Primer Design for Multiplex PCR Using a Genetic Algorithm

Feng-Mao Lin[1] meta@db.csie.ncu.edu.tw
Hsien-Da Huang[2] bryan@mail.nctu.edu.tw
His-Yuan Huang[2] aliken.bi93g@nctu.edu.tw
Jorng-Tzong Horng[1,*] horng@db.csie.ncu.edu.tw

[1]Department of Computer Science and Information Engineering, National Central University
No. 300, Jhongda Rd. Jhong-Li, Taiwan
+886-3-4227151, ext. 35307

[*]Correspondence Author

[2]Department of Biological Science and Technology, National Chiao Tung University
75 Bo-Ai St., Hsin-Chu, Taiwan
+886-3-5712121, ext. 56952

ABSTRACT
Multiplex Polymerase Chain Reaction (PCR) experiments are used for amplifying several segments of the target DNA simultaneously and thereby to conserve template DNA, reduce the experimental time, and minimize the experimental 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. Motivated by the lack of primer design tools for multiplex PCR genotypic assay, we propose a multiplex PCR primer design tool using a genetic algorithm, which is a stochastic approach based on the concept of biological evolution, biological genetics and genetic operations on chromosomes, to find an optimal selection of primer pairs for multiplex PCR experiments. The presented experimental results indicate that the proposed algorithm is capable of finding a series of primer pairs that obeies the design properties in the same tube.

Categories and Subject Descriptors
J.3 [Computer Application]: Life and medical sciences – *biology and genetics.*

General Terms
Algorithms, Design

Keywords
Genetic algorithm, Multiplex PCR, Primer

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 [3].

GECCO'05, June 25-29, 2005, Washington, DC, USA.
ACM 1-59593-010-8/05/0006.

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 desirable to amplify all sequences of interest simultaneously in a "multiple" reaction [4]. 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.

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 [1].

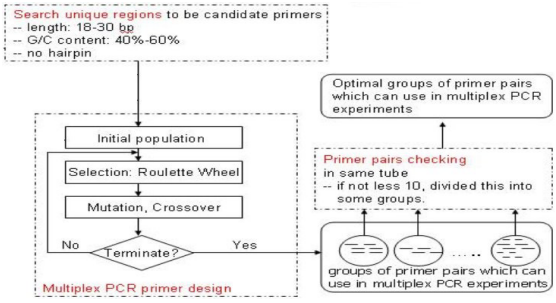

Fig. 1. System flowchart

2. SYSTEM AND METHOD

2.1 Unique Region Searching
The proposed tool contains three parts, which are shown in Fig. 1. The length of the amplified fragment needs to be between 100 and 2000 nucleotides and the melting temperature tolerance should be about five degrees. 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 [6]. 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.

2.2 Multiplex PCR Primer Design Using a Genetic Algorithm

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.

The concept of the system process flow is based on the architecture of a simple genetic algorithm [2]. 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.

To evaluate their fitness, the chromosome must be applied to a sum-of-pairs function [7]. 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.

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 Tm of the group, no member of group, and new group in turn.

There are two kinds of mutation operators in our approach. The mutation operator 1 chooses one of integers to change. After mutating, we test if the Tm 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. Mutation operator 2 chooses one target region to rearrange the group and primer pairs. This action must obey the constraints of primer design.

3. RESULTS

We show the one of the experiments. 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 [5]. Up to the present, twenty-four polymorphisms have been published. These target regions and the *Mycobacterium tuberculosis* genome are the inputs for this case study. The default parameters are used. The constraints for singleplex PCR and multiplex PCR are satisfied.

4. 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.

5. References

[1] Davis, L., Genetic algorithms and simulated annealing, in *1*. 1987.

[2] Goldberg, D.E., Genetic Algorithms in Search, Optimization, and Machine Learning. 1989, New York: Addison-Wesley.

[3] Griffin, H.G. and A.M. Griffin, PCR technology: current innovations. 1994. 6.

[4] Innis, M.A., D.H. Gelfand, and J.J. Sninsky, PCR applications: protocols for functional genomics. 1999.

[5] Ramaswamy, S.V., et al., Single Nucleotide Polymorphisms in Genes Associated with Isoniazid Resistance in Mycobacterium tuberculosis. *Antimicrob. Agents Chemother.*, 47, 4, (2003), 1241-1250.

[6] Schoske, R., et al., Multiplex PCR design strategy used for the simultaneous amplification of 10 Y chromosome short tandem repeat (STR) loci. *Anal Bioanal Chem*, 375, 3, (2003), 333-43.

[7] Setubal, J. and J. Meidanis, Sequence comparison and database search. In Introduction To Computational Molecular Biology. 1997. 47-103.