A Natural PCR-RFLP Primer Design for SNP Genotyping Using a Genetic Algorithm

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

***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 polymorphism (SNP). Before performing PCR-RFLP for SNP genotyping, a feasible primer pair, which must observe numerous constraints, and an available restriction enzyme for discriminating a target SNP, are required. Here, we propose a method which uses a genetic algorithm (GA) to search for optimal natural PCR-RFLP primers and employs the core of SNP-RFLPing to reliably mine available restriction enzymes. The *in silico* simulation of the proposed method in the SNPs of the SLC6A4 gene showed that it is able to stably to design natural PCR-RFLP primers which most fit the common primer constraints and provide available restriction enzymes.**

***Index Terms* — PCR-RFLP, SNP genotyping, primer, restriction enzyme, GA.**

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]. And it is a relatively simple, inexpensive and accurate method for SNP genotyping [9, 10]. When genotyping a SNP uses PCR, the target SNP must be discriminated by digestion with specific restriction enzymes in a process called "Natural PCR-RFLP". Before performing natural PCR-RFLP for SNP genotyping, a feasible primer pair, which must observe several constraints, and an available restriction enzyme for discriminating a target SNP, are required.

Manuscript received January 12, 2010. This work was supported in part by the National Science Council in Taiwan under grants NSC96-2221-E-214-050-MY3, NSC97-2622-E-151-008-CC2, NSC98-2221-E-151-040-, NSC98-2622-E-151-001-CC2, and NSC98-2622-E-151-024-CC3.

Cheng-Hong Yang is with the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan and Department of Network Systems, Toko University, Chiayi, Taiwan (e-mail: chyang@cc.kuas.edu.tw).

Yu-Huei Cheng is with the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan (e-mail: yuhuei.cheng@gmail.com).

Li-Yeh Chuang is with the Department of Chemical Engineering, I-Shou University, Kaohsiung, Taiwan (email: chuang@isu.edu.tw).

There were a few systems developed that provide natural PCR-RFLP primer design. For example, V-MitoSNP identifies the available restriction enzymes from REBASE [11] and designs a PCR-RFLP primer set for RFLPs in all mtSNPs [12]. However, this system uses very simple constraints and generally identifies primers lacking in stringent quality; it can only design primers for mitochondrial SNPs. SNP Cutter uses a pre-selected or customizable list of restriction enzymes and uses Primer3, the most popular non-commercial primer design software [13-15], to look for PCR-RFLP primer sets. Even though it is a helpful tool for PCR-RFLP, it is inconvenient for input which needs a specified format and does not provide optimal PCR-RFLP primers. Prim-SNPing is an improved software tool with a natural PCR-RFLP primer design function for cost-effective SNP genotyping [16]. However, the incorporated window-sliding strategy limits its search efficiency and the quality. Therefore, the development of an improved method for natural PCR-RFLP primer design is still mandated.

Natural PCR-RFLP primer design is a challenging task since numerous primer constraints must be conformed [13, 17-19]. These include the primer length, length difference of primer pair, PCR product length, GC proportion, melting temperature (T_m), GC clamp, dimer (including cross-dimers and self-dimers), hairpin structure, and specificity. Furthermore, it is pivotal that a target SNP must be discriminated by digestion with available restriction enzymes at the same time. Hence, a genetic algorithm (GA) [20, 21] is introduced by us to facilitate the search for natural PCR-RFLP primer sets with available restriction enzymes which provided by the updated core of SNP-RFLPing [22, 23]. GA is a stochastic search algorithm modeled on the process of natural selection underlying biological evolution [24]. It constitutes a randomized search and optimization technique that derives its working principle from natural genetics. In a GA, the evolutionary computations involved – selection, crossover, mutation and replacement – are effective in determining optimal solutions for many natural primer sets. After each run, individuals in a GA share information with each other and superior solutions based on a fitness rule are refined from generation to generation. SNP-RFLPing [22, 23] is a time-saving application for mining the restriction enzymes for RFLP assays. Consequently, an optimal natural primer pair with available restriction enzymes can be found.

ISBN: 978-988-17012-8-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

II. Materials And Methods

A. Problem definition

Let T_D be a template DNA sequence that is made up of base-nucleic acid codes of the DNA with an identified SNP. T_D is defined as follows:

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

where B_i is a base-nucleic acid that can be the regular nucleotide codes (A, T, C, or G) mixed with a SNP represented by IUPAC code (M, R, W, S, Y, K, V, H, D, B or N) or dNTPs format ([dNTP1/dNTP2]). The $\exists!$ symbol represents the existence and uniqueness. For the target SNP, we focused only on true SNPs as described in dbSNP [25] of NCBI, i.e., deletion/insertion polymorphisms (DIPs or Indel) and multi-nucleotide polymorphisms (MNP) are not included.

The natural PCR-RFLP primer design problem consists of finding a pair of sub-sequences of corresponding constraints from T_D that contains at least one restriction enzyme to distinguish 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:

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

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

where F_s and F_e denote the start index and the end index, respectively, of P_f in T_D. R_s and R_e denote the start index and the end index, respectively, of P_r in T_D. P_f and P_r are called a primer pair. $\{\overline{B_i}\}$ is the anti-sense sequence opposite to the sense sequence $\{B\}$. For a sense sequence $\{B\}$ = "GTCTACGTCGAAC", for example, the complement sequence of $\{B\}$ is "CAGATGCAGCTTG", since the complement base of 'A' is 'T' and the complement base of 'C' is 'G'. The anti-sense sequence $\{\overline{B_i}\}$ is the reverse sequence of the complement sequence of $\{B\}$, i.e., $\{\overline{B_i}\}$ = "GTTCGACGTAGAC".

In Fig. 1, the length of the template DNA is L_{T_D}, the minimum PCR product length is P_{min}, the maximum PCR product length is P_{max}, the start position of the forward primer is F_s, the length of the forward primer is F_l, the PCR product length between the start position of the forward primer and the end position of reverse primer is P_l, the PCR product length between the start position of the forward primer and SNP site is P_{l1}, the PCR product length between the SNP site and the end position of the reverse primer is P_{l2}, the length of the reverse primer is R_l, the random range of F_s is F_{s_range}, and the length from F_s to the template DNA end is P_{range}. Now, a simple vector (called "individual" in a GA) consisting of F_s, F_l, P_l and R_l can determine a primer pair. We define this vector as:

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

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

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

Consequently, the forward primer and the reverse primer can be obtained from P_v. P_v is the individual encoding prototype of GA in the natural PCR-RFLP primer design problem, and later sections will use P_v to perform evolutionary computations with the GA.

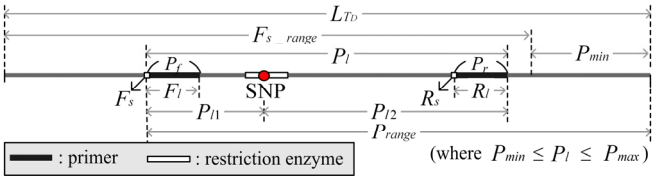

Fig. 1. Parameters for the natural PCR-RFLP primer design problem.

B. Functions for primer constraints

Primer constraints are important for performing a meaningful PCR experiment. The following eight functions were used to account for all possible primers: $Len_{diff}(P_v)$, $Tm(P_v)$, $Tm_{diff}(P_v)$, $GC_{proportion}(P_v)$, $GC_{clamp}(P_v)$, $dimer(P_v)$, $hairpin(P_v)$, $specificity(P_v)$ and $product(P_v)$. These functions are described below.

(1) $Len_{diff}(P_v)$

In the PCR experiment, a length difference of no more than 3 bps between the forward primer and the reverse primer is considered optimal. $Len_{diff}(P_v)$ is used to check whether the length difference of a primer pair exceeds 3 bps.

(2) $Tm(P_v)$ and $Tm_{diff}(P_v)$

The melting temperature $Tm(P)$ used here is calculated by the Wallace formula [26] (Eq. 6). The $Tm(P_v)$ function is used to check whether the melting temperature T_m of a primer pair lies between 50°C and 62°C, and $Tm_{diff}(P_v)$ checks whether the difference of the melting temperature exceeds 5°C.

$$Tm(P) = (\#G+\#C)*4 + (\#A+\#T)*2 \quad (6)$$

(3) $GC_{proportion}(P_v)$

The GC proportion in a primer is denoted the ratio of the nucleotides 'G' and 'C' that appear in a primer. An appropriate GC ratio in a primer lies in the range of 40-60%. The $GC_{proportion}(P_v)$ function is used to check whether the $GC\%(P)$ of the forward and reverse primer is between 40% and 60% or not.

(4) $GC_{clamp}(P_v)$

The function $GC_{clamp}(P_v)$ is used to check whether the 3' terminal end of a primer is base-nucleic acid 'G' or 'C' or not.

ISBN: 978-988-17012-8-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

(5) *dimer*(P_v)

If annealing between two primers takes place, a dimer is formed. The occurrence of dimers during a PCR experiment is problematic. 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 themselves.

(6) *hairpin*(P_v)

A primer may anneal to itself to form 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.

(7) *specificity*(P_v)

The PCR experiment might fail if the primer is not site-specific and appears more than once 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.

(8) *product*(P_v)

In order to ensure that the digested allelic fragments can be easily distinguished by gel electrophoresis, we uses 1:2:3 ratios as a standard to estimate the PCR-RFLP product sizes with a tolerance range of 25 bps and a minimum product size of more than 100 bps.

C. The proposed method

The proposed method contains six separated processes which are (1) SNP-RFLP restriction enzymes mining, (2) judgment on availability of restriction enzymes, (3) creation of a random initial population, (4) fitness evaluation, (5) termination criteria judgment, and (6) selection, crossover, mutation and replacement operations, respectively, are described below. A flowchart of the proposed method is shown in Fig. 2. At first, the restriction enzymes for the target SNP are mined. Then a judgment for whether the restriction enzymes distinguish 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 by a fitness function (described below). And then, selection, crossover and mutation operations are performed, and the worst individuals are replaced by the better individuals based on evolutionary computation. The procedure is repeated in the next iteration until the termination conditions are reached.

(1) SNP-RFLP restriction enzymes mining

The proposed method uses the updated core of SNP-RFLPing [22, 23] which we developed to obtain available restriction enzymes from REBASE [11].

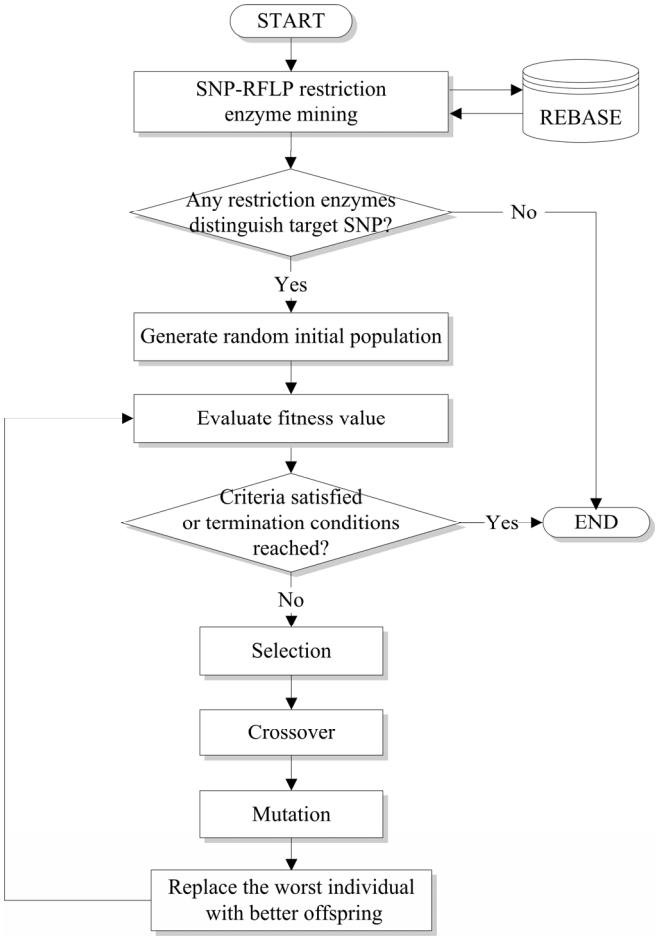

Fig. 2. Flowchart of the GA-based natural PCR-RFLP primer design.

(2) Judgment on availability of restriction enzymes

If no restriction enzyme is available to distinguish the alleles of the target SNP, the natural PCR-RFLP primer design will be insignificant for SNP genotyping. Therefore, the algorithm will be finished when this circumstance occurring. Otherwise, the algorithm keeps on running.

(3) Creation of a random initial population

Individuals $P_v = (F_s, F_l, P_l, R_l)$ are randomly generated as an initial population without duplicates. F_s is randomly generated between 1 and ($L_{T_D} - P_{min} + 1$). F_l is randomly generated between the minimum and the maximum length of a primer. In the present method, the minimum length of a primer was set to 16 bps and the maximum length of a primer was set to 28 bps. In order to limit the PCR product length, the proposed method randomly generates P_l between P_{min} and P_{max}. R_l is randomly generated in the same way as F_l.

(4) Fitness evaluation

The fitness value of every individual is evaluated in turn by a fitness function. In order to let a primer pair satisfies the design constraints, the aforementioned primers constraints are used to evaluate the fitness value and the fitness value is minimized (i.e., a fitness value equal to 0 indicates the best natural primer pair). The fitness function:

ISBN: 978-988-17012-8-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$Fitness(P_v) = 3*(Len_{diff}(P_v) + GC_{proportion}(P_v) + GC_{clamp}(P_v)) + 10*(Tm(P_v) + Tm_{diff}(P_v) + dimer(P_v) + hairpin(P_v)) + 50*specificity(P_v) + 60*product_{ratio}(P_v) \quad (7)$$

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 the requirements of the PCR experiment and can be adjusted by biologists and researchers based on their own experimental requirements. Larger weight values represent a higher importance of a function.

(5) Termination criteria judgment

Two termination criteria are used in the proposed method: an individual fitness value of 0, and reaching a preset number of iterations (generations). The number of iterations set in our experiments is discussed in the following "Parameter settings" section.

(6) Selection, crossover, mutation and replacement operations

The GA process for evolutionary computation includes selection, crossover, mutation and replacement operations. The selection operation here uses tournament selection to select the two best individuals from a population. When the probability of crossover is sufficiently high, the selected two individuals are processed by the uniform crossover operation and two new offsprings are generated. Following the selection operation, one-point mutation is applied to the proposed GA. If the probability of mutation is sufficiently high, one offspring after crossover is randomly selected for mutation. Finally, the worst two individuals are replaced by the new individuals through the replacement operation.

III. Results

A. The environment

The proposed method was run using an Intel(R) Core(TM) 2 CPU of 1.86 GHz and 1GB RAM under the Microsoft Windows XP SP3 and JAVA 5.0 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 [27]; the respective values are 1000, 50, 0.6 and 0.001. Furthermore, the PCR product lengths were set to 1:2:3 ratios and the minimum one more than 100 bps to ensure that the digested allelic fragments can be separated by gel electrophoresis. These ratios were chosen based on previous studies conducted by us. Different population sizes were used to simulate the design of natural PCR-RFLP primers.

C. SNP data set

Recently, a point mutation in the SLC6A4 gene was identified and shown to be associated with psychosis [28], bipolarity [29], and autism spectrum disorders [30]. The SLC6A4 gene is the member 4 for solute carrier family 6 (neurotransmitter transporter, serotonin). The 288 SNPs for the SLC6A4 gene in NCBI dbSNP Build 130 was used in this study, excluding the deletion/insertion polymorphisms (DIP) and multi-nucleotide polymorphisms (MNP). The PCR-RFLP SNPs were retrieved with a 500 bps flanking length (at both sides of the SNP) from SNP-Flankplus (http://bio.kuas.edu.tw/snp-flankplus/) [31].

D. In silico simulation of the proposed method

When we designed natural PCR-RFLP primer using the proposed method for 288 SNPs of the SLC6A4 gene, there were 251 SNPs which had an available restriction enzyme at least designed successfully. The 251 SNPs of the SLC6A4 gene were continued to perform natural PCR-RFLP primer design below.

(1) Based on DeJong and Spears' parameter settings

Dejong and Spears' parameter settings are the standard used for most GAs, and for this reason we used these parameter settings in the present study. Typically, crossover is applied at more than or equal to a rate of 0.6, and the mutation rate is equal to 0.001 [27]. Table 1 shows the results of the proposed method based on DeJong and Spears' parameter settings. Six designed primers violated the primer length difference criterion. Most of the primer length differences were between 0 and 4 bp (data not shown). Eighty-six primers had a GC% of less than 40%, and 73 primers had a ratio higher than 60%; all others were between 40% and 60% (data not shown). Approximately 65.9% of the primers (331/502) satisfied the GC clamp restriction. Ninety-seven percent of the designed primers (490/502) satisfied the T_m criterion, and 94.8% of the primer pairs (238/251) satisfied the T_m difference criterion. The PCR product length and the hairpin criterion were only violated by one primer, respectively. All designed primers were dimer-free and satisfied the specificity criterion.

Table 1. The statistical number of designed natural PCR-RFLP primers for satisfying the common constraints in 251 SNPs of the SLC6A4 gene based on DeJong and Spears' parameter settings using the proposed method. The numerator represents the number of designed natural PCR-RFLP primers that satisfy the common constraints and the denominator is the number of all designed natural PCR-RFLP primers, respectively. The average fitness value represents the estimation criterion for designing natural PCR-RFLP primers. A lower average fitness value is more ideal for designing natural PCR-RFLP primers.

	Constraints									
population size	primer length difference	GC%	GC clamp	T_m	T_m difference	product length	dimer	hairpin	specificity	average fitness
50	245/251	395/502	331/502	490/502	238/251	752/753	753/753	501/502	502/502	4.74

ISBN: 978-988-17012-8-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Table 2. The number of the designed natural PCR-RFLP primers that satisfy the common constraints in 251 SNPs of the SLC6A4 gene based on DeJong and Spears' parameter settings with a different population sizes scaled from 100-1000 bps. The highlighted row contains the best values for designing natural PCR-RFLP primers. It has the lowest average fitness value. The last row gives the mean and standard deviation.

	Constraints									
population size	primer length difference	GC%	GC clamp	T_m	T_m difference	product length	dimer	hairpin	specificity	average fitness
100	248/251	404/502	384/502	496/502	245/251	752/753	753/753	502/502	502/502	3.24
200	248/251	427/502	431/502	497/502	248/251	752/753	753/753	502/502	502/502	2.05
300	247/251	439/502	463/502	496/502	248/251	752/753	753/753	502/502	502/502	1.69
400	248/251	455/502	458/502	498/502	248/251	752/753	753/753	502/502	502/502	1.46
500	248/251	446/502	469/502	498/502	249/251	752/753	753/753	502/502	502/502	1.28
600	249/251	451/502	476/502	498/502	248/251	752/753	753/753	502/502	502/502	1.18
700	249/251	456/502	481/502	497/502	250/251	752/753	753/753	502/502	502/502	1.03
800	249/251	464/502	484/502	498/502	249/251	752/753	753/753	502/502	502/502	0.96
900	249/251	465/502	484/502	498/502	249/251	752/753	753/753	502/502	502/502	0.92
1000	249/251	463/502	490/502	498/502	249/251	752/753	753/753	502/502	502/502	0.82
mean±SD	248±0.70	447±20.44	462±32.35	497±0.84	248±1.34	752±0	753±0	502±0	502±0	1.76±0.73

(2) Based on different population sizes

This paragraph describes the increased population sizes based on the DeJong and Spears' parameter settings to design natural PCR-RFLP primers by the proposed method. The respective population sizes were 100, 200, 300, 400, 500, 600, 700, 800, 900 and 1000 (Table 2). In Table 2, the mean and standard deviations are provided for estimating the difference in different population sizes. An average fitness value of 0.82 was achieved for a population size of 1000; this was is the best value for the population sizes used. For a population size of 1000, only 2 designed primers violated the primer length difference. Most of the primer length differences were between 0 and 3 bps (data not shown). Twenty-one primers had a GC% of less than 40%, and 23 primers had a GC% of more than 60%. For GC clamp, 97.6% of the primers (490/502) satisfied the restriction. Ninety-nine point two percent of the designed primers and primer pairs satisfied the T_m and the T_m difference (498/502 and 249/251). For the PCR product length, only one primer was problematic. Dimers and hairpins were not presented in any primers; the specificity criterion was satisfied by all primers.

IV. Discussion

To date, a few systems, such as V-MitoSNP, SNP Cutter and Prim-SNPing, provide a function for natural PCR-RFLP primer design to genotype SNPs. However, these systems lack of an appropriate algorithm to design optimal natural PCR-RFLP primers. Many primer design approaches have been proposed so far, e.g., dynamic programming [32], genetic algorithm [19, 33], parthenogenetic algorithm MG-PGA [34], greedy algorithm [35], heuristic algorithm [36], and others. Nevertheless, most of these methods do not focus on the design of optimal natural PCR-RFLP primers and the availability of restriction enzymes for SNP genotyping. In this paper, a natural PCR-RFLP primer design method is proposed to facilitate PCR-RFLP. In order to assess the proposed method, we used 251 SNPs of the SLC6A4 gene to *in silicon* simulate the results for designing natural PCR-RFLP primers. The requirement and the influence of the population size in the proposed method for PCR-RFLP primer design are discussed below.

A. The requirement for the PCR-RFLP primer design

In PCR-RFLP for SNP genotyping, the restriction enzymes are the most important information. If restriction enzymes are not available, the PCR-RFLP experiment is meaningless. In our method, the updated core of the SNP-RFLPing program is used to mine for available restriction enzymes to ensure that natural PCR-RFLP primers are meaningful. Before designing PCR-RFLP primers, a feasible template sequence is required. Hence, the SNP-Flankplus program is introduced to obtain a 500 bps flanking sequence for a target SNP. In this study, the typical primer design constraints for primer length, length difference of a primer pair, T_m of a primer, T_m difference of a primer pair, GC proportion of a primer, GC clamp of a primer, PCR product length are used. Secondary structures, such as dimer formation of a primer pair (including cross-dimer and self-dimer), hairpin formation of a primer, and the specificity of a primer pair in a template sequence are also applied to the natural PCR-RFLP primer design.

B. The influence of the population size

The GA is a useful tool to find PCR-RFLP primers which correspond to the primer constraints. The *in silico* simulation of the proposed method showed that it reliably fits the constraints to the primers (Table 1 and Table 2). Dejong and Spears' parameter settings are the standard used for most GAs, and for this reason were used the same parameter settings in the present study. Typically, crossover is applied at more than or equal to the rate of 0.6, and the mutation rate is equal to 0.001 [27]. However, the population size 50 used in Dejong and Spears' parameter settings is too small to provide the necessary sampling accuracy for the design of natural PCR-RFLP primer sets. Consequently, we increased the population size from 50 to 100, 200, 300, 400, 500, 600, 700, 800, 900, and 1000 to ensure accurate sampling. Along with an increased population size, the average fitness value was decreased. In other words, a larger population size is better suited for natural PCR-RFLP primer design (Table 2). The standard deviation shows that the population size greatly influences the GC% (SD=20.44) and the GC clamp (SD=32.35). The population size only slightly influences the primer length

ISBN: 978-988-17012-8-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

difference (SD=0.70), the T_m (SD=0.84), the T_m difference (SD=1.34) and the average fitness value (SD=0.73); it has no influence on the product length, the dimer, the hairpin and the specificity.

V. Conclusion

PCR-RFLP which is a simple, inexpensive, and accurate laboratory technique had been applied to many SNP genotyping experiments. However, the natural PCR-RFLP primer design still is a challenging. In this paper, we propose a method which uses a GA to search for optimal natural PCR-RFLP primers and employs the core of SNP-RFLPing to reliably mine available restriction enzymes to solve the problem. The *in silico* simulation of the proposed method had showed the reliability for certain polymorphisms (251 SNPs in SLC6A4 gene). In conclusion, the proposed method is a reliable method for designing feasible natural PCR-RFLP primers which conform to most of the primer constraints and provides the available restriction enzymes from REBASE.

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." vol. 90: Nature Publishing Group, 2004, pp. 747-751.

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

[5] C. R. Cantor, "The use of genetic SNPs as new diagnostic markers in preventive medicine." vol. 1055, 2005, pp. 48-57.

[6] T. Bernig and S. J. Chanock, "Challenges of SNP genotyping and genetic variation: its future role in diagnosis and treatment of cancer." vol. 6: Future Drugs Ltd London, UK, 2006, pp. 319-331.

[7] W. Sadee, "Pharmacogenomics: Harbinger for the Era of Personalized Medicine?." vol. 5: ASPET, 2005, pp. 140-143.

[8] A. M. Divne and M. Allen, "A DNA microarray system for forensic SNP analysis." vol. 154: Elsevier, 2005, pp. 111-121.

[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. 2857-64, 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,* vol. 33, pp. W489-92, Jul 1 2005.

[11] R. J. Roberts, T. Vincze, J. Posfai, and D. Macelis, "REBASE--enzymes and genes for DNA restriction and modification," *Nucleic Acids Res,* vol. 35, pp. D269-70, Jan 2007.

[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,* 7:379, 2006.

[13] S. Rozen and H. Skaletsky, "Primer3 on the WWW for general users and for biologist programmers," *Methods Mol Biol,* vol. 132, pp. 365-86, 2000.

[14] T. Koressaar and M. Remm, "Enhancements and modifications of primer design program Primer3," *Bioinformatics,* vol. 23, pp. 1289-91, May 15 2007.

[15] F. M. You, N. Huo, Y. Q. Gu, M. C. Luo, Y. Ma, D. Hane, G. R. Lazo, J. Dvorak, and O. D. Anderson, "BatchPrimer3: a high throughput web application for PCR and sequencing primer design," *BMC Bioinformatics,* 9:253, 2008.

[16] 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. 421-31, May 2009.

[17] M. J. McPherson, G. R. Taylor, and P. Quirke, *PCR, a practical approach*: Oxford University Press, USA, 1991.

[18] J. Sambrook and D. W. Russell, "Molecular cloning: a laboratory manual," *Cold Spring Harbor Laboratory Press,* vol. 2, pp. 8.1-8.126, 2001.

[19] J. S. Wu, C. Lee, C. C. Wu, and Y. L. Shiue, "Primer design using genetic algorithm," *Bioinformatics,* vol. 20, pp. 1710-7, Jul 22 2004.

[20] K. D. Jong, "Learning with genetic algorithms: an overview," *Mach. Learning,* vol. 3, pp. 121-138, 1988.

[21] D. E. Goldgerg, *Genetic algorithms in search, optimization, and machine learning*. New York: Addison-Wesley 1989.

[22] 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 Genomics,* 7:30, 2006.

[23] 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. 2001-7, Jul-Aug 2008.

[24] J. Holland, *Adaptation in Nature and Artificial Systems*: MIT Press, 1992.

[25] 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. 308-11, Jan 1 2001.

[26] 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. 3543-57, Aug 10 1979.

[27] K. A. De Jong and W. M. Spears, "An analysis of the interacting roles of population size and crossover in genetic algorithms." vol. 1: Springer, 1990, pp. 38–47.

[28] 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. 103-8, Jun 2009.

[29] L. Mandelli, M. Mazza, G. Martinotti, M. Di Nicola, T. Daniela, E. Colombo, S. Missaglia, D. De Ronchi, R. Colombo, L. Janiri, and A. Serretti, "Harm avoidance moderates the influence of serotonin transporter gene variants on treatment outcome in bipolar patients," *J Affect Disord,* p. Epub ahead of print, Mar 28 2009.

[30] 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. 251-7, Aug 2008.

[31] 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. 147-9, 2008.

[32] T. Kampke, M. Kieninger, and M. Mecklenburg, "Efficient primer design algorithms," *Bioinformatics,* vol. 17, pp. 214-25, Mar 2001.

[33] S. H. Chen, C. Y. Lin, C. S. Cho, C. Z. Lo, and C. A. Hsiung, "Primer Design Assistant (PDA): A web-based primer design tool," *Nucleic Acids Res,* vol. 31, pp. 3751-4, Jul 1 2003.

[34] 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.

[35] J. Wang, K. B. Li, and W. K. Sung, "G-PRIMER: greedy algorithm for selecting minimal primer set," *Bioinformatics,* vol. 20, pp. 2473-5, Oct 12 2004.

[36] 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. 181-8, Apr 2009.

ISBN: 978-988-17012-8-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)