PCR-CTPP design based on Particle Swarm Optimization with Fuzzy Adaptive Strategy

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

***Abstract*—PCR-CTPP (Polymerase chain reaction with confronting two-pair primers) is a time- and cost-effective SNP genotyping method. However, the design of feasible PCR-CTPP primer sets is still challenging. In this study, we propose a PSO (particle swarm optimization) with fuzzy adaptive strategy, named FAPSO, to design feasible PCR-CTPP primer sets. Two hundred and eighty-eight SNPs which exclude the deletion/insertion polymorphism (DIP) and multi-nucleotide polymorphism (MNP) in SLC6A4 gene were tested *in silicon* by the proposed method. The results shown the proposed method provides more feasible PCR-CTPP primers than a GA (genetic algorithm)-based and a native PSO-based method. In conclusion, the FAPSO-based method is useful to assist the biologists and researchers to design feasible PCR-CTPP primer sets.**

***Index Terms*—PCR-CTPP, SNP, GA, PSO, FAPSO**

I. Introduction

SNPs (Single Nucleotide Polymorphisms) are usually used in association studies of diseases and cancers due to their great quantity. Many laboratories have introduced high-throughput platforms of SNP genotyping such as real-time PCR (polymerase chain reaction) [1] and SNP array [2] to validate SNPs or novel mutations, but they are more expensive than the other existing methods. The PCR-restriction fragment length polymorphism (RFLP) genotyping [3-5] is still favorite due to its inexpensive for the small-scale genotyping. However, the chief shortcoming of the PCR-RFLP is usually long digestion time in 2-3 hours for restriction enzymes [6, 7].

Recently, a restriction enzyme-free SNP genotyping technique [8, 9] was developed named PCR-CTPP (PCR with confronting two-pair primers). PCR-CTPP has genotyped many SNPs successfully, such as interleukin-1B (IL-1B) C-31T, interleukin-2 (IL-2) -330G, beta2-adrenergic receptor (beta2-AR) Gln27Glu, aldehyde dehydrogenase 2 (ALDH2) [10], pylori-induced gastric atrophy [11], severe coronary artery disease [12], and esophageal cancer risk [13]. Although PCR-CTPP is suitable and reliable for most cases of SNPs, the existing computation methods for designing feasible PCR-CTPP primer sets are still insufficient.

Many typical primer design constraints need to be considered when designing PCR-CTPP primers, such as primer length, length difference of primer pair, PCR product length, GC proportion, melting temperature (T_m), difference in melting temperature ($T_{m\text{-}diff}$), GC clamp, the existence of dimers (including cross-dimers and self-dimers), hairpin structure, and specificity. The important factor is especially $T_{m\text{-}diff}$ [14]. That makes design the feasible PCR-CTPP primers challenging. In the past, we have introduced a genetic algorithm (GA) to design available PCR-CTPP primer sets [15, 16]. However, the computational results still need to be improved, especially the T_m difference. In order to design the more feasible PCR-CTPP primer sets, we applied particle swarm optimization (PSO) [17] to improve the problem [18]. Particle swarm optimization (PSO) simulates the social behavior of organisms, such as birds in a flock or fish in a school, is a population-based stochastic optimization technique developed by Kennedy and Eberhart [17]. In a PSO, each single candidate solution is described as "an individual bird of the flock", that is, a particle in the search space. Each particle finds the better solution using its own memory as well as knowledge gained by the swarm. Each particle has a fitness value evaluated by an optimized fitness function and a velocity directs the movement of the particles. During movement, each particle adjusts its position in terms of its own experience and the experience of a neighbouring particle, thus making the best position encounter. PSO has been successfully applied in many fields, such as function optimization, artificial neural network training, and fuzzy system control. A comprehensive survey of PSO algorithms and their applications can be found in Kennedy *et al.* [19]. However, a fixed inertia weight or linearly decreasing inertia weight used in PSO simplifies the complex non-linear search process [20, 21]. In order to balance the global and local search ability of PSO, a fuzzy system adapts the inertia weight of PSO dynamically had been implemented [22]. Following, we also introduced the fuzzy adaptive strategy to the PCR-CTPP primer design problem [23]. The PSO with fuzzy adaptive strategy, named FAPSO (Fuzzy Adaptive Particle Swarm Optimization) here, shows the better PCR-CTPP primers designed. In this paper, we will further discuss the results of PCR-CTPP primer design for GA, PSO, and FAPSO.

Manuscript received April 30, 2012. This work was supported in part by the National Science Council in Taiwan under grants NSC99-2622-E-151-019-CC3, NSC99-2221-E-151-056-, NSC100-2221-E-151-049-MY3, and NSC100-2221-E-151-051-MY2.

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

Yu-Huei Cheng is with the Department of Network Systems, Toko University, Chiayi, Taiwan (e-mail: yuhuei.cheng@gmail.com).

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

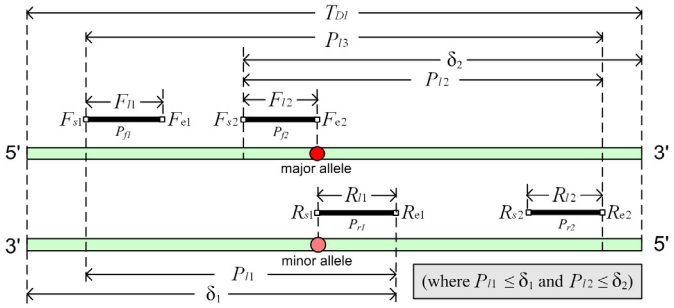

Fig. 1. Parameters of the DNA template and the PCR-CTPP primer set. Symbols indicate: *F*: Forward primer; *R*: Reverse primer; *s*: Start nucleotide position; *e*: End nucleotide position; *P*: Length of PCR product using a primer set (F/R); *l*: Length of primer or product; T_{Dl}: Length of DNA template; δ_1: Length from the upstream end of DNA template to R_{e1}; δ_2: Length from F_{s2} to the downstream end of DNA template.

II. Method

In this paper, we introduced the PSO with fuzzy adaptive strategy for the design of PCR-CTPP primer sets. The problem definition and the proposed method of PCR-CTPP primer design is described below in detail.

A. Problem formulation

Let T_D be the DNA template sequence composed of nucleotide codes with an identified SNP. T_D is defined as:

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

where B_i is the regular nucleotide code ('A', 'T', 'C', or 'G') mixed with a single IUPAC code of SNP ('M', 'R', 'W', 'S', 'Y', 'K', 'V', 'H', 'D', 'B' or 'N') ($\exists!$ is the symbol for existence and uniqueness).

The PCR-CTPP primer design is to find four available primers in T_D based on a defined SNP site as illustrated in Fig. 1 (black bar). The forward primer 1 (P_{f1}) is a short sense sequence in the upstream (5' end) of T_D far from a defined SNP site; the reverse primer 1 (P_{r1}) is a short antisense sequence which contains a nucleotide (the minor allele in the defined SNP site) located at its 3' end; the forward primer 2 (P_{f2}) is a short sense sequence which contains a nucleotide (the major allele in the defined SNP site) located at its 3' end, and the reverse primer 2 (P_{r2}) is the antisense sequence in the downstream of T_D far from a defined SNP site. These four primers are defined as follows:

$$P_{f1} = \{B_i \mid i \text{ is the index of } T_D, \; F_{s1} \le i \le F_{e1}\} \quad (2)$$

$$P_{r1} = \{\overline{B_i} \mid i \text{ is the index of } T_D, R_{s1} \le i \le R_{e1}\} \quad (3)$$

$$P_{f2} = \{B_i \mid i \text{ is the index of } T_D, F_{s2} \le i \le F_{e2}\} \quad (4)$$

$$P_{r2} = \{\overline{B_i} \mid i \text{ is the index of } T_D, R_{s2} \le i \le R_{e2}\} \quad (5)$$

where both P_{f1}/P_{r1} and P_{f2}/P_{r2} are two sets of primer pairs. F_{s1} *vs.* F_{e1} and R_{s1} *vs.* R_{e1} indicate the start index *vs.* the end index of P_{f1} and P_{r1} in T_D, respectively. F_{s2} *vs.* F_{e2} and R_{s2} *vs.* R_{e2} indicate the start index *vs.* the end index of P_{f2} and P_{r2} in T_D, respectively. $\overline{B_i}$ is the complementary nucleotide of B_i, which is described in formula (1).

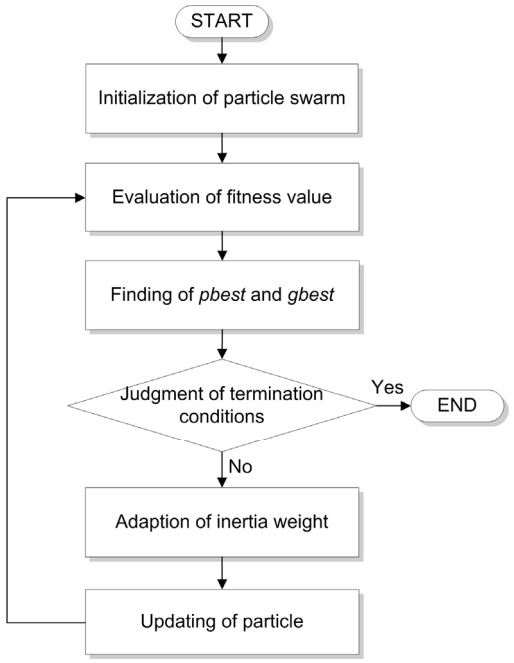

Fig. 2. Flowchart of the FAPSO-based PCR-CTPP primer design. At first, the velocities and positions of a specific number of particles are generated randomly. And then all fitness values of all particles are calculated by the fitness function. Find out the *pbest* from each particle and find out *gbest* from all particles. A judgment on termination conditions is carried out, and if the termination conditions are reached then the algorithm will be finished, else the algorithm proceeds with the following processes. Use fuzzy strategy to adapt the inertia weight according to the previous result, and then update velocities and positions for all particles based on the updating formulas. Repeat related steps shown as the figure until the best solution is found or the preset generation number is reached.

The SNP site defined at the 3' end positions of P_{f2} and P_{r1} indicated by the symbols F_{e2} and R_{s1} in Fig. 1, respectively. A vector (v) with F_{l1}, P_{l1}, R_{l1}, F_{l2}, P_{l2} and R_{l2} is used to represent a PCR-CTPP primer set. This vector is defined as follows:

$$P_v = (F_{l1}, P_{l1}, R_{l1}, F_{l2}, P_{l2}, R_{l2}) \quad (6)$$

F_{l1}, P_{l1}, R_{l1}, F_{l2}, P_{l2} and R_{l2} represent the number of nucleotides of the forward primer 1, product length between P_{f1} and P_{r1}, reverse primer 1, forward primer 2, product length between P_{f2} and P_{r2} and reverse primer 2, respectively. The forward and the reverse primers can be calculated from P_v. Therefore, P_v is used to perform evolutionary computations as described in the following sections.

B. PCR-CTPP design method

The flowchart of the proposed method is shown as Fig. 2. The proposed method consists of six processes: (1) initialization of particle swarm, (2) evaluation of fitness value, (3) finding of *pbest* and *gbest*, (4) judgment of termination conditions, (5) adaption of inertia weight, and (6) updating of particle, are described below.

(1) Initialization of particle swarm

To start the algorithm, particles (P_v) of particular number

are randomly generated for an initial population without duplicates. F_{l1}, R_{l1}, F_{l2} and R_{l2} are randomly generated between the minimum and the maximum length of the primer length constraint. The minimum and maximum primer length constraints are set to between 16 nt and 28 nt, respectively. The PCR product lengths, P_{l1} and P_{l2} are randomly generated between 100 bp and δ_1, and between 100 bp and δ_2, respectively. (δ_1 and δ_2 are maximum tolerant PCR product length of P_{l1} and P_{l2} shown in Fig.1)

(2) Evaluation of fitness value

The fitness value in the fitness function is used to individually ascertain that a particle (i.e., solution) is either good or bad. We use formula (7) [15] to evaluate the fitness values of all particles in the population for related operations later.

$$\begin{aligned} Fitness(P_v) = {} & 3 \times (Len_{diff}(P_v) + GC_{proportion}(P_v) + GC_{clamp}(P_v)) \\ & + 10 \times (dimer(P_v) + hairpin(P_v) + specificity(P_v)) \\ & + 50 \times (T\mathrm{m}(P_v) + T\mathrm{m}_{diff}(P_v)) + 100 * Avg\_T\mathrm{m}_{diff}(P_v) \\ & + 60 * PCR\mathrm{len}_{ratio}(P_v) \end{aligned} \quad (7)$$

The weights (3, 10, 50, 60 and 100) of the fitness function are applied to estimate the importance of the primer constraints. These weights are set by the experiential conditions for PCR-CTPP. They also accept adjustment based on the experimental requirements. The respective function is described as follows:

Primer length

A feasible primer length for a PCR experiment is set between 16 nt and 28 nt. Since the random values of F_{l1}, R_{l1}, F_{l2} and R_{l2} have been limited by the preset constraint condition, the primer length estimation does not be considered to join to the fitness function. A length difference (Len_{diff}) less than or equal to 3 nt between the F_{l1}/R_{l1}, F_{l2}/R_{l2}, and F_{l1}/R_{l2} primer sets is considered optimal. The $Len_{diff}(P_v)$ function is used to judge the constraint.

GC proportion

In general primer design, the typical GC proportion constraint is set between 40% and 60%. However, the designed PCR-CTPP primers contain the target SNP limiting the range of the GC proportion. To relax this constraint, the constraint of GC proportion in a primer is adjusted to between 20% and 80%. The $GC_{proportion}(P_v)$ function is proposed to lead the GC proportion of PCR-CTPP primers corresponding this constraint.

GC Clamp

To meet the presence of 'G' or 'C' at the 3' terminal of a primer to ensure a tight localized hybridization bond, the $GC_{clamp}(P_v)$ function is proposed to meet the criterion.

Melting temperature

The melting temperature (T_m) for each PCR-CTPP primer must be considered carefully for PCR experiments. The T_m calculation formula for a primer is described as follows:

$$\begin{aligned} T\mathrm{m}_{BM}(P) = {} & 81.5 + 16.6 * (\log_{10}[\mathrm{Na}^+]) \\ & + 0.41 * (\mathrm{GC}\%) - 675 / |P| \end{aligned} \quad (8)$$

where P represents a primer and $|P|$ represents the length of primer P; Na^+ is the molar salt concentration. The suffix BM represents the formula which was proposed by Bolton and McCarthy [24].

The $T\mathrm{m}(P_v)$ function is proposed to confine a PCR-CTPP primer set ranging from 45 ^{o}C to 62 ^{o}C. Similar T_m between a primer pair is important when a PCR experiment is performed in a tube. The $T\mathrm{m}_{diff}(P_v)$ function is proposed to guide the difference of the melting temperatures to less than or equal to 1 ^{o}C. In order to balance the T_m values among a PCR-CTPP primers, the $Avg\_T\mathrm{m}_{diff}(P_v)$ function is proposed to calculate the average T_m difference.

Dimer and hairpin

Primer dimers (annealing of two primers), such as cross-dimers (a forward primer and a reverse primer) and self-dimers (two forward primers or two reverse primers) must be avoided. To check for the occurrence of primer dimers, the function $dimer(P_v)$ is proposed. The hairpin represents primer annealing to itself. To check for avoiding the presence of a hairpin structure in PCR-CTPP primers, the $hairpin(P_v)$ function is proposed.

Specificity

The function $specificity(P_v)$ is proposed to check for whether each PCR-CTPP primer reappearance in the template DNA sequence to ensure its specificity. The PCR experiment may fail when a designed primer is not sequence-specific (i.e., it appears more than once in the DNA template).

PCR product length

The $PCR\mathrm{len}_{ratio}(P_v)$ function is proposed to calculate the appropriate PCR product length. Three ratios, i.e., ratio1, ratio2 and ratio3, are introduced to the function of $PCR\mathrm{len}_{ratio}(P_v)$ representing P_{l1}, P_{l2} and P_{l3}, respectively. The minimum PCR product length must be greater than 100 bp.

(3) Finding of pbest and gbest

One of the characteristics of PSO is that each particle has a memory of its own best experience. Each particle finds its personal best position and velocity (called *pbest*) and the global best position and velocity (called *gbest*) when moving. If the fitness of a particle P_v is better than the fitness of *pbest* in the previous generation, *pbest* will be updated to P_v in the current generation. If the fitness of a particle P_v is better than *gbest* in the previous generation and is the best one in the current generation, *gbest* will be updated to P_v. Each particle adjusts its direction based on *pbest* and *gbest* in the next generation.

(4) Judgment of termination conditions

The algorithm is terminated when *gbest* has achieved the best position, i.e., its fitness value is 0, or when a maximum number of generations have been reached.

TABLE I
VARIABLES AND THE RANGE OF VARIABLES

Variables	Functions	Range of variables (x_1, x_2)
NCBPE	$f_{leftTriangle}$	(0, 0.006)
	$f_{Triangle}$	(0.05, 0.4)
	$f_{rightTriangle}$	(0.3, 1)
weight	$f_{leftTriangle}$	(0.2, 0.06)
	$f_{Triangle}$	(0.4, 0.9)
	$f_{rightTriangle}$	(0.6, 1.1)
w_change	$f_{leftTriangle}$	(-0.12, -0.02)
	$f_{Triangle}$	(-0.04, 0.04)
	$f_{rightTriangle}$	(0.0, 0.05)

TABLE II
FUZZY RULES

	If	Then
1	*NCBPE* = low and *weight* = low	*w_change* = medium
2	*NCBPE* = low and *weight* = medium	*w_change* = low
3	*NCBPE* = low and *weight* = high	*w_change* = low
4	*NCBPE* = medium and *weight* = low	*w_change* = high
5	*NCBPE* = medium and *weight* = medium	*w_change* = medium
6	*NCBPE* = medium and *weight* = high	*w_change* = low
7	*NCBPE* = high and *weight* = low	*w_change* = high
8	*NCBPE* = high and *weight* = medium	*w_change* = medium
9	*NCBPE* = high and *weight* = high	*w_change* = low

(5) Adaption of inertia weight

In PSO, the inertia weight is used to balance the global and local search ability. A large inertia weight facilitates a global search while a small inertia weight facilitates a local search [22]. In order to adjust the search ability, the inertia weight is changed dynamically using a fuzzy system. In this paper, we used the fuzzy system proposed by Shi and Eberhart [22]. According to Shi and Eberhart, two variables are used as inputs for the fuzzy system. The first variable is the current best performance evaluation (*CBPE*) and the other is the current inertia weight. The *CBPE* is converted into a normalized format so that it can be applied to apply to the specific primer design problem. The normalized *CBPE* (*NCBPE*) is described by:

$$NCBPE = \frac{CBPE - CBPE_{\min}}{CBPE_{\max} - CBPE_{\min}} \tag{9}$$

where $CBPE_{\min}$ represents the real minimum and $CBPE_{\max}$ represents the real maximum.

Three fuzzy sets "low", "medium" and "high" are used. They are associated to the membership functions of $f_{leftTriangle}$, $f_{Triangle}$ and $f_{rightTriangle}$, respectively. The membership functions are defined as:

$$f_{left\_triangle}(x) = \begin{cases} 1 & , \text{if } x < x_1 \\ \dfrac{x_2 - x}{x_2 - x_1} & , \text{if } x_1 \le x \le x_2 \\ 0 & , \text{if } x > x_2 \end{cases} \tag{10}$$

$$f_{triangle}(x) = \begin{cases} 0 & , \text{if } x < x_1 \\ 2\dfrac{x - x_1}{x_2 - x_1} & , \text{if } x_1 \le x \le \dfrac{x_2 + x_1}{2} \\ 2\dfrac{x_2 - x}{x_2 - x_1} & , \text{if } \dfrac{x_2 + x_1}{2} \le x \le x_2 \\ 0 & , \text{if } x > x_2 \end{cases} \tag{11}$$

$$f_{right\_triangle}(x) = \begin{cases} 0 & , \text{if } x < x_1 \\ \dfrac{x - x_1}{x_2 - x_1} & , \text{if } x_1 \le x \le x_2 \\ 1 & , \text{if } x > x_2 \end{cases} \tag{12}$$

where x_1 and x_2 determine the shape and location of the functions.

Table I and Table II list whole fuzzy system for adapting the inertia weight in the problem dynamically.

(6) Updating of particle

In each generation, the particles will change their position and velocity. Equations (13) and (14) give the updating formulas for each particle.

$$\begin{aligned} v_i^{next} = w \times v_i^{current} + c_1 \times r_1 \times (s_i^p - s_i^{current}) \\ + c_2 \times r_2 \times (s^g - s_i^{current}) \end{aligned} \tag{13}$$

$$s_i^{next} = s_i^{current} + v_i^{next} \tag{14}$$

In equations (13) and (14), v_i^{next} is the updated velocity of the *i*th particle; $v_i^{current}$ is the current velocity of the *i*th particle; c_1 and c_2 are the acceleration constants; w is the inertia weight; r_1 and r_2 are a number which is randomly generated within 0~1; s_i^p is the personal best position of the *i*th particle; s^g is the global best position in the particles; $s_i^{current}$ is the current position of the *i*th particle; s_i^{next} is the updated position of the *i*th particle.

III. RESULTS AND DISCUSSION

A. Template sequence

A point mutation in the SLC6A4 gene was identified and shown to be associated with psychosis [25], bipolar [26], and autism spectrum disorders [27] patients. Here, two hundred and eighty-eight SNPs in SLC6A4 gene were used to estimate the efficiency of the proposed method exclude the deletion/insertion polymorphism (DIP) and multi-nucleotide polymorphism (MNP). The used version is NCBI dbSNP build 130. All SNPs were retrieved with 500 bp flanking length (at both sides of the target SNP) from SNP-Flankplus (http://bio.kuas.edu.tw/snp-flankplus/) [28] as available template sequences.

B. Parameter settings

Four main parameters are set for the GA-based method, i.e., the number of iterations (generations), the population

TABLE III
RESULTS OF PCR-CTPP PRIMER DESIGN

Constraints	GA satisfactory rate (%)	PSO satisfactory rate (%)	FAPSO satisfactory rate (%)
primer length difference	75.12	82.87	86.00
GC%	96.09	96.61	97.31
GC clamp	55.99	57.73	60.94
T_m	86.63	95.40	96.53
T_m difference	23.61	58.80	73.61
product length	71.18	57.06	57.06
dimer	95.56	96.15	96.81
hairpin	85.94	86.63	89.06
specificity	96.96	98.09	97.57

The results of the designed PCR-CTPP primers showing the satisfactory rate (%) based on PSO-based and FAPSO-based methods for SNPs of the SLC6A4 gene.

TABLE IV
RESULTS OF PCR-CTPP PRIMER DESIGN

Constraints	Rate of improvement for FAPSO-GA (%)	Rate of improvement for FAPSO-PSO (%)
primer length difference	10.88	3.13
GC%	1.22	0.70
GC clamp	4.95	3.21
T_m	9.90	1.13
T_m difference	50.00	14.81
product length	-14.12	0.00
dimer	1.25	0.66
hairpin	3.13	2.43
specificity	0.61	-0.52

Comparison of the results of the designed PCR-CTPP primers showing the satisfactory rate (%) based on FAPSO-based method with GA-based and PSO-based methods for SNPs in the SLC6A4 gene.

size, the probability of crossover and the probability of mutation. The respective values are 1000, 50, 0.6 and 0.001; the values are based on DeJong and Spears' parameter settings [29]. Four main parameters are set for both the FAPSO- and PSO-based methods, i.e., the number of iterations (generations), the number of particles, the inertia weight w, and the acceleration constants c_1 and c_2. Their values are set to 50, 10, 0.8, 2 and 2, respectively. Furthermore, we change the number of generations to 100 and 1000, respectively to observe its influences.

C. The results for the GA-, PSO- and FAPSO- based PCR-CTPP primer design methods

We design the primer lengths are all randomly generated between 16 nt and 28 nt for the 288 SNPs. The results of the entire designed PCR-CTPP primers are shown in Table III.

For GA, 75.12% designed primers satisfy the length difference criterion. Most of the primer length differences are between 0 and 5 bp (data not shown). For GC%, 96.09% primers satisfy the criterion; only 30 primers are less than 20%, 25 primers are more than 80% (data not shown). There are 55.99% primers satisfy the GC clamp criterion. Most of the designed primers also satisfy T_m (86.63%); less than half of the primer pairs are satisfied with the T_m difference criteria (23.61%). The criterion for the product length is satisfied in 71.18%. For the criteria for primer dimer and specificity, only few primers are problematic (4.44% and 3.04%, respectively). For hairpin, 14.06% primers are not satisfactory.

For PSO, 82.87% designed primers satisfy the length difference criterion. Most of the primer length differences are between 0 and 5 bp (data not shown). For GC%, 96.61% primers satisfy the criterion; only 15 primers are less than 20%, 28 primers are more than 80% (data not shown). There are 57.73% primers satisfy the GC clamp criterion. Most of the designed primers also satisfy T_m (95.40%); more than half of the primer pairs are satisfied with the T_m difference criteria (58.80%). The criterion for the product length is satisfied in 57.06%. For the criteria for primer dimer and specificity, only few primers are problematic (3.85% and 1.91%, respectively). For hairpin, 13.37% primers are not satisfactory.

For FAPSO, 86.00% designed primers satisfy the length difference criterion. Most of the primer length differences are between 0 and 5 bp (data not shown). For GC%, 97.31% primers satisfy the criterion; only 12 primers are less than 20%, 23 primers are more than 80% (data not shown). There are 60.94% primers satisfy the GC clamp criterion. Most of the designed primers also satisfy the T_m (96.53%) and the T_m difference criteria (73.61%). The criterion for product length is satisfied in 57.06% of the designed primer pairs. For the criteria for primer dimer and specificity, only few primers are problematic (3.19% and 2.43%, respectively). For hairpin, 10.94% primers are not satisfactory.

D. Comparison of the results of FAPSO-based method with GA-based and PSO-based methods

From Table IV, we observe all satisfactory rates of the primer constraints using FAPSO-based method are better than GA-based method eliminating the product length criterion. The satisfied primer length difference of FAPSO-based method is higher 10.88% than GA-based method. The satisfied GC% of FAPSO-based method is higher 1.22% than GA-based method. The satisfied GC clamp of FAPSO-based method is higher 4.95% than GA-based method. The satisfied T_m and T_m difference of FAPSO-based method is higher 9.90% and 50.00% than GA-based method, respectively. The satisfied dimer and hairpin of FAPSO-based method are higher 1.25% and 3.13% than GA-based method, respectively. The satisfied specificity of FAPSO-based method is higher 0.61% than GA-based method. However, the satisfied product length of FAPSO-based method is lower 14.12% than GA-based method.

Furthermore, we also observe all satisfactory rates of the primer constraints using FAPSO-based method are better than PSO-based method eliminating the specificity criterion. The satisfied primer length difference of FAPSO-based method is higher 3.13% than PSO-based method. The satisfied GC% of FAPSO-based method is higher 0.70% than PSO-based method. The satisfied GC clamp of FAPSO-based method is higher 3.21% than PSO-based method. The satisfied T_m and T_m difference of FAPSO-based method is higher 1.13% and 14.81% than PSO-based method, respectively. The satisfied product length of FAPSO-based method is equal to PSO-based method. The

TABLE V
RESULTS OF PCR-CTPP PRIMER DESIGN USING FAPSO

ID	primer length difference	GC%	GC clamp	T_m	T_m difference	product length	dimer	hairpin	specificity	avg. fitness
1	86.00	97.31	60.94	96.53	73.61	57.06	96.81	89.06	97.57	207.79
2	84.95	96.70	62.59	95.83	82.41	58.56	96.74	92.27	97.74	145.36
3	86.11	97.57	64.15	97.66	86.69	57.52	98.13	93.32	98.00	80.00

The results of the designed PCR-CTPP primers showing in three different generations based on FAPSO method for SNPs in the SLC6A4 gene. ID1 with the number of generations is set to 50, ID2 with the number of generations was set to 100, and ID3 with the number of generations was set to 1000. The number of particles, the inertia weight w, and the acceleration constants c_1 and c_2 are all set to 10, 0.8, 2 and 2, respectively.

satisfied dimer and hairpin of FAPSO-based method are higher 0.66% and 2.43% than PSO-based method, respectively. However, the satisfied specificity of FAPSO-based method is lower 0.52% than PSO-based method.

In PCR-CTPP, the T_m difference is the most important factor that determines whether the PCR experiment is performed successfully or not. From the above results, we get the satisfactory rate of T_m difference using GA-based and PSO-based methods are greatly improved by FAPSO-based method (50.00% and 14.81%, respectively). Simultaneously, the other criteria are also improved eliminating product length is worse than GA-based method and the specificity is worse than PSO-based method.

E. Different generations used in FAPSO-based method

From Table V, we observe the satisfactory rates of the GC clamp, T_m difference, product length, hairpin, and specificity are raised (1.65%, 8.80%, 1.50%, 3.21%, and 0.17%, respectively) and the satisfactory rates of the primer length difference, GC%, T_m, and dimer are lightly decreased (1.05%, 0.61%, 0.70%, and 0.07%, respectively) when the number of generations is increased from 50 to 100 (see ID1 and ID2). Furthermore, we also observe all satisfactory rates of the primer length difference, GC%, GC clamp, T_m, T_m difference, product length, dimer, hairpin, and specificity are raised (0.11%, 0.26%, 3.21%, 1.13%, 13.08%, 0.46%, 1.32%, 4.26%, and 0.43%, respectively) when the number of generations is increased from 50 to 1000 (see ID2 and ID3). Moreover, the average fitness is improved from 207.79 to 145.36 and 80.00 when the number of generations is increased from 50 to 100 and 1000. From Table V, we suggest the higher generations the better average fitness.

In summary, the FAPSO-based PCR-CTPP primer design method is superior to the GA-based and PSO-based methods. Although the FAPSO-based method effectively improves the satisfactory rate of T_m difference, it has the shortcoming in the satisfactory rate of product length contrary to GA-based method. The entire solutions are still satisfactory; therefore it is useful to facilitate the design of PCR-CTPP primers.

IV. CONCLUSIONS

The FAPSO-based PCR-CTPP primer design method provides better melting temperature as well as better common primer constraints estimation than GA-based and PSO-based methods. Feasible PCR-CTPP primer sets are always obtained to assist SNP genotyping experiments. The experimental flexibility of the designed PCR-CTPP primers using FAPSO-based method in 288 polymorphisms has been confirmed by *in silicon* simulations. PCR-CTPP with the lower costs and shorter genotyping times, it may be applied to all kinds of high-throughput genotyping experiments in the future. To date, we have developed the PCR-CTPP primer design methods to facilitate PCR-CTPP for validating SNPs or novel mutations. In conclusion, the *in silicon* simulation results indicate that FAPSO-based method applied to PCR-CTPP primer design sets outperform the previously GA-based method and the native PSO method. FAPSO-based method is a useful tool to design feasible PCR-CTPP primers since its conformation of the most of the PCR-CTPP constraints and with a closer melting temperature among designed primers.

REFERENCES

[1] L. Hui, T. DelMonte, and K. Ranade, "Genotyping using the TaqMan assay," *Curr Protoc Hum Genet,* vol. Chapter 2, p. Unit 2 10, Jan 2008.

[2] F. Jasmine, H. Ahsan, I. L. Andrulis, E. M. John, J. Chang-Claude, and M. G. Kibriya, "Whole-genome amplification enables accurate genotyping for microarray-based high-density single nucleotide polymorphism array," *Cancer Epidemiol Biomarkers Prev,* vol. 17, pp. 3499-508, Dec 2008.

[3] 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,* vol. 7, p. 30, 2006.

[4] G. T. Lin, H. F. Tseng, C. H. Yang, M. F. Hou, L. Y. Chuang, H. T. Tai, M. H. Tai, Y. H. Cheng, C. H. Wen, C. S. Liu, C. J. Huang, C. L. Wang, and H. W. Chang, "Combinational polymorphisms of seven CXCL12-related genes are protective against breast cancer in Taiwan," *OMICS,* vol. 13, pp. 165-172, Jan 23 2009.

[5] 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 Bioinformatics,* vol. 11, p. 173, 2010.

[6] 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.

[7] NCBI, "Restriction Fragment Length Polymorphism (RFLP) http://www.ncbi.nlm.nih.gov/genome/probe/doc/TechRFLP.shtml (accessed September 2012)."

[8] N. Hamajima, T. Saito, K. Matsuo, K. Kozaki, T. Takahashi, and K. Tajima, "Polymerase chain reaction with confronting two-pair primers for polymorphism genotyping," *Jpn J Cancer Res,* vol. 91, pp. 865-8, Sep 2000.

[9] A. Tamakoshi, N. Hamajima, H. Kawase, K. Wakai, N. Katsuda, T. Saito, H. Ito, K. Hirose, T. Takezaki, and K. Tajima, "Duplex polymerase chain reaction with confronting two-pair primers (PCR-CTPP) for genotyping alcohol dehydrogenase beta subunit (ADH2) and aldehyde dehydrogenase 2 (ALDH2)," *Alcohol Alcohol,* vol. 38, pp. 407-10, Sep-Oct 2003.

[10] N. Katsuda, N. Hamajima, A. Tamakoshi, K. Wakai, K. Matsuo, T. Saito, K. Tajima, and S. Tominaga, "Helicobacter pylori seropositivity and the myeloperoxidase G-463A polymorphism in combination with interleukin-1B C-31T in Japanese health checkup examinees," *Jpn J Clin Oncol,* vol. 33, pp. 192-7, Apr 2003.

[11] S. Togawa, T. Joh, M. Itoh, N. Katsuda, H. Ito, K. Matsuo, K. Tajima, and N. Hamajima, "Interleukin-2 gene polymorphisms associated with increased risk of gastric atrophy from Helicobacter pylori infection,"

Helicobacter, vol. 10, pp. 172-8, Jun 2005.
[12] K. K. Abu-Amero, O. M. Al-Boudari, G. H. Mohamed, and N. Dzimiri, "The Glu27 genotypes of the beta2-adrenergic receptor are predictors for severe coronary artery disease," *BMC Med Genet,* vol. 7, p. 31, 2006.
[13] S. J. Yang, H. Y. Wang, X. Q. Li, H. Z. Du, C. J. Zheng, H. G. Chen, X. Y. Mu, and C. X. Yang, "Genetic polymorphisms of ADH2 and ALDH2 association with esophageal cancer risk in southwest China," *World J Gastroenterol,* vol. 13, pp. 5760-4, Nov 21 2007.
[14] N. Hamajima, T. Saito, K. Matsuo, and K. Tajima, "Competitive amplification and unspecific amplification in polymerase chain reaction with confronting two-pair primers," *J Mol Diagn,* vol. 4, pp. 103-7, May 2002.
[15] C. H. Yang, Y. H. Cheng, L. Y. Chuang, and H. W. Chang, "Genetic Algorithm for the Design of Confronting Two-Pair Primers," *Ninth IEEE international Conference on BioInformatics and BioEngineering (BIBE),* pp. 242-247, 2009.
[16] C. H. Yang, Y. H. Cheng, L. Y. Chuang, and H. W. Chang, "Confronting two-pair primer design for enzyme-free SNP genotyping based on a genetic algorithm," *BMC Bioinformatics,* vol. 11, p. 509, 2010.
[17] J. Kennedy and R. Eberhart, "Particle swarm optimization," *IEEE International Conference on Neural Networks,* vol. 4, pp. 1942-1948, 1995.
[18] C. H. Yang, Y. H. Cheng, and L. Y. Chuang, "Confronting two-pair primer design using particle swarm optimization," *Lecture Notes in Computer Science: Computational Collective Intelligence. Technologies and Applications.,* vol. 6423, pp. 448-456, 2010.
[19] J. F. Kennedy, R. C. Eberhart, and Y. Shi, *Swarm intelligence*. US: Springer, 2001.
[20] Y. Shi, R. Eberhart, and Y. Chen, "Implementation of evolutionary fuzzy systems." vol. 7, 1999, pp. 109-119.
[21] R. C. Eberhart and Y. Shi, "Tracking and optimizing dynamic systems with particle swarms." vol. 1, 2001.
[22] Y. Shi, R. C. Eberhart, E. Team, and I. N. Kokomo, "Fuzzy adaptive particle swarm optimization." vol. 1, 2001.
[23] Y. H. Cheng, L. Y. Chuang, and C. H. Yang, "Fuzzy Adaptive Particle Swarm Optimization for Confronting Two-Pair Primer Design," *Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2012, IMECS 2012, 14-16 March, 2012, Hong Kong,* vol. I, pp. 100-104, 2012.
[24] J. Sambrook, E. F. Fritsch, and T. Maniatis, *Molecular cloning*: Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY, 1989.
[25] 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.
[26] 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,* vol. 119, pp. 205-9, Mar 28 2009.
[27] 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.
[28] 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.
[29] 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.