Computational intelligence-based polymerase chain reaction primer selection based on a novel teaching-learning-based optimisation

Yu-Huei Cheng

Department of Digital Content Design and Management, Toko University, Chiayi, Taiwan
E-mail: yuhuei.cheng@gmail.com

Abstract: Specific primers play an important role in polymerase chain reaction (PCR) experiments, and therefore it is essential to find specific primers of outstanding quality. Unfortunately, many PCR constraints must be simultaneously inspected which makes specific primer selection difficult and time-consuming. This paper introduces a novel computational intelligence-based method, Teaching-Learning-Based Optimisation, to select the specific and feasible primers. The specified PCR product lengths of 150–300 bp and 500–800 bp with three melting temperature formulae of Wallace’s formula, Bolton and McCarthy’s formula and SantaLucia’s formula were performed. The authors calculate optimal frequency to estimate the quality of primer selection based on a total of 500 runs for 50 random nucleotide sequences of ‘Homo species’ retrieved from the National Center for Biotechnology Information. The method was then fairly compared with the genetic algorithm (GA) and memetic algorithm (MA) for primer selection in the literature. The results show that the method easily found suitable primers corresponding with the setting primer constraints and had preferable performance than the GA and the MA. Furthermore, the method was also compared with the common method Primer3 according to their method type, primers presentation, parameters setting, speed and memory usage. In conclusion, it is an interesting primer selection method and a valuable tool for automatic high-throughput analysis. In the future, the usage of the primers in the wet lab needs to be validated carefully to increase the reliability of the method.

1 Introduction

Polymerase chain reaction (PCR), is a common biotechnology applied in the fields of biomedical science and biological engineering for fast mass duplication of DNA sequences [1]. Before performing PCR, the specific primers which correspond with various PCR constraints need to be determined. In the past, Rozen and Skalentsky [2] proposed Primer3, a primer selection program which considers many different parameters to achieve different goals. Recently, Untergasser et al. [3] presented Primer3Plus, a new web interface to Primer3 as an enhanced alternative for the CGI-scripts that come with Primer3. Untergasser et al. [4] described Primer3’s current capabilities, emphasising recent improvements. Moreover, Bashir et al. [5] designed algorithms provide good solutions to primer approximation multiplex PCR (PAMP) based on simulated annealing and integer programming. Koressaar et al. [6] developed a novel computational method that identifies species-specific repeats from microbial organisms and automatically designs species-specific PCR primers for these repeats. Kitchen et al. [7] implemented Metropolis-Hastings Markov Chain Monte Carlo for optimising primer reuse. In addition, Mann et al. [8] developed Pythia, in which DNA binding affinity computations are directly integrated into the primer design process. NCBI developed Primer-BLAST [9] designs primers based on Primer3 and provides specificity check by using BLAST [10]. Chuang et al. [11] proposed a user-friendly web-based tool URPD, which uses the NCBI Reference Sequences (RefSeq) to design primers, and combines them with UCSC In-Silico PCR for redesigning the primers into more feasible ones. Batnyam et al. [12] developed UniPrimer, a web-based tool that designs PCR- and DNA-sequencing primers. Gans et al. [13] developed a software platform that enables PCR-assay design at an unprecedented scale. Karnik et al. [14] developed site-directed mutagenesis (SDM)-Assist which creates SDM primers adding a specific identifier. Yang et al. [15] developed Drug-SNPing, which provides a platform for the integration of drug information, protein–protein interactions, tagSNP selection and genotyping information with PCRrestriction fragment length polymorphism (RFLP) primer design and TaqMan probes. Although the above primer selection methods and tools were meaningful provided, a computational intelligence-based method for automatic science and engineering is still required to provide specific primers and high throughput analysis in the modern molecular technology.

In primer selection, we need to consider numerous primer constraints, such as primer length, length difference, GC content, melting temperature ($T_m$), difference of melting temperature ($T_{m-diff}$), GC clamp, dimer, hairpin and specificity [1, 2, 16]. Manual primer selection is tedious and easily yields incorrect results because of human carelessness...
and error. Consequently, select primers using automatic computation are preferable. In recent years, computational intelligence-based methods have been developed maturely and applied effectively to primer selection problems with promising results. Wu et al. [17] used a genetic algorithm (GA) that imitates nature’s process of evolution and genetic operations on chromosomes to achieve feasible primers. Yang et al. [18] applied a memetic algorithm (MA) to design specific primers to improve the shortcoming of the GA which easily attains entrapment. Moreover, Yang et al. [19] also used GA designed CTPP primers (confronting two-pair primers), natural [20] and mutagenic [21] PCR-RFLP primers for SNP genotyping. However, these GA-based methods do not have good searchability. To screen more specific and feasible primers, and promote the performance of primer selection, we applied a novel computational intelligence-based method named ‘Teaching-Learning-Based Optimisation (TLBO)’ [22, 23] to solve this problem.

2 Methods

2.1 Problem definition

The primer selection problem is to be solved by finding out two sub-sequences that mostly correspond to the primer constraints from a DNA template sequence. One sub-sequence is called forward primer and the other is called reverse primer.

Fig. 1 illustrates the parameters and the positions of the primer pair in a DNA template sequence. $T_D$ is a DNA template sequence; $T_1$ represents the length of $T_D$; $P_{\min}$ represents the preset minimum PCR product length; $P_{\max}$ represents the preset maximum PCR product length; $F_i$ is the start position of a forward primer; $F_i$ is the length of a forward primer; $F_{\max}$ is the end position of a forward primer; $R_i$ is the start position of a reverse primer; $R_i$ is the length of a reverse primer; $R_{\max}$ is the end position of a reverse primer; $P_i$ is the PCR product length between the forward primer and the reverse primer; $F_{\text{range}}$ represents the range of $F_{\max}$ which from the start of the DNA template sequence to the length of the DNA template sequence minus $P_{\min}$, and $P_{\text{range}}$ represents the range of PCR product length from $F_{\min}$ to the end of the DNA template sequence. We give a vector $L_v$ that includes four elements of $F$, $F_i$, $P_i$ and $R_i$ to represent a primer pair. The vector $L_v$ is described as

$$L_v = (F, F_i, P_i, R_i)$$

2.2 TLBO method for primer selection

The TLBO is a method based on the notion of ‘teaching’ and ‘learning’ in the subjects in a class. The TLBO assigns the number of subjects as the design variables for the primer selection problem, and the number of learners as the solutions of this problem. The flowchart of the TLBO method for primer selection is shown in Fig. 2. Five separate processes of (i) initialisation of the population of learners, (ii) learning result evaluation for the population, (iii) teacher phase, (iv) learner phase and (v) judgment of the termination conditions, are described as follows:

1. Initialisation of the population of learners: Initially, dozens of learners $L_v = (F, F_i, P_i, R_i)$ are randomly generated as an initial population without duplicates. $F_i$ is randomly generated between 1 and $(T_D - P_{\min} + 1)$. $F_i$ is randomly generated between the minimum and the maximum length of the primer according to the common primer constraints. To limit the PCR product length, the method randomly generates $P_i$ between $P_{\min}$ and $P_{\max}$. $R_i$ is randomly generated in the same way as $F_i$.

2. Learning result evaluation for the population: The TLBO primer selection method requires an evaluation function to evaluate the learning result of a learner in order to check whether the primers satisfy the primer constraints or not. The primer selection constraints are used as values for the evaluation function, and the produced value of the learning result is minimised (i.e. zero is the best learning result). The evaluation function is described as follows

$$\text{Evaluation}(L_v) = 3 \times (\text{Len}_{\text{diff}}(L_v) + \text{GC}_{\text{proportion}}(L_v)) + \text{GC}_{\text{clamp}}(L_v) + 10 \times (T_m(L_v) + T_{\text{melt}}(L_v)) + \text{dimer}(L_v) + \text{hairpin}(L_v) + 50 \times \text{specificity}(L_v)$$

The evaluation function is the same as the fitness function that has been proved by practical PCR experiments in the MA primer design method [18] and URPD [11]. We set the minimum length of primer as 16 nt and the maximum length of the primer was 28 nt. The $\text{Len}_{\text{diff}}(L_v)$ is used to evaluate the condition of primer length difference. The primer length difference smaller than or equal to 5 nt is considered better. The $\text{GC}_{\text{proportion}}(L_v)$ is used to evaluate the proportion of the ‘G’ and ‘C’ nucleotides in the forward primer and the reverse primer. The proportion of the ‘G’ and ‘C’ nucleotides in the forward primer and the reverse primer is considered better in the range of 40–60%. The $\text{GC}_{\text{clamp}}(L_v)$ is used to evaluate the condition of the primer 3’ terminal end. When the primer 3’ terminal end is either ‘G’ or ‘C’, the primer is GC clamp. GC clamp helps the primer to anneal to the DNA templates robustly. The $T_m(L_v)$ is used to evaluate the condition of primer melting temperature $T_m$ by using three known formulae of Wallace’s formula [24], Bolton and McCarthy’s formula [25] and SantaLucia’s formula [26]. The primer melting temperature $T_m$ in the range from 50 to 62°C is acceptable. $T_{\text{melt}}(L_v)$ evaluates the condition of the primer melting temperature difference. The preferable primer melting temperature difference is lower than 5°C. The dimers ($L_v$) used to check the condition when the forward primer and the reverse primer are annealed to each other or annealed to themselves. The hairpin ($L_v$) is used to check for this
Firstly, the initial population of the learners is randomly generated. All the learners in the population are then evaluated for their learning results. In the teacher phase, the best learner in the population is firstly found out as a teacher. Difference between the existing mean result of each subject and the corresponding result of the teacher for each subject is calculated. And then, the results of the learners in each subject are obtained and the better learning results are left. Two learners in the population are selected randomly to perform the learner phase. A learner learns knowledge from another learner, and the better learning result is eventually left. After the teacher and the learner phases, the learners in the population have the updated knowledge. Finally, the conditions of termination are judged if the stop criteria is reached. If not, the population of learners is evaluated again, and the above steps are repeated, else the final learning results of the learners are obtained.
condition if a primer anneals to itself. The existing dimers and hairpins are both considered to prevent successful PCR. Finally, the specificity \( (L_v) \) is used to check for the condition when a primer reappears in the DNA template sequence. The primer appearing once in the DNA template sequence is a specific primer and meant to ensure specific PCR products.

(3) **Teacher phase:** In this phase, a teacher attempts to increase the mean result of the class in the subject based on the teacher’s capability. At any iteration \( i \), assume that the ‘\( s’ \) represents a particular subject, there are ‘\( m’ \) number of subjects (i.e. \( s = 1, 2, \ldots, m \)); the ‘\( l’ \) represents a particular learner, there are ‘\( n’ \) number of learners (i.e. \( l = 1, 2, \ldots, n \)) and \( M_{s,l} \) is the mean result of the learners in a particular subject ‘\( s’ \).

Every learner has his learning result in each subject at iteration \( i \), that is, \( R_{s,l,i} \); each learner has their total learning result in all the subjects at iteration \( i \), that is, \( R_{\text{total-}l,i} \). We calculate the total learning result of every learner in all the subjects and compare them to find out the best one, that is, \( R_{\text{total-best},i} \). The best overall result \( R_{\text{total-best},i} \) considers that all the subjects together obtained in the entire learners can be taken as the result of the best learner ‘\( l_{\text{best}}\) at iteration \( i \). The identified best learner is therefore considered by the TLBO method as the teacher. The difference between the existing mean result of each subject and the corresponding result of the teacher for each subject is shown in the following equation

\[
\text{Diff}_m_{s,l,i} = r_i (R_{s,l_{\text{best}},i} - T_p M_{s,l,i})
\]  

where \( R_{s,l_{\text{best}},i} \) is the result of the best learner (i.e. teacher) in subject \( s \), \( T_p \) is the teaching factor which decides the value of the mean to be changed, and \( r_i \) is the random number in the range \([0, 1]\). The value of \( T_p \) is either 1 or 2 which is decided randomly with equal probability shown in (4). The teaching factor \( (T_p) \) is a internal parameter in the TLBO method, and its value is not given as an input to the primer selection method.

\[
T_p = \text{round} \left[ 1 + \text{and} \left( 0, \ 1 \right)(2 - 1) \right] \quad (4)
\]

According to the \( \text{Diff}_m_{s,l,i} \), the result of the learner in subject \( s \) updated in the teacher phase is shown in the following equation

\[
R'_{s,l,i} = R_{s,l,i} + \text{Diff}_m_{s,l,i}
\]  

In (5), \( R'_{s,l,i} \) is the updated value of \( R_{s,l,i} \). The method accepts \( R'_{s,l,i} \) if it is better than \( R_{s,l,i} \). All the accepted \( R'_{s,l,i} \) values at the end of the teacher phase are maintained and these values are applied to the learner phase described as follows:

(4) **Learner phase:** In the learner phase, a learner can interact randomly with other learners to enhance his knowledge. Through the interaction with other learners, a learner can obtain new things or skills from the others learners. The learning conditions of this phase are described as follows:

Two learners \( X \) and \( Y \) are selected randomly from the population of learners such that \( R'_{\text{total-}X,i} \neq R_{\text{total-}Y,i} \) and \( R_{\text{total-}X,i} \) and \( R'_{\text{total-}Y,i} \) are, respectively, the updated values of \( R_{\text{total-}X,i} \) and \( R_{\text{total-}Y,i} \) at the end of the teacher phase. The final learning result \( R'_{s,l,i} \) is shown in the following equation (see (6))

For a more detailed teacher phase and learner phase, please refer to the original TLBO literatures [22, 23].

(5) **Judgement of the termination conditions:** The method is terminated when the best learner ‘\( l_{\text{best}}\) has achieved the best learning result, that is, its learning result is 0, or when a preset maximum number of iterations have been reached. We estimate the running time by using the preset maximum number of iterations reached to observe the optimal frequency (OF) of the TLBO, the GA and the MA, for primer selection. Furthermore, we estimate the running time by using the learning result 0 reached compared with Primer3.

### 2.3 Evaluation of primer selection by using OF

To evaluate the quality of the computational intelligence-based primer selection methods, we assume that the result of primer selection must conform to the PCR primer constraints. The evaluation manner is similar to the accuracy of primer selection referred to in the literature [18]. When the learning result reaches the best value (i.e. 0) that means the method has found the optimal solution, the count of the optimal solution is an added one. Through the proportion between the count of the optimal solution and the total number of primer selection runs, we can evaluate the quality of the computational intelligence-based method for primer selection. The evaluation equation for primer selection is therefore shown in the following equation

\[
\text{OF(} \text{optimal frequency} \text{)} = \frac{n}{N} \times 100\%
\]  

where \( n \) represents the number of the final learning result when the best learner reaches zero, and \( N \) represents the total number of primer selections performed.

### 2.4 Datasets and environment

The quality of different DNA template sequences influences the results of primer selection. To reduce the influences of different template sequences for primer selection, we use 50 random nucleotide sequences of ‘Homo species’ with lengths between 1900 and 2100 bp. These template sequences are referred to in the literature [18]. The 50 nucleotide sequences can be retrieved from the NCBI Reference Sequences (http://www.ncbi.nlm.nih.gov/RefSeq/). To fairly compare the TLBO method with the GA and the MA methods in primer selection, we use the same execution environment as the literature [18], that is, a Pentium 4 CPU 3.4 GHz and 1 GB of RAM under
Microsoft Windows XP SP3, to perform primer selection. Furthermore, we also perform a fair comparison of the TLBO with Primer3 on Intel(R) Core(TM) i7-3770 CPU 3.4 GHz × 2 and 8 GB of RAM under Microsoft Windows 7 with 64 bits operation system.

2.5 Parameter settings

Two main parameters were set for the TLBO method. They are the number of iterations (i.e. generations) and the number of learners (i.e. population size). Their values were set to 100 and 8, respectively. Four main parameters were set for the MA method, namely the number of iterations, the population size, the probability of crossover and the probability of mutation. The respective values used in this study were 100, 100, 1.0 and 0.01. The GA method used the same parameters as the MA method except for the number of iterations which was set to 500. The parameters of the GA and the MA are referred to in the literature [18].

To obtain specific PCR product length, we set the

![Box plots](image)

Fig. 3 Results for the OF and the running time using the GA, the MA and the TLBO primer selection methods with the PCR product lengths of 150–300 and 500–800 bp using Wallace’s formula, Bolton and McCarthy’s formula and SantaLucia’s formula for 50 random nucleotide sequences of Homo species

Box plots

a and b Optimal frequency and the running time, respectively, with the PCR product length set to 150–300 bp

c and d Optimal frequency and the running time, respectively, with the PCR product length set to 500–800 bp
Table 1  Five-number summaries of the optimal frequency – OF (%) and the running time – t (s) for the GA, the MA and the TLBO primer selection methods with PCR product lengths of 150–300 and 500–800 bp by using Wallace’s formula, Bolton and McCarthy’s formula and SantaLucia’s formula for the 50 nucleotide sequences of ‘Homo species’

| Tm formulae       | Numbers | Primer selection methods |
|-------------------|---------|--------------------------|
|                   | 150–300 bp | 500–800 bp |
|                   | OF, %     | T, s   | OF, %     | T, s   | OF, %     | T, s   |
| Wallace’s formula |          |        |          |        |          |        |
| smallest value    | 20.60    | 786.86 | 53.40    | 764.02 | 49.80    | 727.22 |
| lower quartile    | 43.00    | 836.97 | 79.90    | 800.76 | 83.50    | 751.02 |
| median            | 50.90    | 873.98 | 88.00    | 820.95 | 90.60    | 775.62 |
| upper quartile    | 64.90    | 884.15 | 93.60    | 842.36 | 96.10    | 789.37 |
| largest value     | 76.00    | 933.03 | 98.00    | 894.41 | 98.00    | 812.66 |
| avg.              | 52.24    | 861.56 | 85.41    | 819.27 | 87.92    | 771.06 |
| Bolton and McCarthy’s formula |          |        |          |        |          |        |
| smallest value    | 0.40     | 1094.39| 18.20    | 816.41 | 1.60     | 1082.30|
| lower quartile    | 9.95     | 1216.47| 59.75    | 850.50 | 8.80     | 1205.35|
| median            | 16.60    | 1254.94| 70.70    | 870.02 | 15.50    | 1229.82|
| upper quartile    | 22.50    | 1290.34| 83.00    | 892.05 | 21.20    | 1273.49|
| largest value     | 31.20    | 1352.84| 92.60    | 902.63 | 30.40    | 1318.96|
| avg.              | 15.86    | 1250.23| 67.56    | 870.76 | 76.35    | 880.26 |
| SantaLucia’s formula |          |        |          |        |          |        |
| smallest value    | 32.40    | 1585.92| 84.60    | 865.34 | 60.20    | 765.88 |
| lower quartile    | 59.65    | 1708.15| 95.65    | 917.45 | 53.60    | 1075.28|
| median            | 69.20    | 1758.81| 97.50    | 934.76 | 66.30    | 1113.87|
| upper quartile    | 78.50    | 1798.36| 99.60    | 960.14 | 76.75    | 1192.13|
| largest value     | 91.60    | 1875.45| 100.00   | 997.47 | 90.20    | 1856.23|
| avg.              | 67.72    | 1750.71| 96.48    | 939.46 | 89.93    | 829.87 |

Bold type indicates the best value for the optimal frequency and the running time among the three primer selection methods.
parameters $P_{\text{min}}$ and $P_{\text{max}}$, respectively, as 150 and 300 bp, and 500 and 800 bps. Furthermore, we performed 500 runs of primer selection for each nucleotide sequence to decrease the stochastic effects of these computational intelligence-based methods, and to evaluate their optimal frequency based on Wallace’s formula, Bolton and McCarthy’s formula and SantaLucía’s formula. For Primer3, we set the identical primer constraints compared with the TLBO method based on SantaLucía’s formula.

3 Results

3.1 Comparison of the TLBO with the GA and the MA primer selection methods

We use box plots to represent the OF and the running time for primer selection. The comparisons among the TLBO, the GA and the MA methods for primer selection [18] are shown in Fig. 3 based on Wallace’s formula, Bolton and McCarthy’s formula and SantaLucía’s formula. The box plots summarily display the numerical data in a convenient way by graphically and intuitively depicting it in groups. Five numbers, that is, the smallest value, lower quartile, median, upper quartile and the largest value, are shown in a box plot. Table 1 completely shows the numerical values used in Fig. 3. From Fig. 3, both the optimal frequency and the running time of the TLBO method are better than those of the GA and MA methods in Wallace’s formula and Bolton and McCarthy’s formula. The optimal frequency of the TLBO method is slightly worse than those of the MA, but the running time of the TLBO method is better than the MA in SantaLucía’s formula. More details are shown in Supplementary 1.

3.2 Efficiency of the TLBO against that of the GA and the MA primer selection methods

When using Wallace’s formula, the optimal frequency of the TLBO with PCR product lengths of 150–300 and 500–800 bp is, respectively, 35.68 and 35.09% higher than those of the GA. The running time of the TLBO with PCR product lengths of 150–300 and 500–800 bp is, respectively, 90.50 s (10.50% saving) and 87.53 s (10.00% saving) faster than those of the GA. In addition, the optimal frequency of the TLBO with PCR product lengths of 150–300 and 500–800 bp is, respectively, 2.51 and 1.41% higher than those of the MA. The running time of the TLBO with PCR product lengths of 150–300 and 500–800 bp is, respectively, 48.21 s (5.88% saving) and 32.09 s (3.91% saving) faster than those of the MA.

On the other hand, when using Bolton and McCarthy’s formula, the optimal frequency of the TLBO with PCR product lengths of 150–300 and 500–800 bp is, respectively, 60.49 and 59.81% higher than those of the GA. The running time of the TLBO with PCR product lengths of 150–300 and 500–800 bp is, respectively, 369.97 s (29.59% saving) and 352.46 s (28.58% saving) faster than those of the GA. Moreover, the optimal frequency of the TLBO with PCR product lengths of 150–300 and 500–800 bp is, respectively, 8.79 and 11.79% higher than those of the MA. The running time of the TLBO with PCR product lengths of 150–300 and 500–800 bp is, respectively, 9.5 s (1.09% wasting) and 6.98 s (0.79% saving) faster than those of the MA.

Finally, when using SantaLucía’s formula, the optimal frequency of the TLBO with PCR product lengths of 150–300 and 500–800 bp is, respectively, 22.21 and 23.75% higher than the GA. The running time of the TLBO with PCR product lengths of 150–300 and 500–800 bp is, respectively, 920.84 s (52.60% saving) and 435.48 s (34.79% saving) faster than those of the GA. Moreover, the optimal frequency of the TLBO with PCR product lengths of 150–300 and 500–800 bp is, respectively, 6.55 and 7.29% lower than those of the MA. The running time of the TLBO with PCR product lengths of 150–300 and 500–800 bp is, respectively, 108.59 s (11.57% wasting) and 133.20 s (14.03% saving) faster than those of the MA.

All the values compared above show that the ability of the TLBO method for search optimal primers is greatly better than the GA method and lightly superior to the MA method in terms of the optimal frequency and the running time based on both the melting temperature formulae of Wallace’s formula and Bolton and McCarthy’s formula. Although the optimal frequency of the TLBO is gently worse than the MA, the running time of the TLBO is extremely better than the GA and preferable to the MA based on SantaLucía’s formula.

3.3 Comparison of the TLBO with Primer3

Primer3 is a comprehensively used and popular primer design method. Primer3 always design primers in an inflexible way. The inflexible way means that the designed primers are always fixed based on the same primer constraints. When using Primer3 designs primers once, twice and more times, the results are output with the same primers. Although the designed primers fail the PCR experiments, the users can readjust the primer constraint parameters for the updated primers or set the parameter of ‘Number To Return’ to obtain more pairs. In contrast to Primer3, the TLBO provides a flexible way. Primers which correspond to or are close to the preset primer constraints are effectively picked for the users. When the primers fail the PCR experiments, redesign primers can be directly achieved. No matter what the inflexible or the flexible way; the goal is to find...
available primers. The comparison of the TLBO for method type, primers presentation, parameters setting, speed and memory usage with Primer3 is listed in Table 2.

4 Discussion

The quality of the primers directly influences the PCR experiment. Although many primer selection methods had been proposed, the creative and computational abilities are still demanded. In this study, we use the novel TLBO method to select primers which correspond with mostly common primer constraints, and perform computer experiments to estimate the performance of the proposed method. The qualities of the TLBO for primer selection were evaluated based on both the fitness function inside and the frequency of the optimal solution outside. The obtained results are very encouraging and should be verified in future real PCR experiments.

(1) The melting temperature calculations in primer selection methods: In this study, we use three different melting temperature formulae to perform primer selection in the same template sequences, that is, 50 random nucleotide sequences of ‘Homo species’. The different melting temperature formulae give extremely different results (see Fig. 3 and Table 1). Wallace’s formula calculates the melting temperature with the easy calculations only by using the sum of 2 and 4°C to evaluate the nucleotides of a primer, therefore it has small-scale search space for primer selection. The Bolton and McCarthy’s formula has more complicated calculations, and the running time is naturally increased. Furthermore, the complicated calculations make the value of the melting temperature more refined and thus the search space for primer selection is certainly huge resulting in difficulty in finding the optimal primers. The SantaLucia’s formula is considered a more accurate calculation for primer melting temperature. It uses nearest neighbour method and thermodynamic parameters for estimating primer melting temperature close to practical primer melting temperature. The elaborate calculations make the optimal primers difficult to find, and the running time is reasonably raised.

(2) Explanation of the parameter settings among the computational intelligence-based methods: In the test experiments of primer selection, the parameter settings are obviously different among the computational intelligence-based methods. In terms of the generations, the GA uses 500 iterations, but both the MA and the TLBO use 100 iterations. In terms of the population, both the GA and the MA only use a population size of 100, but the TLBO only use population size (i.e. the number of learners) of 8. Furthermore, the parameters of the probability of crossover and the probability of mutation in the GA and the MA are unnecessary in the TLBO method. Here, we give a reasonable explanation for these parameters settings. In this study, we mainly compare three methods according to their performance for primer selection. By the intuitive comparison, that is, the fastest and the best result is the winner under the normal conditions, we can determine which the better method is. The GA and the MA had been compared in the literature [18] in terms of their accuracy and efficiency. In the same parameter settings, the result of the MA with the iterations set to 100 is better than the result of the GA with the equal iterations and more iterations (i.e. iterations are 500). Simultaneously, the running time of the MA is also faster than the one of the GA in the two different iterations. Therefore the MA is certainly better than the GA for primer selection. The TLBO is not yet compared with any primer selection methods. To prove that the performance of the TLBO primer selection method is better than the GA and the MA primer selection methods, we set the iterations as 100, the same as the GA and the MA primer selection methods, and use the smaller population size (i.e. the number of learners are 8) for the TLBO primer selection method. The number of learners is selected based on the way of trial and error. All the optimal frequency and the running time of the TLBO method are better than the GA and the MA methods except for the optimal frequency based on SantaLucia’s formula. Therefore we infer that the TLBO method for primer selection is indeed a valuable method.

(3) High-throughput for primer selection: We observe that the efficiency of the TLBO primer selection method is the best among the three primer selection methods. Many molecular biotechnologies require high-throughput analysis. The faster running time and higher optimal frequency of the TLBO primer design is very useful for large template sequences, it is especially valuable in high-throughput analysis for modern biotechnology. Consequently, the TLBO method is very suitable for high-throughput PCR primer selection.

(4) Free of algorithm-specific parameters for TLBO primer selection: In many computational intelligence-based algorithms, the algorithm-specific parameters are always the central issue applied to solve a specific problem. Too many algorithm-specific parameters are not good for solving a problem because these parameters must be considered and adjusted to correspond with the specific problem. That usually makes a process elaborate. In the parameter settings, the TLBO does not require any algorithm-specific parameters. The TLBO only uses two general parameters, that is, the number of iterations (generations) and the number of learners (population size), it is unlike the GA and the MA with the probability of crossover and the probability of mutation. Less parameters setting of the TLBO is user-friendly for the users and the researchers.

(5) Test in the wet lab needs to be performed in the future: The paper provides many descriptions and comparisons for the proposed method for primer selection. Although the method can obtain primers which conform to the primer constraints and in a relatively short running time than the other methods, it lacks the test in the wet lab. On the contrary, Primer3 has proved its use in over a decade in several thousand labs around the world. Therefore for now it is only an interesting primer selection method which needs still to pass the test in the wet lab in the future.

5 Conclusions

The computational intelligence-based methods have been applied to PCR primer selection with satisfactory results. However, the different computational methods provide distinguishable performance under the same circumstance. The ability for providing primers which conform to the primer constraints is important for an automatic primer selection. In this study, we compared the novel TLBO method to the GA and the MA for primer selection based on Wallace’s formula, Bolton and McCarthy’s formula and SantaLucia’s formula. Furthermore, we also compared the TLBO to Primer3 according to their method type, primers presentation, parameters setting, speed and memory usage.
Our results indicate that the application of the TLBO for primer selection is able to obtain desirable primer sets in a relatively short running time. That shows that it is a valuable tool for automatic high-throughput analysis for select feasible primer sets. However, the usage of the primers in the wet lab is not validated yet. Now, it is the only interesting primer selection method and needs still to pass the test in the wet lab in the future.

6 Acknowledgments

This work is partly supported by the National Science Council in Taiwan under grant NSC101-2221-E-464-001- and NSC102-2221-E-464-004-.

7 References

1 Mullis, K.B., Faloona, F.A.: ‘Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction’, Methods Enzymol., 1987, 155, pp. 335–350
2 Rozen, S., Skaltsky, H.: ‘Primer3 on the www for general users and for biologist programmers’, Methods Mol. Biol., 2000, 132, pp. 365–386
3 Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A.: ‘Primer3plus, an enhanced web interface to Primer3’, Nucleic Acids Res., 2007, 35, (Web Server issue), pp. W71–74
4 Untergasser, A., Cutcutache, I., Koressaar, T., et al.: ‘Primer3 – new capabilities and interfaces’, Nucleic Acids Res., 2012, 40, (15), pp. e115
5 Bashir, A., Liu, Y.T., Raphael, B.J., Carson, D., Bafna, V.: ‘Optimization of primer design for the detection of variable genomic lesions in cancer’, Bioinformatics, 2007, 23, (21), pp. 2807–2815
6 Koressaar, T., Joers, K., Remm, M.: ‘Automatic identification of species-specific repetitive DNA sequences and their utilization for detecting microbial organisms’, Bioinformatics, 2009, 25, (11), pp. 1349–1355
7 Kitchen, J.L., Moore, J.D., Palmer, S.A., Allaby, R.G.: ‘Meme-odpr: primer design optimization using Markov chain Monte Carlo sampling’, BMC Bioinformatics, 2012, 13, pp. 287
8 Mann, T., Humbert, R., Dorschner, M., Matamoros, J., Noble, W.S.: ‘A thermodynamic approach to pcr primer design’, Nucleic Acids Res., 2009, 37, (13), pp. e95
9 Ye, J., Coulouris, G., Zaretskaya, I., Rozen, S., Madden, T.L.: ‘Primer-blast: a tool to design target-specific primers for polymerase chain reaction’, BMC Bioinformatics, 2012, 13, pp. 134
10 Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J.: ‘Basic local alignment search tool’, J. Mol. Biol., 1990, 215, (3), pp. 403–410
11 Chuang, L.Y., Cheng, Y.H., Yang, C.H.: ‘Urdp: a specific product primer design tool’, BMC Res. Notes, 2012, 5, (1), pp. 306
12 Batniam, N., Lee, J., Lee, J., Hong, S.B., Oh, S., Han, K.: ‘Uniprimer: a web-based primer design tool for comparative analyses of primate genomes’, Comp. Funct. Genomics, 2012, 2012, pp. 520732
13 Gans, J.D., Dunbar, J., Eichorst, S.A., Gallegos-Graves, L.V., Wolinsky, M., Kuske, C.R.: ‘A robust pcr primer design platform applied to the detection of acidobacteria group 1 in soil’, Nucleic Acids Res., 2012, 40, (12), pp. e96
14 Karnik, A., Karkik, R., Grefen, C.: ‘Sdm-assist software to design site-directed mutagenesis primers introducing ‘silent’ restriction sites’, BMC Bioinformatics, 2013, 14, (1), pp. 105
15 Yang, C.H., Cheng, Y.H., Chuang, L.Y., Chang, H.W.: ‘Drug-snping: an integrated drug-based, protein interaction-based tagsnp-based pharmacogenomics platform for snp genotyping’, Bioinformatics, 2013, 29, (6), pp. 758–764
16 Chuang, L.Y., Cheng, Y.H., Yang, C.H.: ‘Specific primer design for the polymerase chain reaction’, Biotechnol Lett, 2013, 35, (10), pp. 1541–1549
17 Wu, J.S., Lee, C., Wu, C.C., Shiuie, Y.L.: ‘Primer design using genetic algorithm’, Bioinformatics, 2004, 20, (11), pp. 1710–1717
18 Yang, C.H., Cheng, Y.H., Chuang, L.Y., Chang, H.W.: ‘Specific pcr product primer design using memetic algorithms’, Bioinformat. Prog., 2009, 25, (3), pp. 745–753
19 Yang, C.H., Cheng, Y.H., Chuang, L.Y., Chang, H.W.: ‘Confronting two-pair primer design for enzyme-free snp genotyping based on a genetic algorithm’, BMC Bioinformatics, 2010, 11, pp. 509
20 Chuang, L.Y., Cheng, Y.H., Yang, C.H.: ‘Associate Per-Rflp assay design with snps based on genetic algorithm in appropriate parameters estimation’, IEEE Trans. Nanobiotechnol., 2013, 12, (2), pp. 119–127
21 Yang, C.H., Cheng, Y.H., Yang, C.H., Chuang, L.Y.: ‘Mutagenic primer design for mismatch Pcr-Rflp snp genotyping using a genetic algorithm’, IEEE/ACM Trans. Comput. Biol. Bioinform., 2012, 9, (3), pp. 837–845
22 Rao, R.V., Savsani, V.J., Vakharia, D.P.: ‘Teaching-learning-based optimization: a novel method for constrained mechanical design optimization problems’, Comput.–Aided Des., 2011, 43, (3), pp. 303–315
23 Rao, R.V., Savsani, V.J., Vakharia, D.P.: ‘Teaching-learning-based optimization: an optimization method for continuous non-linear large scale problems’, Inf. Sci., 2012, 183, (1), pp. 1–15
24 Wallace, R.B., Shaffer, J., Murphy, R.F., Bonner, J., Hirose, T., Itakura, K.: ‘Hybridization of synthetic oligodeoxirribonucleotides to Phi Chi 174 DNA: the effect of single base pair mismatch’, Nucleic Acids Res., 1979, 6, (11), pp. 3543–3557
25 Sambrook, J., Fritsch, E.F., Maniatis, T.: ‘Molecular Cloning’ (Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY, 1989)
26 Santalucia Jr, J.: ‘A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics’. Proc. Natl Acad Sci, USA, 1998, vol. 95, no 4, pp. 1460–1465