Fuzzy Adaptive Particle Swarm Optimization for Confronting Two-Pair Primer Design

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

Abstract—PCR-CTPP (Polymerase chain reaction with confronting two-pair primers) is a simple SNP genotyping method with time- and cost-effective. Many SNPs have been successfully genotyped by PCR-CTPP. However, the design of feasible PCR-CTPP primer sets is still challenging. In this study, we propose a FAPSO (fuzzy adaptive particle swarm optimization)-based method 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 native PSO (particle swarm optimization)-based method. In conclusion, the FAPSO-based method is useful to assist the biologists and researchers to design feasible CTPP primer sets.

Index Terms—PCR-CTPP, FAPSO, PSO, SNP

I. INTRODUCTION

SNPs (Single Nucleotide Polymorphisms) are usually used in association studies of diseases and cancers due to its great quantity. Many high-throughput platforms of SNP genotyping such as real-time PCR (polymerase chain reaction) [1] and SNP array [2] have been introduced to validate SNPs or novel mutations by some laboratories, but the PCR-restriction fragment length polymorphism (RFLP) genotyping [3-5] is still favorite due to its inexpensive for the small-scale genotyping. However, 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 with confronting two-pair primers (PCR-CTPP). PCR-CTPP 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]. PCR-CTPP is suitable and reliable for most cases of SNPs. However, the lack of an effective computation method for designing feasible PCR-CTPP primer sets.

In the past, we have introduced a genetic algorithm to design available PCR-CTPP primer sets [14, 15]. However, the computational results still need to be improved, especially the factor of Tm (melting temperature) difference. PCR-CTPP primers are only tolerant of a small difference in melting temperature between the four primers [16]. The fuzzy adaptive particle swarm optimization (FAPSO) [17] is therefore proposed to apply to the problem.

II. METHOD

Particle swarm optimization (PSO) developed by Kennedy and Eberhart is a population-based stochastic optimization technique [17]. PSO simulates the social behavior of organisms, such as birds in a flock or fish in a school. 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 best 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. [18]. However, a fixed inertia weight or linearly decreasing inertia weight used in PSO simplifies the complex non-linear search process [19, 20]. 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 [21]. In this paper, we introduced the fuzzy adaptive PSO (FAPSO) 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 Td be the DNA template sequence composed of nucleotide codes with an identified SNP. Td is defined as:

\[ T_d = \{ B_i | i \text{ is the index of DNA sequence}, 1 \leq i \leq t, \exists! B_i \in \text{IUPAC code of SNP} \} \]

The PCR-CTPP primer design is to find two available short sequences in $T_D$ based on a defined SNP site as illustrated (Fig. 1). 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:

$$P_{f1} = \{ B_i | i is the index of T_D, F_{f1} \leq i \leq F_{s1} \}$$  \hspace{1cm} (2)$$

$$P_{r1} = \{ B_i | i is the index of T_D, R_{r1} \leq i \leq R_{s1} \}$$  \hspace{1cm} (3)$$

$$P_{f2} = \{ B_i | i is the index of T_D, F_{f2} \leq i \leq F_{s2} \}$$  \hspace{1cm} (4)$$

$$P_{r2} = \{ B_i | i is the index of T_D, R_{r2} \leq i \leq R_{s2} \}$$  \hspace{1cm} (5)$$

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

The SNP site defined at the 3’ end positions of $P_{f2}$ and $P_{r1}$ indicated by the symbols $F_{s2}$ and $R_{s1}$ in Fig. 1, respectively. A vector (v) with $F_{f1}$, $P_{f1}$, $R_{r1}$, $F_{f2}$, $P_{f2}$ and $R_{r2}$ is used to represent a PCR-CTPP primer set. This vector is defined as follows:

$$P_v = (F_{f1}, P_{f1}, R_{r1}, F_{f2}, P_{f2}, R_{r2})$$  \hspace{1cm} (6)$$

$F_{f1}$, $P_{f1}$, $R_{r1}$, $F_{f2}$, $P_{f2}$ and $R_{r2}$ 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. Consequently, the forward and the reverse primers can be calculated from $P_v$. $P_v$ 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.
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 constraint condition, the primer length estimation does not be considered to join to the fitness function. A length difference ($\text{Len}_{\text{adj}}$) less than or equal to 3 bp between the $F_{l1}/R_{l1}$, $F_{l2}/R_{l2}$, and $F_{l1}/R_{l2}$ primer sets is considered optimal. The $\text{Len}_{\text{adj}}(P_v)$ function is used to judge the constraint.

**GC content**
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 $\text{GC}_{\text{prop}}(P_v)$ 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 $\text{GC}_{\text{clamp}}(P_v)$ 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 experiment. The $T_m$ calculation formula for a primer is described as follows:

$$T_m(BM)(P) = 81.5 + 16.6 \times \left(\log_{10}[\text{Na}^+]\right) + 0.41 \times \left([\text{GC}^\%] - 675 / |P|\right)$$

(8)

where $P$ represents a primer and $|P|$ represents the length of primer $P$; $\text{Na}^+$ is the molar salt concentration. The suffix BM represents the formula which was proposed by Bolton and McCarthy [22].

The $T_m(P_v)$ function is proposed to confine a PCR-CTPP primer set ranging from 45°C to 62°C. Similar $T_m$ between a primer pair is important when a PCR experiment is performed in a tube. The $\text{Len}_{\text{adj}}(P_v)$ function is proposed to guide the difference of the melting temperatures to less than or equal to 1°C. In order to balance the $T_m$ values among a PCR-CTPP primers, the $\text{Avg}_{\text{Len}}(P_v)$ 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 also be avoided. To check for the occurrence of primer dimers, the function $\text{dimer}(P_v)$ is proposed. In addition, the hairpin check is also implemented to avoid annealing to itself. To check for the presence of a hairpin structure in PCR-CTPP primers, the $\text{hairpin}(P_v)$ function is proposed.
\[
\begin{align*}
\dot{v}_i^p &= w \times v_i^{\text{prev}} + c_1 \times r_1 \times (s^g_i - s_i^{\text{prev}}) \\
&\quad + c_2 \times r_2 \times (s^s_i - s_i^{\text{prev}}) \\
\end{align*}
\]

(9)

\[
\begin{align*}
\v_i\text{next} &= s_i^{\text{prev}} + \dot{v}_i^{\text{next}} \\
\end{align*}
\]

(10)

In equations (9) and (10), \(\dot{v}_i^{\text{next}}\) is the updated velocity of the \(i\)th particle; \(v_i^{\text{prev}}\) 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; \(s^g_i\) is the personal best position of the \(i\)th particle; \(s^s_i\) is the global best position of the particles; \(s_i^{\text{prev}}\) is the current position of the \(i\)th particle; \(s_i^{\text{next}}\) is the updated position of the \(i\)th particle. In order to prevent a particle from overshooting the limits of \(F_s, F_v, P_l\) and \(R_l\) during the update process, we randomly reset the particle according to the primer constraints.

III. RESULTS AND DISCUSSION

A. Template sequence

A point mutation was identified and shown to be associated with psychosis [23], and bipolar [24] patients in the SLC6A4 gene. Here, two hundred and eighty-eight SNPs were used to estimate the efficiency of the proposed method excluding the deletion/insertion polymorphism (DIP) and multi-nucleotide polymorphism (MNP) in SLC6A4 gene. 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/) [25] as available template sequences.

B. Parameter settings

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 were set to 50, 10, 0.8, 2 and 2, respectively.

C. The results for the 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 1. For PSO, 82.87% designed primers satisfy the length difference criterion. Most of the length difference requirements were between 0 and 5 bp (data not shown). For GC%, 96.61% primers satisfy the criterion; only 15 primers were less than 20%, 28 primers were more than 80% (data not shown). There are 57.73% primers satisfy the GC clamp criterion. Most of the designed primers also satisfy the \(T_m\) (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, 86.0% designed primers satisfy the length difference criterion. Most of the primer length differences were between 0 and 5 bp (data not shown). For GC%, 97.31% primers satisfy the criterion; only 12 primers were less than 20%, 23 primers were 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 was satisfied in 57.06% of the designed primer pairs. For the criteria for primer dimer and specificity, only few primers were problematic (3.19% and 2.43%, respectively). For hairpin, 10.94% primers are not satisfactory.

D. Comparison of the results of FAPSO-based with PSO-based PCR-CTPP method

From Table 1, all satisfactory rates of the primer constraints using FAPSO-based method is 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 0.7% 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 4.81% than PSO-based method, respectively. The satisfied product length of FAPSO-based method is equal to PSO-based method. The 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 affects experiments. From the above results, we observe the satisfactory rate of \(T_m\) difference is greatly improved by FAPSO-based method. Simultaneously, the other criteria are also improved eliminating the specificity. Although satisfactory rate of the specificity is slightly decreased, it still arrives 97.57%. In summary, the FAPSO-based PCR-CTPP primer design method is useful to facilitate SNPs genotyping.

IV. CONCLUSIONS

The FAPSO-based PCR-CTPP primer design method provides better melting temperature as well as better common
primer constraints estimation than PSO-based method. Feasible PCR-CTPP primer sets are always obtained to assist SNP genotyping experiments. The experimental flexibility of the FAPSO-based designed PCR-CTPP primers in 288 polymorphisms has been confirmed by in silico simulations. In the future, PCR-RFLP may replace PCR-RFLP due to its lower costs and shorter genotyping times [16]. 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 silico simulation results indicate that FAPSO applied to PCR-CTPP primer design sets outperform a native PSO method. FAPSO-based method is a useful tool to design feasible PCR-CTPP primers since its confirmation of the most of the PCR-CTPP constraints and with a closer melting temperature among designed primers.

REFERENCES