4-bit binary mask is then randomly generated. When the mask is value 1, the corresponding value is exchanged. The exchanged values are judged as to whether they violate the constraint condition or not. If they do, the original value is restored, else the process continues to the next step, i.e., the mutation operation.

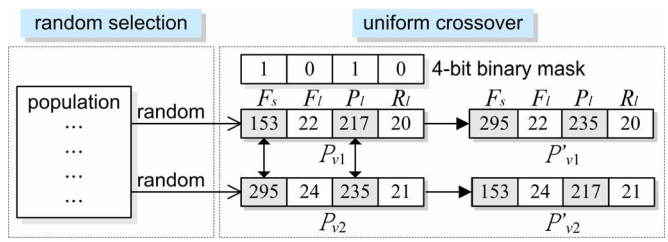

Fig. 4. An example of selection and crossover operations for PCR-RFLP assay design. First, two P_v from the population are randomly selected from population, for example $P_{v1} = (153, 22, 217, 20)$ and $P_{v2} = (295, 24, 235, 21)$. Then, a mask of 4 binary bits is randomly generated, i.e., 1010, so that the values of F_s and P_l are exchanged accordingly. The offsprings $\boldsymbol{P}'_{v1} = (295, 22, 235, 20)$ and $\boldsymbol{P}'_{v2} = (153, 24, 217, 21)$ are examined as to whether the variable values violate the constraints or not. If the variable values violate the constraints, the violating variable value is not exchanged only exchange the correct variable. Finally, the new offsprings perform the mutation process in order to create diverse individuals.

6) Selection, Crossover, Mutation and Replacement Operations: The GA process for evolutionary computation includes selection, crossover, mutation, and replacement operations. The selection operation here selects two individuals randomly from the population. When the probability of crossover is sufficiently high, the selected two individuals are processed by the uniform crossover operation to generate two new offsprings. Fig. 3 shows the flowchart of the selection and crossover processes, and an example is shown in Fig. 4. When the probability of mutation is sufficiently high, one of the two new offsprings is randomly selected for one point mutation and a new diverse offspring is thus generated. The mutation process flowchart is

Authorized licensed use limited to: b-on: Instituto Politecnico de Tomar. Downloaded on April 19,2024 at 18:43:55 UTC from IEEE Xplore. Restrictions apply.

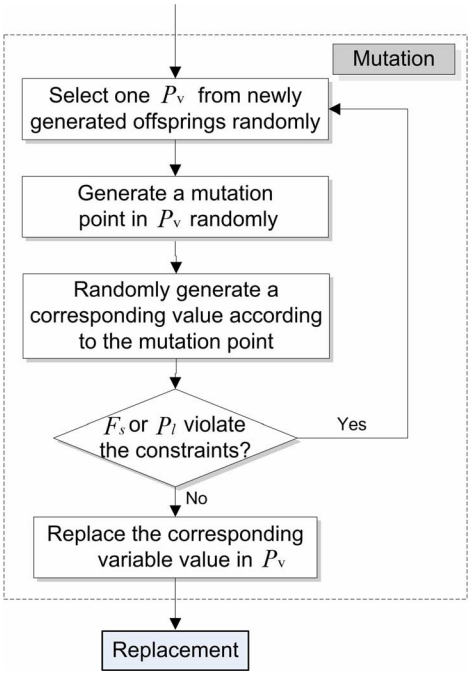

Fig. 5. Mutation flowchart for PCR-RFLP assay design. When the mutation operation is performed, one individual from the offsprings after the crossover is selected. One mutation point is then designated and the corresponding value generated. If the mutation point is F_s or P_l, the constraint condition must be examined as to whether it violates the primer regulations. If it conforms to the primer regulations, the corresponding variable value is replaced; else the above process is repeated until the value conforms to the constraint. Finally, the replacement operation is performed.

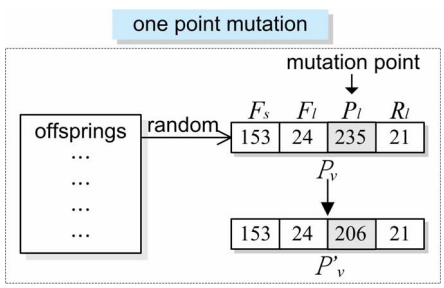

Fig. 6. An example of a mutation operation for PCR-RFLP assay design. Initially, a $P_v = (153, 24, 235, 21)$ is randomly selected from the newly generated offsprings after the crossover process. Then one mutation point of the position of F_s, F_l, P_l, or R_l is randomly generated. In the example, the position of the mutation point is P_l. The value of P_l is randomly generated between $P_{\min}$ and $P_{\max}$, and then replaces the corresponding variable value in P_v. The position of the mutation is F_s or P_l must be examined as to whether the random value violates the constraints or not. If no constraints are violated, the corresponding variable value in P_v is replaced, else the above processes are repeated until the constraints are satisfied. Finally, the new offspring $P'_v = (153, 24, 206, 21)$ is formed and then the replacement process is performed. The replacement process replaces the worst individuals by better offsprings generated during the above processes.

shown in Fig. 5, and an example of the mutation operation is given in Fig. 6. In Fig. 5, we examine only F_s and P_l for violation of the constraints. F_s and P_l are key to determine the position and length of PCR product and must therefore be examined to avoid exceeding the length of the template sequence. Since any the change in primer length within this range does not exceed the length of template sequence primer design is not influenced, and therefore we do not consider examining them.

III. Results

A. The Computational Environment

The proposed algorithm was run on an Intel(R) Core(TM) 2 CPU of 1.86 GHz and 1 GB RAM under the Microsoft Windows XP SP3 and JAVA 6.0 platforms.

B. Parameter Settings

Four main parameters are set for the proposed method, i.e., the number of iterations (generations), the population size, the crossover rate and the mutation rate. The parameter settings are based on the DeJong and Spears parameter settings [33]; the respective values are 1000, 50, 0.6 and 0.001. The PCR product lengths used to recognize the target SNPs are set to a 1 : 2 : 3 ratio, which ensures that the digested allelic fragments can be separated by the gel electrophoresis. Furthermore, different population sizes, crossover rates and mutation rates were used to evaluate the design of PCR-RFLP assay and we found the better parameter settings from the evaluation and used it to this problem.

C. SNP Data Set

Recently, a point mutation in the SLC6A4 gene was identified and shown to be associated with psychosis [34], bipolarity [35], and autism spectrum disorders [36]. The SLC6A4 gene is the member 4 of the solute carrier family 6 (neurotransmitter transporter, serotonin). All of the 288 SNPs of the SLC6A4 gene in NCBI dbSNP Build 130 were used in this study, excluding only the deletion/insertion polymorphisms (DIPs) and multi-nucleotide polymorphisms (MNPs). The DNA template sequences of the SLC6A4 SNPs were retrieved with a 500 bp flanking length (at both sides of the SNP) from SNP-Flankplus (http://bio.kuas.edu.tw/snp-flankplus/) [37]. The reference cluster IDs of these SNPs are shown in http://bio.kuas.edu.tw/ganpd/dataset.jsp.

D. In Silico Simulation for the Proposed Method

The common constraints used in the proposed method, i.e., primer length between 16 and 28 nt, GC% between 40 and 60%, primer $T\mathrm{m}$ between 45 and 62 °C, primer $T\mathrm{m}$ difference less than 5 °C, PCR product length larger than 100 bp, and easily separated PCR bands on gel electrophoresis with a ratio of 1 : 2 : 3 were used. All designed PCR-RFLP assay primer sets are provided in http://bio.kuas.edu.tw/ganpd/results.jsp, and the statistical results based on the common constraints are shown in supplemental file—Tables S1 to S5. For the SLC6A4 gene, our method found 251 PCR-RFLP SNPs with available restriction enzymes, and designed 502 primers (251 pairs of primer sets). The number of GC%, GC clamp, $T\mathrm{m}$, hairpin, and specificity is 502 (i.e., a primer with one GC%, one GC clamp, one $T\mathrm{m}$, one hairpin, and one specificity); the number of length difference and $T\mathrm{m}$ difference is 251 (i.e., a pair primer with one length difference and one $T\mathrm{m}$ difference); the number of PCR product length and dimer is 753 (i.e., a primer pair with three product lengths and three dimers which contains one cross-dimer and two self-dimers). Additional file describes the *in silico* results for the proposed method with different parameter settings and

Authorized licensed use limited to: b-on: Instituto Politecnico de Tomar. Downloaded on April 19,2024 at 18:43:55 UTC from IEEE Xplore. Restrictions apply.

TABLE I
THE CHARACTERISTICS DESCRIBED IN SNP CUTTER, PRIM-SNPING AND THE PROPOSED METHOD FOR THE SLC6A4 GENE. SNP CUTTER ONLY ACCEPTS INPUT IN SPECIFIED FORMATS OR DBSNP REFERENCE IDS. FOR THE SLC6A4 GENE, SNP CUTTER DESIGNED PCR-RFLP ASSAY PRIMERS FOR 55 SNPS WITH AVAILABLE RESTRICTION ENZYMES; 25 SNPS WITHOUT AVAILABLE RESTRICTION ENZYMES ARE PROVIDED, AND 196 SNPS WERE MISSING FROM THEIR DATABASES; 12 SNPS COULD NOT BE DESIGNED DUE TO UNKNOWN ERRORS. PRIM-SNPING IS CAPABLE OF PROVIDING SEQUENCE-BASED INPUT. FOR THE SLC6A4 GENE, PRIM-SNPING DESIGNED PCR-RFLP ASSAY PRIMERS FOR 129 SNPS WITH AVAILABLE RESTRICTION ENZYMES; 55 SNPS WITHOUT AVAILABLE RESTRICTION ENZYMES ARE PROVIDED, AND 104 SNPS WERE NO PRIMER WAS DESIGNED. THE PROPOSED METHOD DESIGNED 251 PCR-RFLP ASSAY PRIMERS WITH AVAILABLE RESTRICTION ENZYMES; ONLY 37 SNPS WITHOUT AVAILABLE RESTRICTION ENZYMES ARE PROVIDED

Characteristics	Program name		
	SNP Cutter	Prim-SNPing	The proposed method
SNP PCR-RFLP assay design with available restriction enzymes	55	129	251
SNP PCR-RFLP assay design without available restriction enzymes	25	55	37
Missing in dbSNP	196	0	0
No primers designed	0	104	0
Unknown errors	12	0	0

the better parameter settings. Moreover, the result using the better parameter settings are also shown.

E. Comparison of the Proposed Method With SNP Cutter and Prim-SNPing

SNP Cutter and Prim-SNPing were also used to design the PCR-RFLP assay primers for all SNPs in the SLC6A4 gene. SNP Cutter can design PCR-RFLP assay primers successfully for 55 SNPs with available restriction enzymes from the 288 SNPs. A further 25 SNPs were found without available restriction enzymes and the rest of the SNPs (208) were problematic (either missing in dbSNP or unknown errors). Prim-SNPing can design PCR-RFLP assay primers successfully for 129 SNPs with available restriction enzymes from the 288 SNPs. A further 55 SNPs were found without available restriction enzymes, and for the rest of the SNPs (104 SNPs) primers could not be designed. The proposed method, however, designed PCR-RFLP assay primers for 251 of 288 SNPs from this gene successfully (see the Table I).

F. Validation of PCR-RFLP Experiment

SNP-restriction fragment length polymorphism (SNP-RFLP) or PCR-RFLP markers can be used to rapidly and accurately differentiate *Staphylococcus aureus* strains based on their penicillin-binding protein gene (*pbp*). For validation the performance of the proposed method, the penicillin-binding protein genes of *Staphylococcus aureus* were selected as the test sequences. Following the nucleotide sequence alignment of the *pbp* genes, the different SNP genotype (A/G) is shown as follows:

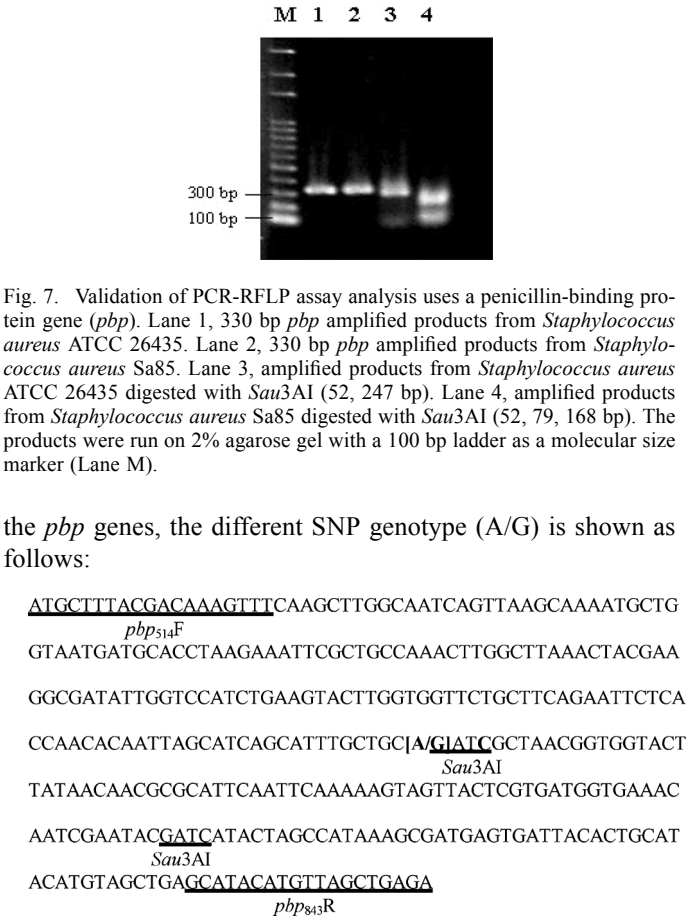

Fig. 7. Validation of PCR-RFLP assay analysis uses a penicillin-binding protein gene (*pbp*). Lane 1, 330 bp *pbp* amplified products from *Staphylococcus aureus* ATCC 26435. Lane 2, 330 bp *pbp* amplified products from *Staphylococcus aureus* Sa85. Lane 3, amplified products from *Staphylococcus aureus* ATCC 26435 digested with *Sau*3AI (52, 247 bp). Lane 4, amplified products from *Staphylococcus aureus* Sa85 digested with *Sau*3AI (52, 79, 168 bp). The products were run on 2% agarose gel with a 100 bp ladder as a molecular size marker (Lane M).

where [A] is the SNP found in the standard strain of *Staphylococcus aureus* (ATCC 26435) and [G] is the SNP found in the clinical strain of *Staphylococcus aureus* (Sa85).

PCR primers were designed for the SNP of the *pbp* gene by the user-friendly interface (http://bio.kuas.edu.tw/ganpd/availability.jsp) we implemented. The primer set consisting of pbp_{514} F (5'-ATGCTTTACGACAAAGTTTT-3') and pbp_{843} R (5'-TCTCAGCTAACATGTATGC-3') was used for the *pbp* gene amplification. The amplification conditions were as follows: 25 ng DNA samples were added to the PCR reaction mixture (25 μL) containing 2.5 μL 10X reaction buffer (with 20 mM Mg^{2+}), 0.25 μL of 5U/μL *taq* DNA polymerase, 1.0 μL 10 mM dNTPs Mix, 10 pmol of each pbp_{514} primer, 2.0 μL sample DNA, 17.25 μL ddH_2O. Initial denaturalization at 94 °C for 10 min, followed by 35 cycles of 94 °C for 1 min, 55 °C for 1 min, and 72 °C for 1 min and a final elongation step of 72 °C for 10 min was performed. The amplified PCR product of 330 bp was submitted for digestion with the restriction enzyme *Sau*3AI selected by the proposed method. *Sau*3AI cleaves the DNA at a GATC sequence. The DNA fragments obtained were resolved by electrophoresis on 2% agarose gel stained by ethidium bromide. The results are shown in Fig. 7.

As Fig. 7 shows, the PCR amplification yielded a unique PCR product size that is not capable of discriminating both strains of *Staphylococcus aureus* (Lane 1 and 2). After PCR, the amplicons were subjected to restriction digestion for PCR-RFLP analysis. The PCR-RFLP results from the standard strain of *Staphylococcus aureus* (ATCC 26435) showed two distinguishable DNA fragments of 73 and 257 bp in length. On the

Authorized licensed use limited to: b-on: Instituto Politecnico de Tomar. Downloaded on April 19,2024 at 18:43:55 UTC from IEEE Xplore. Restrictions apply.

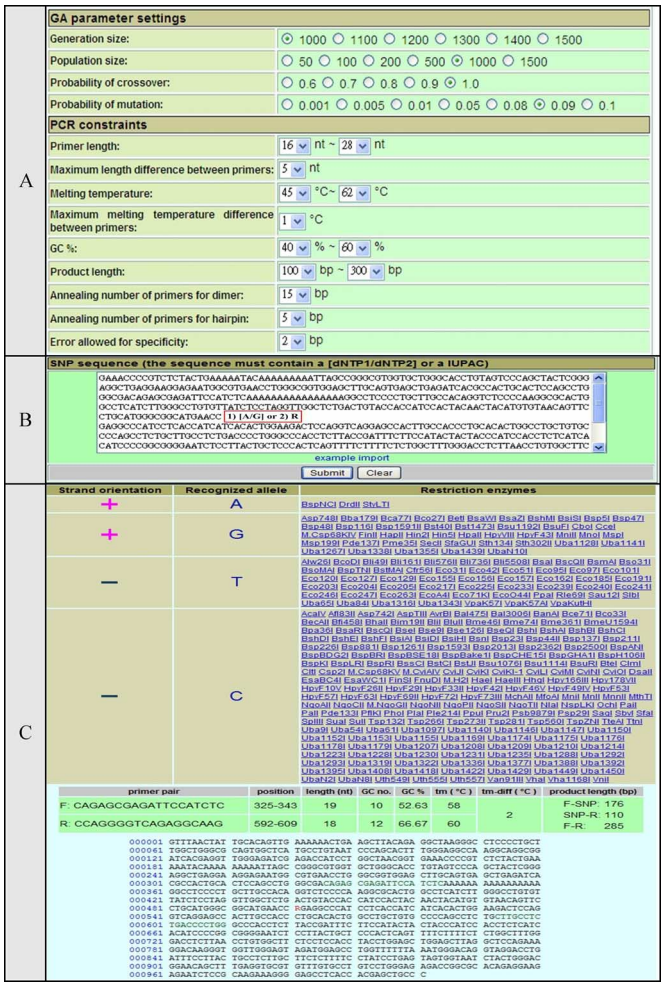

Fig. 8. The user-friendly interface of GANPD. A) GA fmeter settings and PCR constraints. B) SNP sequence input with [dNTP1/dNTP2] or IUPAC format. C) Output of GANPD including available restriction enzymes, a feasible PCR-RFLP assay primer set and its information, and sequence representation.

other hand, the clinical strain of *Staphylococcus aureus* (Sa85) showed three DNA fragments of 71, 79, and 180 bp in length. The PCR-RFLP analysis demonstrated that both tested strains showed the expected PCR-RFLP patterns. It is suggested that all of the results obtained are thus correct. Overall, the PCR-RFLP assay analysis proved that the proposed method is capable of providing highly reliable SNP genotyping, and that it is a convenient and efficient tool for designing PCR experiments.

G. The Available User-Friendly Interface

GANPD, a web-based PCR-RFLP assay designer, provides the feasible PCR-RFLP primers and available restriction enzymes based on the proposed method. GANPD provides users a user-friendly interface to set related parameters including GA parameters and PCR constraints [Fig. 8(A)] and a brief input box to paste template sequence with a specific SNP site. Two SNP types are available: 1) [dNTP1/dNTP2] and 2) IUPAC format (e.g., [A/G] or R) [Fig. 8(B)]. After performing PCR-RFLP assay analysis by the proposed method, all available restriction enzymes from REBASE, primer information and a feasible PCR-RFLP primer set are provided for PCR-RFLP experiments [Fig. 8(C)].

IV. Discussion

To date, only a few systems, such as V-MitoSNP, SNP Cutter and Prim-SNPing, provide the PCR-RFLP assay design function to genotype SNPs. However, these systems lack an appropriate method for PCR-RFLP assay design. In this paper, a PCR-RFLP assay design method is proposed to facilitate PCR-RFLP experiments for SNP genotyping. In order to assess the proposed method, we used 288 SNPs of the SLC6A4 gene to simulate different parameter settings for the population size, the crossover rate, and the mutation rate, and finally got the better parameter settings for PCR-RFLP assay design. Furthermore, wet laboratory validation was performed and a user-friendly interface was provided.

A. Requirements for PCR-RFLP Primer Design

In PCR-RFLP for SNP genotyping, the information about the availability of restriction enzymes is paramount. If no restriction enzymes are available for a target SNP, the PCR-RFLP experiment is meaningless. In our method, the updated core of the SNP-RFLPing program [29] is used to mine for restriction enzymes in order to ensure that PCR-RFLP assay can be performed. When designing PCR-RFLP assay for the target SNP, a feasible SNP template sequence is required. Hence, the SNP-Flankplus program [37] is introduced to obtain a flanking sequence of 500 bps in length for a target SNP. The typical primer design constraints primer length, primer length difference, Tm, Tm difference, GC proportion, GC clamp, PCR product length are used in this study. Furthermore, secondary structures, such as dimer formation of a primer pair (including cross-dimers and self-dimers), hairpin formation of a primer, and the specificity of primers in a template sequence are also applied to design PCR-RFLP primers.

B. Influence of the Parameter Settings

1) Population Size: The GA is used to find PCR-RFLP primers that correspond to the primer design constraints. The *in silico* simulation of the proposed method showed that it reliably fits the primer design constraints. Dejong and Spears parameter settings are applied in the simulation of the proposed method (Table S1 in supplemental file). However, the population size of 50 used by Dejong and Spears is too small to provide the necessary sampling accuracy for the design of PCR-RFLP assay primer sets. Consequently, we gradually increased the population sizes from 50 to 1000 to ensure accurate sampling (Table S2 in supplemental file). With gradually increased population sizes, the average fitness values gradually decrease, i.e., the solutions got increasingly better. This means that a larger population size is better suited for PCR-RFLP assay design. The standard deviation shows that a change in the population size greatly influences the GC% value $(\mathrm{SD} = 20.44)$ and the GC clamp value $(\mathrm{SD} = 32.35)$. The change in population size only slightly influences the primer length difference $(\mathrm{SD} = 0.70)$, the $Tm(\mathrm{SD} = 0.84)$, the Tm difference (SD=1.34) and the average fitness value $(\mathrm{SD} = 0.73)$; the change has no influence on the product length, the dimer, the hairpin and the specificity value.

2) Crossover Rate and Mutation Rate: In order to observe the influence of the crossover rate and the mutation rate, we

Authorized licensed use limited to: b-on: Instituto Politecnico de Tomar. Downloaded on April 19,2024 at 18:43:55 UTC from IEEE Xplore. Restrictions apply.

changed these parameters to simulate PCR-RFLP assay design (Tables S3 and S4 in supplemental file). If a crossover rate is set to be higher than the crossover rate of Dejong and Spears used, the average fitness value decreases. In other words, a crossover rate larger than the crossover rate of Dejong and Spears is better suited to design PCR-RFLP primers. The standard deviations shows that the change in the crossover rate influences the GC% $(\mathrm{SD} = 10.23)$ and the GC clamp $(\mathrm{SD} = 22.9)$ values; it slightly affects the primer length difference $(\mathrm{SD} = 2.10)$, the $T\mathrm{m}(\mathrm{SD} = 4.24)$, the $T\mathrm{m}$ difference $(\mathrm{SD} = 5.70)$, the product length $(\mathrm{SD} = 0.57)$, the hairpin $(\mathrm{SD} = 1.50)$, and the specificity $(\mathrm{SD} = 0.32)$ values, and has no influence on the formation of dimers. The mutation rate does not seem to have any regular influence on the average fitness values. The standard deviation also shows that the changed mutation rate slightly influences all common constraints except for the dimer and the specificity constraints. When the population size was adjusted to 1000 and the crossover rate was adjusted to 1.0, the influence of the mutation rate is lower. Only the standard deviation of GC% $(\mathrm{SD} = 3.97)$ and GC clamp $(\mathrm{SD} = 1.48)$ were greater than 1. The results indicate that an increase in the population size and the crossover rate improves PCR-RFLP assay design.

C. Review of the Related PCR-RFLP Assay Design

V-MitoSNP and Prim-SNPing are two web-based programs we developed to assist in PCR-RFLP assay design. However, V-MitoSNP only focuses on mitochondrial SNPs and does not provide a sequence or SNP ID input function, thus limiting its application. V-MitoSNP uses simple constraints and generally identifies primers which lack stringent quality criteria. Prim-SNPing provides a user-friendly interface. However, the window-sliding strategy used makes its search always inefficiently. Furthermore, it uses an exact method to design primers that must conform to all primer constraints. This often results in none of output. SNP Cutter is another tool used for PCR-RFLP assay design. It uses Primer3 to provide suitable primers and identifies available restriction enzymes. However, it is limited by the specific input formats required and many SNP IDs are not found in its database. In addition, the latest restriction enzymes are not provided and the Primer3 program is not updated, thus limiting its applicability for PCR-RFLP assay design. The proposed method uses a GA, an evolutionary computation concept, to design PCR-RFLP primers and a useful restriction enzyme mining program, SNP-RFLPing, is integrated to it. The proposed method screens the possible solutions based on the validated approaches, so that feasible primers and available restriction enzymes can always be obtained.

V. Conclusion

PCR-RFLP is a useful laboratory technique for in small-scale basic research studies of complex genetic diseases that are associated with SNP. This study presents a reliable method for designing feasible PCR-RFLP assay. The primer constraints and secondary structures were carefully considered. The updated core of SNP-RFLPing is integrated into the proposed method to mine available restriction enzymes for the target SNP. The flexibility of the proposed method has been demonstrated for many polymorphisms by *in silico* simulated experiments in 288 SNPs of the SLC6A4 gene and given a wet laboratory validation. We ensure the method is capable of finding available restriction enzymes and feasible primers for SNP genotyping. A convenient and user-friendly interface is provided to assist researchers in PCR-RFLP assay design. In conclusion, the method is a useful tool for PCR-RFLP SNP genotyping.

References

[1] J. G. Hacia, J. B. Fan, O. Ryder, L. Jin, K. Edgemon, G. Ghandour, R. A. Mayer, B. Sun, L. Hsie, C. M. Robbins, L. C. Brody, D. Wang, E. S. Lander, R. Lipshutz, S. P. Fodor, and F. S. Collins, "Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays," *Nat. Genet.*, vol. 22, pp. 164–7, Jun. 1999.

[2] V. Mooser, D. M. Waterworth, T. Isenhour, and L. Middleton, "Cardiovascular pharmacogenetics in the SNP era," *J. Thromb. Haemost.*, vol. 1, pp. 1398–402, Jul. 2003.

[3] H. C. Erichsen and S. J. Chanock, "SNPs in cancer research and treatment," *Brit. J. Cancer*, vol. 90, pp. 747–751, 2004.

[4] L. J. Engle, C. L. Simpson, and J. E. Landers, "Using high-throughput SNP technologies to study cancer," *Oncogene*, vol. 25, pp. 1594–1601, 2006.

[5] C. R. Cantor, "The use of genetic SNPs as new diagnostic markers in preventive medicine," *Ann. N.Y. Acad. Sci.*, vol. 1055, pp. 48–57, 2005.

[6] T. Bernig and S. J. Chanock, "Challenges of SNP genotyping and genetic variation: Its future role in diagnosis and treatment of cancer," *Expert Rev. Mol. Diagnostics*, vol. 6, pp. 319–331, 2006.

[7] W. Sadee, "Pharmacogenomics: Harbinger for the Era of personalized medicine?," *Mol. Interventions*, vol. 5, pp. 140–143, 2005.

[8] A. M. Divne and M. Allen, "A DNA microarray system for forensic SNP analysis," *Forensic Sci. Int.*, vol. 154, pp. 111–121, 2005.

[9] M. Ota, H. Fukushima, J. K. Kulski, and H. Inoko, "Single nucleotide polymorphism detection by polymerase chain reaction-restriction fragment length polymorphism," *Nat. Protoc.*, vol. 2, pp. 2857–2864, 2007.

[10] R. Zhang, Z. Zhu, H. Zhu, T. Nguyen, F. Yao, K. Xia, D. Liang, and C. Liu, "SNP Cutter: A comprehensive tool for SNP PCR-RFLP assay design," *Nucleic Acids Res.*, pp. W489–W492, Jul. 1, 2005.

[11] R. J. Roberts, T. Vincze, J. Posfai, and D. Macelis, "REBASE—A database for DNA restriction and modification: Enzymes, genes and genomes," *Nucleic Acids Res.*, vol. 38, pp. D234–D236, Jan. 2010.

[12] L. Y. Chuang, C. H. Yang, Y. H. Cheng, D. L. Gu, P. L. Chang, K. H. Tsui, and H. W. Chang, "V-MitoSNP: Visualization of human mitochondrial SNPs," *BMC Bioinformatics*, vol. 7, pp. 379–, 2006.

[13] S. Rozen and H. Skaletsky, "Primer3 on the WWW for general users and for biologist programmers," *Methods Mol. Biol.*, vol. 132, pp. 365–386, 2000.

[14] H. W. Chang, L. Y. Chuang, Y. H. Cheng, Y. C. Hung, C. H. Wen, D. L. Gu, and C. H. Yang, "Prim-SNPing: A primer designer for cost-effective SNP genotyping," *Biotechniques*, vol. 46, pp. 421–431, May 2009.

[15] M. J. McPherson, G. R. Taylor, and P. Quirke, *PCR, a Practical Approach*. New York: Oxford Univ. Press, 1991.

[16] J. Sambrook and D. W. Russell, *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press, 2001, vol. 2, pp. 8.1–8.126.

[17] J. S. Wu, C. Lee, C. C. Wu, and Y. L. Shiue, "Primer design using genetic algorithm," *Bioinformatics*, vol. 20, pp. 1710–1717, Jul. 22, 2004.

[18] T. Kampke, M. Kieninger, and M. Mecklenburg, "Efficient primer design algorithms," *Bioinformatics*, vol. 17, pp. 214–225, Mar. 2001.

[19] C. H. Yang, Y. H. Cheng, L. Y. Chuang, and H. W. Chang, "Specific PCR product primer design using memetic algorithm," *Biotechnol. Prog.*, vol. 25, pp. 745–753, May–Jun. 2009.

[20] L. Y. Chuang, Y. H. Cheng, and C. H. Yang, "URPD: A specific product primer design tool," *BMC Res. Notes*, vol. 5, pp. 306–, Jun. 19, 2012.

[21] C. H. Yang, Y. H. Cheng, H. W. Chang, and L. Y. Chuang, "Primer design with specific PCR product using particle swarm optimization," *Int. J. Chem. Biomol. Eng.*, vol. 3, pp. 18–23, 2010.

[22] J. Wu, J. Wang, and J. Chen, "A practical algorithm for multiplex PCR primer set selection," *Int. J. Bioinform. Res. Appl.*, vol. 5, pp. 38–49, 2009.

[23] J. Wang, K. B. Li, and W. K. Sung, "G-PRIMER: Greedy algorithm for selecting minimal primer set," *Bioinformatics*, vol. 20, pp. 2473–2475, Oct. 12, 2004.

Authorized licensed use limited to: b-on: Instituto Politecnico de Tomar. Downloaded on April 19,2024 at 18:43:55 UTC from IEEE Xplore. Restrictions apply.

[24] Y. F. Chen, R. C. Chen, Y. K. Chan, R. H. Pan, Y. C. Hseu, and E. Lin, "Design of multiplex PCR primers using heuristic algorithm for sequential deletion applications," *Comput. Biol. Chem.*, vol. 33, pp. 181–188, Apr. 2009.
[25] C. H. Yang, Y. H. Cheng, C. H. Yang, and L. Y. Chuang, "Mutagenic primer design for mismatch PCR-RFLP SNP genotyping using a genetic algorithm," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 9, pp. 837–845, May–Jun. 2012.
[26] C. H. Yang, Y. H. Cheng, and L. Y. Chuang, "A natural PCR-RFLP primer design for SNP genotyping using a genetic algorithm," in *Proc. Int. MultiConf. Eng. Comput. Scientists 2010*, Hong Kong, pp. 124–129.
[27] H. W. Chang, C. H. Yang, P. L. Chang, Y. H. Cheng, and L. Y. Chuang, "SNP-RFLPing: Restriction enzyme mining for SNPs in genomes," *BMC Genom.*, vol. 7, pp. 30–, 2006.
[28] L. Y. Chuang, C. H. Yang, K. H. Tsui, Y. H. Cheng, P. L. Chang, C. H. Wen, and H. W. Chang, "Restriction enzyme mining for SNPs in genomes," *Anticancer Res.*, vol. 28, pp. 2001–2007, Jul.–Aug. 2008.
[29] H. W. Chang, Y. H. Cheng, L. Y. Chuang, and C. H. Yang, "SNP-RFLPing 2: An updated and integrated PCR-RFLP tool for SNP genotyping," *BMC Bioinform.*, vol. 11, pp. 173–, 2010.
[30] S. T. Sherry, M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, and K. Sirotkin, "dbSNP: The NCBI database of genetic variation," *Nucleic Acids Res.*, vol. 29, pp. 308–311, Jan. 1, 2001.
[31] R. B. Wallace, J. Shaffer, R. F. Murphy, J. Bonner, T. Hirose, and K. Itakura, "Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: The effect of single base pair mismatch," *Nucleic Acids. Res.*, vol. 6, pp. 3543–3557, Aug. 10, 1979.
[32] V. R. Prezioso*, General Notes on Primer Design in PCR**. Westbury, NY, USA: BioSystems Laboratory, Brinkmann™ Instruments Inc., 2013.
[33] K. A. De Jong and W. M. Spears*, An Analysis of the Interacting Roles of Population Size and Crossover in Genetic Algorithms*. Berlin/Heidelberg, Germany: Springer, 1990, vol. 496.
[34] T. E. Goldberg, R. Kotov, A. T. Lee, P. K. Gregersen, T. Lencz, E. Bromet, and A. K. Malhotra, "The serotonin transporter gene and disease modification in psychosis: Evidence for systematic differences in allelic directionality at the 5-HTTLPR locus," *Schizophr. Res.*, vol. 111, pp. 103–108, Jun. 2009.
[35] L. Mandelli, M. Mazza, G. Martinotti, M. Di Nicola, D. Tavian, E. Colombo, S. Missaglia, D. De Ronchi, R. Colombo, and L. Janiri, "Harm avoidance moderates the influence of serotonin transporter gene variants on treatment outcome in bipolar patients," *J. Affect. Disord.*, vol. 119, pp. 205–209, 2009.
[36] T. Sakurai, J. Reichert, E. J. Hoffman, G. Cai, H. B. Jones, M. Faham, and J. D. Buxbaum, "A large-scale screen for coding variants in SERT/SLC6A4 in autism spectrum disorders," *Autism Res.*, vol. 1, pp. 251–257, Aug. 2008.
[37] C. H. Yang, Y. H. Cheng, L. Y. Chuang, and H. W. Chang, "SNP-Flankplus: SNP ID-centric retrieval for SNP flanking sequences," *Bioinformation*, vol. 3, pp. 147–149, 2008.

Li-Yeh Chuang received her M.S. degree from Department of Chemistry at University of North Carolina, Greensboro, NC, USA, in 1989 and Ph.D. degree from Department of Biochemistry at North Dakota State University, Fargo, ND, USA, in 1994. She is a professor of the Department of Chemical Engineering & Institute of Biotechnology and Chemical Engineering at I-Shou University, Kaohsiung, Taiwan. Her main areas of research are bioinformatics, biochemistry, and genetic engineering.

Yu-Huei Cheng (M'12) received the M.S. degree and Ph.D. degree from the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Taiwan, in 2006 and 2010, respectively. He is an assistant professor of the Department of Digital Content Design and Management, Toko University, Chiayi, Taiwan. He has rich experiences in computer programming, database design and management, and systems programming and design. His research interests include bioinformatics, biomedical engineering, computational biology, computational intelligence, electronic circuit, evolutionary computation, fuzzy systems, information retrieval, and solar energy.

Cheng-Huei Yang received the B.S. degree from National Taipei Institute of Technology, Taiwan, in 1978 and the M.S. degree from Northeastern University, Boston, MA, USA, in 1987. He received his Ph.D. degree in electrical engineering from National Cheng Kung University, Tainan, Taiwan, in 2001. He is a professor of Electronic Communication Engineering, National Kaohsiung Marine University, Kaohsiung, Taiwan. His current research interests are network communication, electronic instrument systems, and image processing.

Cheng-Hong Yang (SM'03) received his M.S. and Ph.D. degrees in computer engineering from North Dakota State University, Fargo, ND, USA, in 1990 and 1992, respectively. He is a professor and serves as president of National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan. He is the IEEE senior member and his main areas of research are evolutionary computation, bioinformatics, and assistive tool implementation.

Authorized licensed use limited to: b-on: Instituto Politecnico de Tomar. Downloaded on April 19,2024 at 18:43:55 UTC from IEEE Xplore. Restrictions apply.