Soft Comput (2007) 11:855–863
DOI 10.1007/s00500-006-0137-8

ORIGINAL PAPER

# Primer design for multiplex PCR using a genetic algorithm

**Li-Cheng Wu · Jorng-Tzong Horng · Hsi-Yuan Huang · Feng-Mao Lin · Hsien-Da Huang · Meng-Feng Tsai**

Published online: 24 October 2006


**Abstract** Multiplex Polymerase chain reaction (PCR) is the term used when more than one pair of primers is used in a polymerase chain reaction. The goal of multiplex PCR is to amplify several segments of target DNA simultaneously and thereby to conserve template DNA, save time, and minimize expense. The success of the experiment is dependent on primer design. However, this can be a dreary task as there are many constrains such as melting temperatures, primer length, GC content and complementarity that need to be optimized to obtain a good PCR product. In our investigations, we found few primer design tools for multiplex PCR and there was no suitable tool for our partners who want to use a multiplex PCR genotypic assay. The tool draws on a genetic algorithm where stochastic approaches based on the concept of biological evolution, biological genetics and genetic operations on chromosomes are used to find an optimal solution for multiplex PCR. The presented experimental results indicate that the proposed algorithm is able to find a set of primer pairs that not only obey the design properties but also work in the same tube.

L.-C. Wu · J.-T. Horng (✉)
Institute of System Biology and Bioinformatics,
National Central University, 320 Taoyuan, Taiwan
e-mail: horng@db.csie.ncu.edu.tw

J.-T. Horng · F.-M. Lin · M.-F. Tsai
Department of Computer Science and Information
Engineering, National Central University,
320 Taoyuan, Taiwan

H.-Y. Huang · H.-D. Huang
Department of Biological Science and Technology,
Institute of Bioinformatics, National Chiao Tung University,
300 HsinChu, Taiwan

## 1 Introduction

Polymerase chain reaction (PCR) is a very powerful technique in molecular biology and is widely used today for an increasing number of applications such as in clinical diagnostics, in identification of individuals, in vitro DNA amplification and so on (Griffin and Griffin 1994). It was discovered in 1983 by Kary Mullis who was working at that time for the Cetus Corporation. The first publication of the procedure appeared in 1985 and Mullis was awarded the Nobel Prize for Chemistry in 1993.

A PCR experiment is a method for the fast and mass amplification of a specific DNA sequence. It is an iterative process, consisting of three steps: denaturation of the double-stranded DNA by heat, annealing of the oligonucleotide primers to the single-stranded target sequences, and extension of the annealing primers by a thermostable DNA polymerase. These three steps are repeated between 24 and 45 times usually in order to complete the DNA amplifying process and this is able then to generate enough sequences to allow subsequence experimental protocols. The PCR procedure is shown in Fig. 1. In addition, in reverse transcription-PCR (RT-PCR) (Burke 1996), it is used to quantify mRNA levels from much smaller samples. The process is roughly same as the PCR experiment. The additional first step in RT-PCR is the production of a single-strand complementary DNA (cDNA) of the RNA through the action of the retroviral enzyme, reverse transcriptase (Sambrook and Russell 2001).

Multiplex PCR is the term used when more than one pair of primers are included in a polymerase chain reaction. Many research and diagnostic assays involve the analysis of multiple loci. Rather than perform singleplex PCR amplification reactions for each locus, it is often

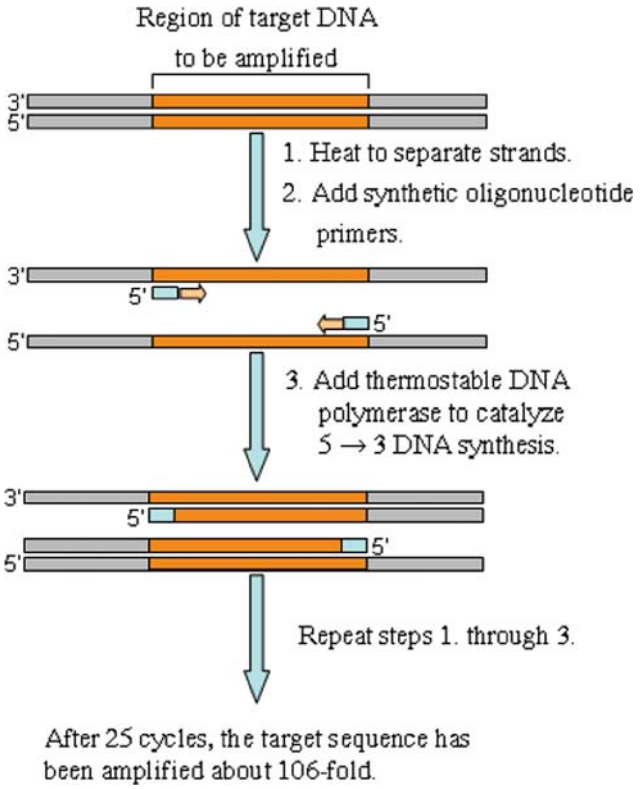


**Fig. 1** The PCR procedure

desirable to amplify all sequences of interest simultaneously in a "multiple" reaction (Innis et al. 1999). Multiplex PCR thus conserves template DNA, save time, and minimize expense. Reducing the number of tubes to which aliquots of DNA need to be added also minimizes the possibility of contamination and sample mix-up during reaction setup.

The success of the above-mentioned methods is dependent on primer design. Developing a better tool for this has become an active research issue. Therefore, various kinds of approaches and tools for the design of primers have been proposed in the last few decades. The ideal length of the primers is expected to be in a range from 18 to 30 bases. The percentage of GC content ought to be between 40 and 60% to obtain specific binding, yet also allow efficient melting during the PCR. In addition, the base distribution of the primers should be random, with polypurine and poly pyrimidine tracts being avoided (Griffin and Griffin 1994). The primers must avoid forming any self-complementary and complementary sequences. The specificity of PCR depends strongly on the melting temperature (Tm) of the primers. Generally speaking, good results are obtained when the difference in Tm within one primer pair does not exceed 5°C. PCR products that are between 100 and 2,000 base pairs long are desired. Although the above mentioned tools have satisfied various parameters, most of them do not relate to multiplex PCR (Sambrook and Russell 2001).

Multiplex PCR primer design is a great challenge. Multiple primer annealing events need to occur under the same annealing conditions without interfering with one another. In order to analyze the PCR products by electrophoresis, the products should have different lengths. In practice, it is preferable that the difference in the lengths of the PCR products is more than 50 bp. In addition, the constraints of PCR have to be satisfied.

Traditionally, the researcher finds a primer that satisfies the primer design constraints using manual primer design. However, it may take a lot of time to find a good primer and this will lead to lower accuracy and unreliable results through human error. Since experiments are expensive, and a minor mistake may cause the experiment to fail, the manual primer design method is considered to be potentially a poor approach.

To save time, and minimize problems, the use of computer programs to optimize the design, selection, and placement of oligonucleotide primers is supported (Sambrook and Russell 2001). Many different primer design programs have been developed. According to our surveys, there are approximately 70 programs available for primer design. Some of them are used on the internet, and some of them are simple stand-alone programs. In addition, some are free, while others require payment. "Primer3" (Rozen and Skaletsky 2000) picks primers from a DNA sequence, and it can avoid choosing primers in transposable elements and can pick an oligonucleotide as a probe or primers. The "Amplicon" (Jarman 2004) is a program for designing PCR primers on aligned groups of DNA sequences. The most important application for Amplicon is the design of 'group specific' PCR primer sets that amplify a DNA region from a given taxonomic group, but do not amplify orthologous regions from other taxonomic groups. The "CODEHOP" (Rose et al. 1998) was developed for PCR amplification of distantly related gene sequences. An interactive program has been written to design CODEHOP PCR primers from conserved blocks of amino acids within multiply aligned protein sequences. "ExPrimer" (Sandhu and Acharya 2005) is a web-based computer program to design primers mainly from a specified exon–exon junction (E-E-jn) of a gene of interest. The tool suggests the optimum primer-pair(s) of which the right (reverse) primer represents a particular E-E-jn of the mRNA. "Expeditor" (Hu et al. 2005), that can be used to combine known gene structure information from human and coding sequence information from farm animal species for a streamlined primer design in target farm animal species. Although these existing tools can find feasible

solutions for many problems in PCR, they do not satisfy some of the constraints, for example, specificity or multiplex PCR.

Compared with primer design for singleplex PCR, multiplex PCR primer design is difficult. In our surveys, we only found approximately six programs for multiplex PCR primer design. Most of them require payment. "MultiPLX"[1] groups PCR primers groups according to their compatibility. The program calculates the optimal combinations of primer pairs for PCR primer multiplexing. Primer compatibility is tested against each other as well as against each other's products. Additionally, the Genome Test is performed with all possible PCR primer pairs in each multiplex group. This avoids the appearance of unwanted PCR products from multiplexed groups. "Primer Primer 5"[2] is able to design primer pairs for PCR and multiplex experiments. It integrates multiple-sequence alignment with primer design to facilitate the design of primers for amplification. A proprietary algorithm is used to calculate a minority consensus that uses degenerate bases to represent all possible bases in a particular sequence position. Based upon this consensus, primers are designed in highly conserved regions of the sequences. "Fast PCR"[3] has specific, ready-to-use templates for many PCR and sequencing applications: standard and long PCR, inverse PCR, degenerate PCR directly on amino acid sequence and multiplex PCR. The program is convenient when searching homology in a personal database by local alignment and other bioinformatics tools are included.

In general, the above programs are not free, nor do they consider all of the necessary constraints. If a primer is able to anneal in several locations of a DNA template, specificity is violated. In addition, some software packages state that they can design primers for multiplex PCR, but they do not. "Primer Primer 5" is a conspicuous example. It does assist the user in selecting multiple or nested primers from a pool by ensuring that they are free of cross dimers, but multiplex PCR primer design is not supported. The main motivation for this study was the need by a group at the Taipei Veterans General Hospital, Taiwan, to carry out multiplex PCR as part of their tuberculosis research. The group was examining isoniazid resistance in *Mycobacterium tuberculosis* by single nucleotide polymorphism analysis of various genes. They wanted to use a multiplex PCR genotypic assay to increase the efficiency of their experiments. Therefore, a new tool for multiplex PCR primer design needed to be developed.

The aim was to develop a primer design tool for multiplex PCR. The user can input the multiple regions or the multiple loci of interest. The tool will return the optimal and specific groups of primers quickly according to user's requirements such as primer length, GC content, and so on. It will reduce the time and error in developing the assay. Finally, the tool is also designed with a friendly interface that will allow easy use by scientists.

In this paper, we use the genetic algorithm (GA) to design primers for multiplex PCR. Genetic algorithms were formally introduced in the United States in the 1970s by John Holland at University of Michigan. He described the "genetic algorithm", as a control structure with representations and operations that can be managed in order to evolve bit strings that were adapted to the problem to be solved. Genetic algorithms tend to converge on solutions that are globally optimal or nearly so (Davis 1987).

## 2 System and Methods

### 2.1 Unique region searching

The proposed tool contains three parts, which are shown in Fig. 2. They are "unique regions searching", "multiplex PCR primer design" and "result verification". First, we find unique regions in the target sequence. In addition, these regions need to also satisfy some parameters of primer design such as the length of primers at 18–30 nucleotides and GC content between 40 and 60%. We select candidate primers from these regions.

The length of the amplified fragment needs to be between 100 and 2,000 nucleotides and the melting temperature tolerance should be about 5°. If the primer pairs satisfy these limitations, the pairs are legal primers. In multiplex PCR primer design, we determine the legality of primer pair repetitively. In order to reduce execution time, we record the legality between candidate primer and candidate primer in advance.

Excessive regions of complementarity between primers should be avoided as they allow the formation of primer-dimers, where the primers bind to one another instead of the template (Schoske et al. 2003). Therefore, we calculate the number of matching nucleotides between primers. The matching nucleotides are one of the calculated fitness components in multiplex PCR primer design. Table 1 shows a set of primer pair interactions with the highest degree of cross reactivity.

[1] MultiPLX http://www.asperbio.com.

[2] Primer Primer 5 http://www.premierbiosoft.com/primerdesign/index.html.

[3] Fast PCR http://www.biocenter.helsinki.fi/bi/bare-1/html/oligos.htm.

**Fig. 2** System flowchart

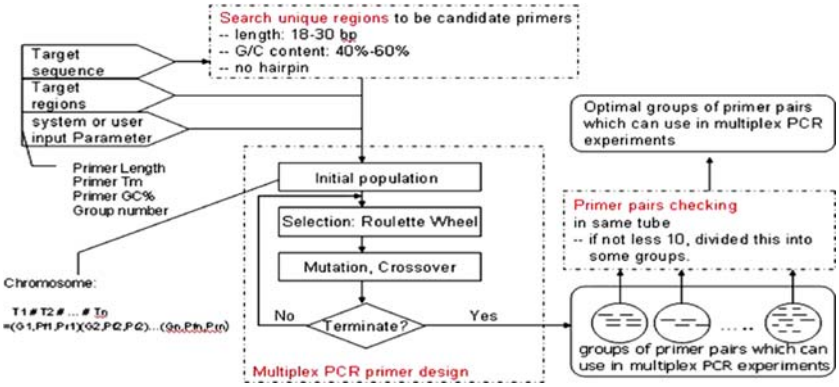


**Table 1** Cross-checking

| Sequence Information | Potential Interaction |
|---|---|
| A1 vs. A2<br>Matches = 12<br>Alignment score = 5 | 3–TAGTAGATAGACAGAGGTGGATACA–5<br>\|\|\|\|\|\|\|\|\|  \|\|<br>5–TCCATATCATCTATCCTCTGCCTA–3 |

*Alignment score which is defined as the number of complementary base pairs minus the number of mismatched base pairs between two primers

## 2.2 Multiplex PCR primer design using a genetic algorithm

In this section, we describe the main components used in our genetic algorithm, namely, crossover, mutation, and fitness.

## 2.3 Chromosome

Each chromosome is one of the solutions of the multiplex PCR primer design. It was defined as serial integers and the number of integers is three times the number of target regions. Every three contiguous integers, including the number of the forward primer, the number of the reverse primer and experimental tube number, are presented to amplify one target region. For example, five target regions cost 15 integers in chromosome.

**Definition 2.3.1** (*Chromosome*) *A chromosome is composed of target regions* $T_1, T_2, \ldots, T_n$*, denoted as* $T_1\#T_2\#\cdots\#T_n$*, can be represented by*

$$T_1\#T_2\#\cdots\#T_n = (G_1, P_{f1}, P_{r1})(G_2, P_{f2}, P_{r2})\cdots(G_n, P_{fn}, P_{rn})$$

where

- $n$ is the number of target region;
- $T_i$ is the $i$th target region of genome;
- the symbol # is represented for concatenation;
- $G_i$ is the tube of multiplex PCR for the $i$th target region;
- $P_{fi}$ is the number of forward primer for the $i$th target region;
- $P_{ri}$ is the number of reverse primer for the $i$th target region;

$1 \leqq i \leqq n$

Figure 3 shows that there are five target regions we want to amplify. Among the chromosome, the 21st primer and the 54th primer can amplify the 1st target region and the 4th primer and the 23rd primer can amplify the 4th target region. These primers both act in the 3rd tube.

### *2.3.1 GA process flow*

The system processing flow is depicted in Algorithm 1. The symbol $|P|$ represents the size of the population. The concept of the system process flow is based on the architecture of a simple genetic algorithm (Goldberg 1989). In each chromosome, the length of amplified fragment should be between 100 and 2,000 bases and the difference in melting temperature between each primer

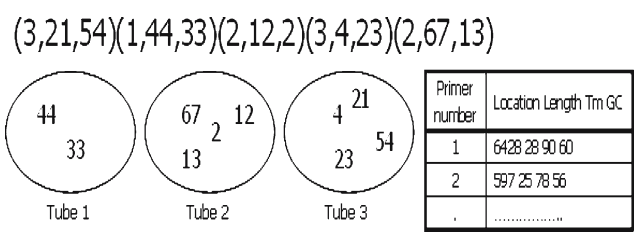


**Fig. 3** Example of chromosome

**Algorithm 1** The Flow of Our Approach

```
Generate the initial population P
Let P_m be the probability of mutation; P_e be the probability of crossover;
while not satisfy the termination condition do
    for i = 1 to |P| do
        Select two chromosomes X and Y from population by the Roulette Wheel method
        Let X' = X and Y' = Y
        Mutation( X' )
        Mutation( Y' )
        Crossover( X' , Y' )
        Add X' and Y' into mating pool
    end for
    Select the best top |P| chromosomes to replace the original population
end while
```

should not exceed 5° for any target regions. In addition, the primer pairs in same tube should have a similar melting temperature.

Elitism is used. This means, that at least one of the best solutions in each generation is preserved without change to a new population. Therefore, the best solution can survive to the succeeding generation. The crossover and mutation operators are repeatedly applied until the termination conditions are satisfied. The following are termination conditions:

- The number of generations exceeds the maximum number of generations permitted.
- The best fitness does not improve over a given number of generations. The default is 500 generations.

### *2.3.2 Fitness*

To evaluate their fitness, the chromosome must be applied to a sum-of-pairs function (Setubal and Meidanis 1997). The sum-of-pairs function is defined as the sum of the scores of all primer pairs in same tube. If the difference of length of amplified fragment is less 50 bases and the number of complementary sequences is not zero, then the fitness receives a lower score. The fitness values of chromosomes are recomputed after the mutation and crossover process.

### *2.3.3 Crossover operator*

The purpose of crossover is to exchange information from the chromosomes to produce offspring, which it is hoped will possess an advantage over the parental generation. However, the offspring also can inherit a disadvantage from the parental generation. The crossover does not promise to produce good offspring. Then, based on the principle of survival-of-the-fittest, the worse offspring are eliminated by competition.

In the crossover process, two parent chromosomes, denoted as $X$ and $Y$, are selected by Roulette Wheel Selection and are used to produce two daughter chromosomes, denoted as $X'$ and $Y'$. The common cutting point is randomly selected in parent chromosomes. It will cut every chromosome into two parts, called the longer part and the shorter part. We reserve the longer part and exchange the shorter part. The identifier of tube must be reassigned. The assignment order is the same $T_m$ of the group, no member of group, and new group in turn. Algorithm 2 describes, in detail, the crossover operator 1.

An example of crossover is shown in Fig. 4. Two parent chromosomes $X$ and $Y$ are used to produce two daughter chromosomes $X'$ and $Y'$. When the cutting point is selected, we exchange the shorter part and reassign the group. In the chromosome $Y'$, the melting temperature of the 12th primer and the 2th primer is 66°C.

**Algorithm 2** Crossover(chromosome $X$ , chromosome $Y$)

Randomly select a cutting point $L$, $L$ % 3 = 0, $1 < L < 3n$

**if** $L \geq 3n - L$ **then**

Exchange the rear-end part of two chromosomes and Reassign groups.

In reassigning process, we select the group which has the same $T_m$ with the primer pair first. If the group can't be found, the empty groups will be considered. We can select one from them. When the above-mentioned methods can't work, we will assign a new group until the maximum number of group is reached.

**else if** $L < 3n - L$ **then**

Exchange the front-end part of two chromosomes and Reassign groups.

The reassigning process is same with the above-mentioned process.

**end if**

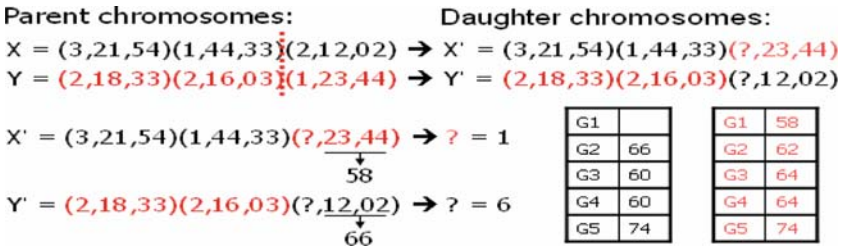


**Fig. 4** Crossover operator example

According to the assignment order, we finally assign a new group to it.

### 2.3.4 Mutation operator

Evolution can't produce novel individuals by crossover and reproduction alone because the offspring only mix the properties of parents. Evolution under these circumstances only moves ahead slowly and is limited to a small group of individuals. In nature, organisms use mutation to create new variants. Therefore, we made use of a mutation operator to increase the diversity of population and this allows evolution to act fully through diversification.

There are two kinds of mutation operators in our approach. In the mutation process, each selected chromosome is mutated by randomly using one of the following mutation operators.

The mutation operator 1 chooses one of integers to change. After mutating, we test if the $T_m$ of corresponding target region is the same as the original. If the mutation point is the group, we use another group to replace it. First, we find the candidate groups, the $T_m$ of which is the same as the group that we want to change. We randomly choose one of candidate groups to replace the mutation group. This action doesn't change the $T_m$ of target region. It is aimed to let the primer pair change their reaction environment. If the mutation point is the primer, we randomly choose another primer that must obey the constraints of primer design. Algorithm 3 describes the detail of mutation operator 1.

An example is given in Fig. 5. Figure 5a shows mutation operator 1 where it chooses a group to change. After deciding the mutation point, we identify the other

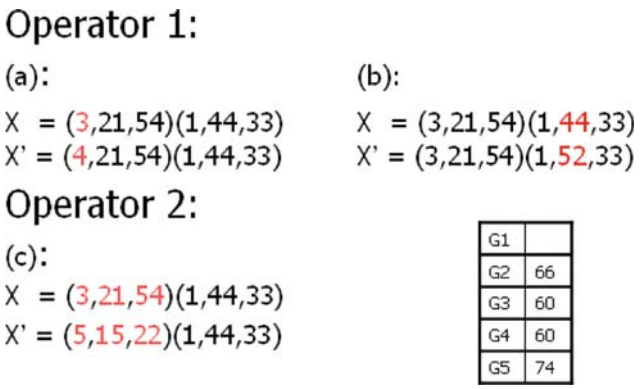


**Fig. 5** Mutation operator example

groups whose $T_m$ is 60°C as the candidate groups. Then, we randomly select one group from them to replace the original group. Unfortunately there are no candidate groups, so we assign a new group or mutate another location. In the same way, mutation operator 1 can choose a primer to change as depicted in Fig. 5b.

Mutation operator 2 whose detail is depicted in Algorithm 4 replaces group and primer at the same time. We choose one target region to rearrange the group and primer pairs. This action must obey the constraints of primer design.

### 2.3.5 Result verification

Finally, we check the solutions again to optimize the number of primer pairs in same tube and to avoid non-specific PCR amplicons.

**Algorithm 3** Mutation operator 1(chromosome X)

Randomly select a mutation point $L$, $0 \leqq L \leqq 3n$

**If** $L$ % 3 = 0 **then**
    Find groups whose $T_m$ is the same with the group we want to change
    Randomly select one to alter the origin
**else**
    Randomly choose the other primer which must obey the constraints of primer design to alter the origin
**end if**

**Algorithm 4** Mutation operator 2(chromosome $X$)

Randomly select a mutation point $L$, $0 < L < 3n$, $L \% 3 = 0$
Randomly select a group $G$
According to the $T_m$ of the group $G$, find the primer pair $P_r$ and $P_f$ for target region $T_L$
Replace the $G$, $P_r$, and $P_f$ to $L$, $L + 1$, and $L + 2$ respectively

**Table 2** The result of case study I

| Target | Region | Forward primer | | | | Reverse primer | | | | Product size |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Start | Len | Tm | GC | Start | Len | Tm | GC | |
| Group 0 → Tm: 86 | | | | | | | | | | |
| katG (110) | 2155779~2155782 | 2154867 | 25 | 86 | 59 | 2156566 | 25 | 90 | 55 | 1672 |
| kasA (77) | 2518341~2518344 | 2517673 | 25 | 88 | 57 | 2519532 | 25 | 86 | 59 | 1831 |
| kasA (413) | 2519349~2519352 | 2519283 | 25 | 90 | 50 | 2519532 | 25 | 86 | 59 | 219 |
| Group 1 → Tm: 62 | | | | | | | | | | |
| inhA (194) | 1674779~1674782 | 1674046 | 25 | 64 | 52 | 1675177 | 25 | 62 | 40 | 1110 |
| katG (463) | 2154720~2154723 | 2154062 | 25 | 64 | 60 | 2155953 | 25 | 62 | 55 | 1871 |
| katG (90) | 2155839~2155842 | 2155454 | 25 | 66 | 57 | 2155953 | 25 | 62 | 55 | 478 |
| ahpC (176) | 2726341~2726344 | 2725564 | 25 | 62 | 55 | 2726502 | 25 | 62 | 55 | 918 |
| Group 2 → Tm: 80 | | | | | | | | | | |
| ndh (110) | 2102710~2102713 | 2101383 | 25 | 82 | 51 | 2102976 | 25 | 80 | 48 | 1566 |
| ahpC (174) | 2726716~2726719 | 2725630 | 25 | 82 | 57 | 2727461 | 25 | 80 | 60 | 1805 |
| Group 3 → Tm: 76 | | | | | | | | | | |
| kasA (312) | 2519046~2519049 | 2517688 | 25 | 76 | 58 | 2519273 | 25 | 80 | 60 | 1561 |
| ahpC (61) | 2726371~2726374 | 2726304 | 25 | 76 | 58 | 2727303 | 25 | 76 | 58 | 975 |
| ahpC (51) | 2726476~2726479 | 2726304 | 25 | 76 | 58 | 2727461 | 25 | 80 | 60 | 1133 |
| Group 4 → Tm: 86 | | | | | | | | | | |
| katG (434) | 2154807~2154810 | 2154691 | 25 | 88 | 51 | 2154867 | 25 | 86 | 59 | 147 |
| kasA (269) | 2518917~2518920 | 2517673 | 25 | 88 | 57 | 2519532 | 25 | 86 | 59 | 1831 |
| Group 5 → Tm: 64 | | | | | | | | | | |
| inhA (21) | 1674260~1674263 | 1674046 | 25 | 64 | 52 | 1675157 | 25 | 64 | 60 | 1090 |
| katG (397) | 2154918~2154921 | 2154062 | 25 | 64 | 60 | 2156068 | 25 | 68 | 54 | 1986 |
| katG (315) | 2155164~2155167 | 2154062 | 25 | 64 | 60 | 2155454 | 25 | 66 | 57 | 1372 |
| katG (91) | 2155836~2155839 | 2155454 | 25 | 66 | 57 | 2157329 | 25 | 64 | 52 | 1854 |
| Group 6 → Tm: 78 | | | | | | | | | | |
| katG (529) | 2154522~2154525 | 2154302 | 25 | 82 | 57 | 2154727 | 25 | 78 | 56 | 399 |
| kasA (66) | 2518308~2518311 | 2517678 | 25 | 78 | 56 | 2519273 | 25 | 80 | 60 | 1570 |
| Group 7 → Tm: 70 | | | | | | | | | | |
| ndh (268) | 2102236~2102239 | 2101770 | 25 | 70 | 52 | 2102708 | 25 | 70 | 59 | 915 |
| katG (438) | 2154795~2154798 | 2154224 | 25 | 70 | 59 | 2155514 | 25 | 70 | 59 | 1268 |
| katG (336) | 2155101~2155104 | 2153914 | 25 | 74 | 60 | 2155514 | 25 | 70 | 59 | 1577 |
| katG (138) | 2155695~2155698 | 2155514 | 25 | 70 | 59 | 2157283 | 25 | 72 | 56 | 1747 |

## 3 Results

The tool runs on a PC with an AMD K7-1200 Mhz CPU, 750 MB RAM and an OS consisting of a Linux 9.0 platform. It is written by C++.

### 3.1 Case study I

Isoniazid (INH) is a central component of drug regimens used worldwide to treat tuberculosis. Previous studies show that a variety of single nucleotide polymorphisms in multiple genes are found exclusively in INH-resistant clinical isolates. These genes are either involved in mycolic acid biosynthesis or are overexpressed as a response to the buildup or cellular toxicity of INH (Ramaswamy et al. 2003). Up to the present, 24 polymorphisms have been published. These target regions and the *Mycobacterium tuberculosis* genome are the inputs for this case study. The default parameters are used. Table 2 shows the detailed results for this case. In the target region column, we display the start and end locations of every target region. The detail primer information includes start location, length, melting temperature, GC content and the length of amplified fragment for any primer pair. These are shown in the second and third column. The constraints for singleplex PCR and

**Table 3** The result of case study II

| Target | Region | Forward primer | | | | Reverse primer | | | | Product size |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Start | Len | Tm | GC | Start | Len | Tm | GC | |
| Group 0 → Tm: 72 | | | | | | | | | | |
| ybbH | 191182∼192033 | 191032 | 25 | 74 | 48 | 192279 | 25 | 72 | 44 | 1222 |
| ybdE | 219594∼220019 | 218454 | 25 | 72 | 44 | 220067 | 25 | 74 | 48 | 1588 |
| mdr | 332260∼333540 | 331789 | 25 | 72 | 44 | 333796 | 25 | 76 | 52 | 1982 |
| Group 1 → Tm:76 | | | | | | | | | | |
| gcaD | 56350∼57720 | 55905 | 25 | 78 | 56 | 57928 | 25 | 76 | 52 | 1998 |
| rpmC | 139922∼140122 | 138521 | 25 | 76 | 52 | 140483 | 25 | 76 | 52 | 1937 |
| Group2 → Tm: 74 | | | | | | | | | | |
| ksgA | 50638∼51516 | 50092 | 25 | 74 | 48 | 51979 | 25 | 76 | 52 | 1862 |
| yazB | 87398∼87607 | 86124 | 25 | 74 | 48 | 87862 | 25 | 76 | 52 | 1713 |
| adaB | 204337∼204876 | 204227 | 25 | 74 | 48 | 205572 | 25 | 76 | 52 | 1320 |
| Group 3 → Tm:72 | | | | | | | | | | |
| holB | 40663∼41652 | 39923 | 25 | 72 | 44 | 41838 | 25 | 76 | 52 | 1890 |
| gltX | 111044∼112495 | 110748 | 25 | 74 | 48 | 112512 | 25 | 72 | 44 | 1739 |
| Group 4 → Tm: 72 | | | | | | | | | | |
| ybxB | 121065∼121670 | 120125 | 25 | 74 | 48 | 121888 | 25 | 72 | 44 | 1738 |
| ycbD | 268838∼270304 | 268474 | 25 | 76 | 52 | 270466 | 25 | 72 | 44 | 1967 |
| Group 5 → Tm: 74 | | | | | | | | | | |
| phoD | 283555∼285225 | 283286 | 25 | 76 | 52 | 285243 | 25 | 74 | 48 | 1932 |
| yceA | 309554∼310396 | 308841 | 25 | 78 | 56 | 310623 | 25 | 74 | 48 | 1757 |
| Group 6 → Tm: 78 | | | | | | | | | | |
| dnaX | 26812∼28503 | 26659 | 25 | 78 | 56 | 28636 | 25 | 78 | 56 | 1952 |
| rpsI | 154299∼154691 | 153407 | 25 | 78 | 56 | 154895 | 25 | 78 | 56 | 1463 |
| Group 7 → Tm: 70 | | | | | | | | | | |
| mpr | 245179∼246120 | 244896 | 25 | 74 | 48 | 246483 | 25 | 70 | 40 | 1562 |
| ycbA | 265537∼266700 | 265448 | 25 | 74 | 48 | 266905 | 25 | 70 | 40 | 1432 |

multiplex PCR are satisfied. For example, the difference in product size between each target is at least 50 bases in same tube. We also observe other phenomena. Some primers can be shared and some primer pairs can amplify regions that include many targets. These are due to the short distance between some target regions.

### 3.2 Case study II

We select *Bacillus subtilis* genome as our material and choose 30 genes as target regions. The designated primer length was 25 nucleotides. The result also satisfies the constraints of singleplex PCR and multiplex PCR as shown in Table 3. In this experiment, we find that there are no legal primer pairs close to certain the target regions, thus no primers for these genes are generated. Our approach filters this in advance.

## 4 Discussion and Conclusion

We present a novel tool to design primers for multiplex PCR. It can design primers for target regions and group the primer pairs to achieve the purpose of multiplex PCR primer design. We were able to find unique regions in the target sequence and select candidate primers from these. We use this method to avoid primers annealing in several locations. Most programs don't pay attention to the area of specificity.

Our tool uses "unique regions searching" to find candidate primers. This action ensures uniqueness of the primers. Primers will not anneal in several locations of a template. However, there is a defect in this method. If there is no unique region close to the target regions, we will fail to find primer pairs for these regions. We need new approaches to solve this problem. Additionally, we need to consider further constraints, such as the fact that the base at the 3′ end of each primer should be G or C. Such changes, we believe, will increase the accuracy of any experiments carried out after primer design for multiplex PCR using the program outlined here.

## References

Burke JF (1996) PCR: essential techniques.
Davis L (1987) Genetic algorithms and simulated annealing. In: 1
Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. Addison-Wesley, New York
Griffin HG, Griffin AM (1994) PCR technology: current innovations

Hu ZL, Glenn K, Ramos AM, Otieno CJ, Reecy JM, Rothschild MF (2005) Expeditor: a pipeline for designing primers using human gene structure and livestock animal EST information. J Hered 96:80–82

Innis MA, Gelfand DH, Sninsky JJ (1999) PCR applications: protocols for functional genomics

Jarman SN (2004) Amplicon: software for designing PCR primers on aligned DNA sequences. Bioinformatics, pp. bth121

Ramaswamy SV, Reich R, Dou S-J, Jasperse L, Pan X, Wanger A, Quitugua T, Graviss EA (2003) Single nucleotide polymorphisms in genes associated with isoniazid resistance in mycobacterium tuberculosis. Antimicrob Agents Chemother 47:1241–1250

Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. Nucleic Acids Res 26:1628–1635

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386

Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual, 3 edn

Sandhu KS, Acharya KK (2005) ExPrimer: to design primers from exon–exon junctions. Bioinformatics 21:2091–2092

Schoske R, Vallone PM, Ruitberg CM, Butler JM (2003) Multiplex PCR design strategy used for the simultaneous amplification of 10 Y chromosome short tandem repeat (STR) loci. Anal Bioanal Chem 375:333–343

Setubal J, Meidanis J (1997) Sequence comparison and database search. In: Introduction to computational molecular biology