A Greedy Algorithm for Minimizing the Number of Primers in Multiple PCR Experiments

Koichiro Doi doi@is.s.u-tokyo.ac.jp
Hiroshi Imai imai@is.s.u-tokyo.ac.jp
Department of Information Science, Faculty of Science, University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Abstract

The selection of a suitable set of primers is very important for polymerase chain reaction (PCR) experiments. Most existing algorithms for primer selection are concerned with producing a primer pair for each DNA sequence. However, when all the DNA sequences of the target objects are already known, like the approximately 6,000 yeast ORFs, we may want to design a small set of primers to PCR amplify all the targets, which can then be resolved electrophoretically in a series of experiments. This would be quite useful, because decreasing the number of primers greatly reduces the cost of an experiment.

This paper extends the problem of primer selection for a single experiment presented in Doi and Imai [3] to primer selection for multiple PCR experiments, and proposes algorithms for the extended problem. The algorithms design primer sets one at a time. We extend the greedy algorithm for one PCR experiment in [3] by handling amplified segments in DNA sequences that have been identified by primer pairs already selected and by changing the priorities in the greedy algorithm. This algorithm is applied to real yeast data. The number of primers equaled 85% of the number of identified DNA sequences, which represented more than 90% of all the target DNA sequences. This is 42% the number of primers needed for multiplex PCR. Furthermore, the length of each primer is less than half the length of multiplex PCR primers so the cost of producing the primers is reduced to 20% of the cost in the multiplex PCR case.

1 Introduction

The polymerase chain reaction (PCR) amplifies DNA sequences efficiently and has many applications. Many more will be developed as more DNA sequence data becomes available. PCR applications include genome typing and DNA fingerprinting. As these require PCR amplification of many different targets, amplifying the targets in groups can greatly reduce experimental costs. Such savings can be achieved by designing the correct primers for each PCR experiment.

We explain the background and requirements of this problem in the introduction, and then present our results.

PCR PCR amplifies the DNA sequence between a pair of primers, one called the sense primer and the other the antisense primer, that have an orientation. PCR is usually used to amplify a particular region in a target sequence. Therefore, a primer pair that matches one portion of the target sequence is selected. If these primers match many positions in the DNA genome, many subsequences may be amplified. The primers used for the human genome are usually > 17 mers, because all the random permutations of 17 mers (4^{17} is approximately 1.6×10^{10}) exceeds the size of the human genome (3×10^9 bp), and there should be few random pairings. The cost of a primer is proportional to its length.

PCR with Many Primers As already stated, it is often necessary to amplify many different target objects and many methods have been proposed for doing this. Multiplex PCR amplifies multiple genes using multiple sets of primers in a single reaction (Chamberlain, Gibbs, Ranier, Nguyen and Caskey [1]). In multiplex PCR, each pair of primers is designed for one amplified region and all the primers are distinct from each other. In this multiplex PCR, the primer length is set between 17 and 20 and the number of primers is exactly twice the number of target objects. Welsh and McClelland [10] proposed a method called Arbitrarily Primed PCR (AP-PCR), which uses short (8 to 12 nucleotides) arbitrary primers. These short primers match several targets and produce more fragments, so the number of primers necessary to amplify all the targets may be much smaller than in multiplex PCR. AP-PCR is applicable when no prior sequence information is required. In both these cases, the PCR products are visualized by agarose gel electrophoresis and utilized for DNA fingerprinting, etc. The number of applications is really enormous. For example, Ito and Sakaki [4] proposed Fluorescent Differential Display (FDD) using arbitrarily primed RT-PCR.

Primer Selection: Necessity of Reducing the Number of Primers Now that the entire genomes of yeast, *E. coli*, and other organisms have been sequenced, the application of AP-PCR may be extended further. If the AP-PCR primers are replaced with primers obtained from the known sequence data, the predicted and experimental PCR products can be compared to identify missing sequences. Thus, primer design is becoming more important, especially in designing primers for multiple targets when there are repeated amplifications. If the minimum set of primers required to amplify all the target sequences in a distinguishable way can be obtained, the experimental costs will be decreased. For example, yeast has about 6,000 ORFs. With current electrophoresis technology, this would provide too many targets to obtain distinguishable amplified regions of different lengths for all the ORFs in a single PCR experiment. Doi [2] and Doi and Imai [3] considered minimizing the number of primers for a single PCR experiment. This paper designs primer sets for a series of PCR experiments by extending the algorithms in [3].

Obtaining Gene Expression Patterns In recent years, many gene expression patterns have been determined by using oligonucleotide chips and DNA microarrays. Many studies make use of the data produced by these methods. However, the data produced by these technologies are relatively imprecise, and may not be applicable when greater precision is needed. AP-PCR and its extensions can produce more precise data, and the primer sets generated by our method are valuable for obtaining data when only a small margin of error is permitted.

Related Work on Primer Selection Some computer programs for primer selection have been designed ([6, 9]). These programs are used to design pairs of primers for a single target gene and to find pairs obeying given biological constraints and display information about the amplified product. Nicodème and Steyaert [7] designed a program for multiplex PCR. As mentioned above, the number of primers in multiplex PCR is fixed at double the number of targets. Their program designs sets of primers that amplify regions of all the target genes, so as to minimize the number of amplifications. This is the main function of our proposed algorithm. The lengths of the amplified regions differ in each amplification in a series.

Pearson *et al.* [8] formulated and analyzed a minimum primer set that covers all DNA sequences. This is a much simpler version of the method in [2, 3]. Their paper does not consider primer orientation. From the viewpoint of computational complexity, they analyze the formulated problem, prove the approximability of the problem by the direct use of set cover reduction, and construct an exact branch and bound algorithm and an approximation algorithm with a $\ln n$ approximation ratio, where n is the number of DNA sequences.

Doi and Imai [2, 3] formulated primer selection problems that required minimizing the number of primers in a single PCR experiment meeting some biological conditions and proposed algorithms to

produce primers.

Our Results This paper extends the primer selection problems in [3] to primer selection for multiple PCR amplifications, and proposes algorithms for the extended problem. We explain the primer selection problem for a single amplification, which was presented in [3], and extend the problem to primer selection for multiple PCR amplifications. We propose algorithms for the extended problem in section 3. The algorithms design primer sets one at a time using the greedy algorithm for a single PCR amplification presented in [3].

In section 4, we extend the greedy algorithm and apply the extended algorithm to real yeast data. First, we controlled which segments of identified DNA sequences were amplified by using previously selected primer sets. However, the experimental results using the extended algorithm were not very good. Therefore, we modified the priority order the greedy algorithm uses for a single PCR experiment and weighted the DNA sequences. This produced a better solution. The number of primers equaled 85% the number of identified DNA sequences, which represented more than 90% of all the target DNA sequences. This is 42% the number of primers that would be needed in multiplex PCR. Furthermore, the primers were less than half as long as the multiplex primers, reducing the cost of experiments to 20% of the cost in the case of multiplex PCR.

We evaluated the predicted results for the yeast data. The primer sets produced by the greedy algorithm identified fewer DNA sequences as the series of amplifications proceeded. Ultimately, more primers were generated than in the case of multiplex PCR, as indicated by a number of newly identified DNA sequences. This result was improved using a simple algorithm.

2 Biological Constraints and Problem Formulation

2.1 Biological Constraints for PCR Primer Selection

This paper considered the following biological constraints on primers:

GC content Primers with a 40-60% GC content are widely used.

Primer length We use short (8-12) primers so that each primer can match multiple positions in the DNA genome.

Complementary sequences The set of primers must not contain any self-complementary or complementary sequences. (For example, 5'-GCCTAGGC-3' is a self-complementary sequence, 5'-GACAATGC-3' and 5'-GCATTGTC-3' are complementary sequences.)

Length of amplified regions PCR products that are between 50 and 500 base pairs long are desired. We considered the lower limit for the amplified regions.

Different lengths of amplified segments 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 5.)

Details of the biological constraints for designing primers are described in [5]. We will consider biological constraints in another paper.

2.2 Formulations and Properties

Doi [2] formulated some problems for various biological constraints, and demonstrated their complexity. In this subsection, we explain these results and formulate the primer selection problem for multiple PCR experiments.

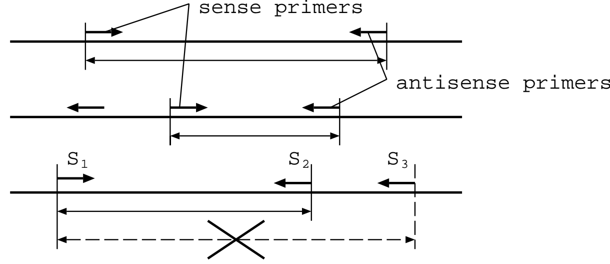

Figure 1: Amplification Constraint.

Given n DNA sequences, the length of DNA sequence j is denoted by m_j $(1 \leq j \leq n)$. The position p of DNA sequence j, is indicated by (j,p). Primer S_i is a DNA sequence with a constant length that is much shorter than m_j. $|S_i|$ denotes the length of primer S_i. Consider a primer set $S = \{S_1, ..., S_i, ..., S_l\}$. $(j, p, \mathrm{r/l}) \in S_i$ means that primer i matches position p of DNA sequence j (where the direction is r(right) or l(left)). That is, for $k = 1$ to $|S_i|$ the $(p+k-1)$th character (resp. the $(p-k+1)$th character) of sequence j is identical to the kth character of primer i in the case of r (resp. l).

Definition 1 (Primer Selection Problem (PSP)) *The Primer Selection Problem (PSP) involves finding the minimum primer subset $S' \subseteq S$ that includes all the desired DNA sequences, and putting a DNA subsequence between sense and antisense primers for each DNA sequence. That is,*

$$\textit{minimize } |S'| \textit{ s.t. } \forall j\ \exists S_i, S_{i'} \in S',\ (j, p_1, \mathrm{r}) \in S_i, (j, p_2, \mathrm{l}) \in S_{i'},\ p_2 - p_1 + 1 > 0.$$

(The length of the amplified segment is $p_2 - p_1 + 1$.)

Theorem 1 ([2]) *The PSP has a polynomial approximation greedy algorithm with a factor $\ln mn$ $(m = \max(m_1, ..., m_n))$.*

Theorem 2 ([2]) *The PSP cannot be approximated in polynomial time within a factor $(1-\epsilon)\ln n$ for any $\epsilon > 0$ unless $NP \subset TIME(n^{O(\log\log n)})$.*

To consider the length of DNA amplified, we must impose the next constraint.

Definition 2 (Amplification Constraint between Sense and Antisense Primers) *If a DNA sequence is amplified by primers that match position p_1 in direction* r *and position p_2 in direction* l*, primers that match position p $(p_1 \leq p \leq p_2)$ of DNA sequence j cannot exist in this PCR experiment.*

In Fig. 1, the sequence between S_1 and S_2 is amplified, while the sequence between S_1 and S_3 is not amplified. In practice the latter sequence may be amplified slightly, but it can be neglected as the former sequence predominates.

Definition 3 (Primer Selection Problem with Minimum and Distinguishable Length Constraints (PSPMDLC)) *The Primer Selection Problem with Minimum and Distinguishable Length Constraints requires finding the minimum primer subset $S' \subseteq S$ that includes all the target DNA sequences, as in the PSP, so that the amplified DNA subsequence has a length of least k, and the difference in the lengths of the amplified DNA subsequences is at least d. That is,*
minimize $|S'|$
s.t.

$\forall j\ \exists S_i, S_{i'} \in S',\ \exists (j, p_1, \mathrm{r}) \in S_i, (j, p_2, \mathrm{l}) \in S_{i'},$

1. $p_2 - p_1 + 1 \geq k,$

2. *$\forall p'$ $(p_1 \leq p' \leq p_2)$ $(j, p', \mathrm{r}), (j, p', \mathrm{l})$ are not in sets (which are regarded as primers) which are elements of primer set S' (amplification constraint).*

3. *$\forall j' (\neq j)$ If $(j', p_3, \mathrm{r}) \in S_i, (j', p_4, \mathrm{l})$ are in sets (which are regarded as primers) which are elements of primer set S' and satisfy the amplification constraint (2), $|(p_4 - p_3 + 1) - (p_2 - p_1 + 1)| \geq d$ is also satisfied.*

In the PSPMDLC, we consider the minimum length and the difference in the lengths of the amplified segments. Then we extend the PSPMDLC to the primer selection problem for multiple PCR experiments. The PSPMDLC sets constraints for each amplification, but a single PCR does not amplify all the target DNA sequences. However, there is an amplification for each DNA sequence.

S^e is the set of primer pairs.

Definition 4 (Primer Selection Problem for Multiple PCR Experiments (PSPME)) *The primer Selection Problem for Multiple PCR Experiments minimizes $\sum_e |S^e|$*
s.t.

$\forall j \exists e \exists S_i, S_{i'} \in S^e,\ (j, p_1, \mathrm{r}) \in S_i, (j, p_2, \mathrm{l}) \in S_{i'},$

1. *$p_2 - p_1 + 1 \geq k$.*

2. *$\forall p' (p_1 \leq p' \leq p_2)$ $\forall S_{i''} \in S^e$ $(j, p', \mathrm{r}), (j, p', \mathrm{l}) \notin S_{i''}$ (amplification constraint).*

3. *$\forall j' (\neq j)$ If $(j', p_3, \mathrm{r}), (j', p_4, \mathrm{l}) \in S^e$ satisfies the amplification constraint, $|(p_4 - p_3 + 1) - (p_2 - p_1 + 1)| \geq d$ is also satisfied.*

This paper primarily considers the case of $d = 1$. In the case of $d = 1$, there must be amplified segments in each DNA sequence whose lengths differ from segments amplified in the other DNA sequences. We propose algorithms for this problem below.

3 Algorithms

3.1 Framework of Algorithms

This section proposes a framework for the algorithms using the algorithms for a single PCR amplification, which are explained in section 3.2. We call the algorithm used for a single PCR experiment *A1*.

The structure of the algorithm is:

1. Apply *A1* until the finishing condition is satisfied.

2. Remove DNA sequences that are newly identified by the primer set selected by *A1* in step 1.

3. Return to step 1, until all the DNA sequences are identified by at least one primer set or no more DNA sequences are distinguished by the primer set selected by *A1*.

We used the term 'Remove' in step 2. However, subsequent primer pairs may amplify DNA sequences already 'removed'. Therefore, we also have to consider amplified segments in DNA sequences that have already been identified. We consider how to deal with these in section 4.1.

The finishing condition of *A1* in step 1 is when 100 amplified segments between 50 and 500 bases long are obtained. 100 amplified segments are the practical limit that can be distinguished by electrophoresis.

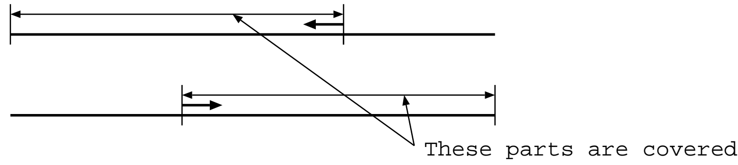

Figure 2: Covered elements for a primer.

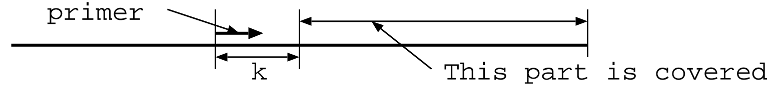

Figure 3: Modifying covered elements for a primer.

3.2 Greedy Algorithm for a single PCR Amplification

In this subsection, we explain the greedy algorithm *A1* in [3]. The algorithm regards the primer selection problem as a set cover problem. Position p of DNA sequence j (j, p) is regarded as an element of the underground set. The primers are regarded as subsets. Fig. 2 shows the elements of a primer.

To be exact, (j,p) $(1 \leq j \leq n, 1 \leq p \leq m_j + 1)$ are elements in the underground set for the set cover problem. (We add $(j, p+1)$ for all DNA sequences j.) If $(j,p,\mathrm{r}) \in S_i$, primer S_i covers $(j,p{+}1),(j,p{+}2),...,(j,m_j),(j,m_j{+}1)$. If $(j,p,\mathrm{l}) \in S_i$, primer S_i covers $(j,1),(j,2),...,(j,p{-}1),(j,p)$.

All (j,p) $(1 \leq j \leq n, 1 \leq p \leq m_j + 1)$ are covered if and only if amplified sequences exist in all the target DNA sequences.

The greedy algorithm is as follows:

1. Scan the DNA sequences and identify primers that satisfy the biological conditions. Make a list of candidate primers.
2. Select primer S_i that includes the largest number of uncovered elements from the candidate primers, provided that it does not meet any of the prohibitions placed on primers. Delete S_i from the set of candidates and add S_i to solution.
3. Repeat step 2 until all the elements are covered or there are no more candidate primers.

The algorithm only considers constraints on the GC content and primer length in step 1, which produces candidate primers. The restrictions and modifications considered in step 2 are:

complementary condition A sequence whose complementary sequence has been selected cannot be selected.

covered elements modification If $(j,p,\mathrm{r}) \in S_i$, primer S_i covered $(j,p{+}1),(j,p{+}2),...,(j,m_j),(j,m_j{+}1)$ in *A1*. But in this algorithm, if $(j,p,\mathrm{r}) \in S_i$, primer S_i covered $(j,p+k+1),(j,p+k+2),...,(j,m_j+k),(j,m_j+k+1)$. (Given k for the minimum length of the amplified segment. See Fig. 3.)

short length condition If primer S_i $((j,p,\mathrm{r}) \in S_i)$ was selected, S_i' $((j,p',\mathrm{r/l}) \in S_i'$, $p \leq p' < p+t)$ cannot be selected. If primer S_i $((j,p,\mathrm{l}) \in S_i)$ was selected, S_i' $((j,p',\mathrm{r/l}) \in S_i'$, $p-t < p' \leq p)$ cannot be selected. (Given t for the minimum length of the amplified segment.)

resolution condition If the lengths of the segments that are amplified by a candidate sequence and the primers that have already been selected by the algorithm are the same, the candidate segment cannot be selected. We focus on amplified segments with lengths between 50 and 500,

Table 1: Effects of extending the algorithm to already identified DNA sequences.

extension	no	yes
# experiments	42	31
# primers	292	944
average # of primers	7.0	30.5
$n_d(50, 500)$	2034	3128

so only the lengths of segments between 50 and 500 bases long need to differ from each other. We call this the prohibitive condition for between 50 and 500 only. (This differs from the definition in [3].)

4 Extension of the Algorithm and Experimental Results

We applied the greedy algorithm to the 6,218 yeast ORFs stored in GenBank, and modified the algorithm accordingly. We set the GC content of a primer to 45-55% and the length of a primer to 8. We focused on amplified segments with lengths between 50 and 500 bases.

We set k (the parameter for *the covered element modification*)$= 50$ and t (the parameter for *the short length condition*) $= 50$. In the following tables, $n_d(50, 500)$ is the number of DNA sequences that have amplified segments with lengths between 50 and 500 that differ from those of other amplified segments.

4.1 Extension for Already Identified DNA Sequences

It is important for our algorithm to be able to handle DNA sequences that are amplified by primer pairs already selected. Therefore, we propose an extension to the algorithm to deal with DNA sequences that have already been identified.

This extension selects primers that cover the largest number of uncovered elements. It is preferable that the number of segments of already identified DNA sequences amplified is small. Therefore, we select the primer that covers the largest number of uncovered elements from the candidate primers that amplify the smallest number of sequences from previously identified DNA sequences.

Next, the prohibitive conditions are considered. *The short length condition* must only be applied to unidentified DNA sequences, because the identified DNA sequences do not need long amplified segments to be distinguished. Since the lengths of the amplified segments in both identified and unidentified DNA sequences must be different to identify the unidentified segments, we consider both identified and unidentified DNA sequences in *the resolution condition.*

Table 1 shows the effects of extending the algorithm to already identified DNA sequences. This decreases the number of amplifications required and increases the number of DNA sequences identified. Nevertheless, these primer sets distinguish only half of all the yeast DNA sequences, so the solution is inadequate. This result occurs because *A1* in section 3.2 tends to amplify specific DNA sequences. Although we attempt to make it more difficult to amplify already identified DNA sequences; some specific DNA sequences are still amplified. Therefore it is necessary to modify *A1* as presented in section 3.2.

4.2 Order of Priority of the Greedy Algorithm

In the case presented in [3], it was not necessary to consider the number of DNA sequences covered, because the solution covered all of the DNA sequences. In the current case, however, we consider

Table 2: Effect of modifying the priority in the greedy algorithm.

changing the order of priority	no	yes
# experiments	31	98
# primers	944	4853
average # of primers	30.5	49.5
$n_d(50, 500)$	3128	5750

obtaining a small set of primers that amplify some (but not all) of the DNA sequences. We must consider selecting single primers that cover many DNA sequences.

The greedy algorithm can be written,

1. Scan the DNA sequences and identify primers that satisfy the biological conditions. Make a list of candidate primers.
2. Select the primer S_i from the candidate primers, if the largest number of uncovered DNA sequences are amplified by S_i and the primers that have already been selected by the algorithm, and it does not meet any of the prohibitions placed on primers. If several primers cover the same number of uncovered DNA sequences, select primer S_i that includes the largest number of uncovered elements. Delete S_i from the candidate set and add S_i to the solution.
3. Repeat 2 until all the elements are covered or there are no more candidate primers.

The elements are the same as in subsection 3.2. The prohibitions in step 2 are *the complementary condition*, *the short length condition*, and *the resolution condition*. We also apply the extension outlined in subsection 4.1.

We incorporated another idea into our algorithm: The idea of assigning a weight to elements and DNA sequences. In the previous section, the greedy algorithm assigns priority according to the number of elements. This is very simple; however, long DNA sequences are easier to amplify than short DNA sequences. We control that by assigning a weight for the elements and DNA sequences. The weight of one element is set to 1/(length of the DNA sequence containing the element). The weight of a DNA sequence is set to 1/(length of the DNA sequence). The algorithm uses these weights instead of the number of elements and the number of DNA sequences.

Table 2 shows that altering the priority produces a better solution. The number of primers is 85% the number of identified DNA sequences, which represents more than 90% of all the target DNA sequences, and hence is 42% of the number needed in the multiplex PCR. Furthermore, each primer is less than half as long as usual, thereby reducing the cost of the experiments to 20% of the cost of multiplex PCR. Approximately 470 DNA sequences are not distinguishable. Multiplex PCR cannot amplify more than 100 DNA sequences with the given biological constraints, although it can amplify about 350 DNA sequences for which our algorithm fails to find primer sets. We should be able to obtain a better solution by relaxing the biological constraints or by altering the parameters fixed in our algorithm: GC content or primer length. (Of course, if long primers are used, almost all the sequences amplified by multiplex PCR will also be amplified.) Other algorithms must be used to set the biological constraints.

Fig. 4 shows the number of primers and the number of newly identified DNA sequences obtained with each primer pair for the result on the right side of Table 2. Our algorithms obtain primer sets one at a time. Therefore, at the beginning of the algorithm, good primer sets are obtained, but the final primer pairs obtained by the algorithm are not as good. As the number of primers increases, the number of newly identified DNA sequences decreases. By the end of the algorithm, the primer pairs identified are worse than those for multiplex PCR, based on the number of newly identified DNA

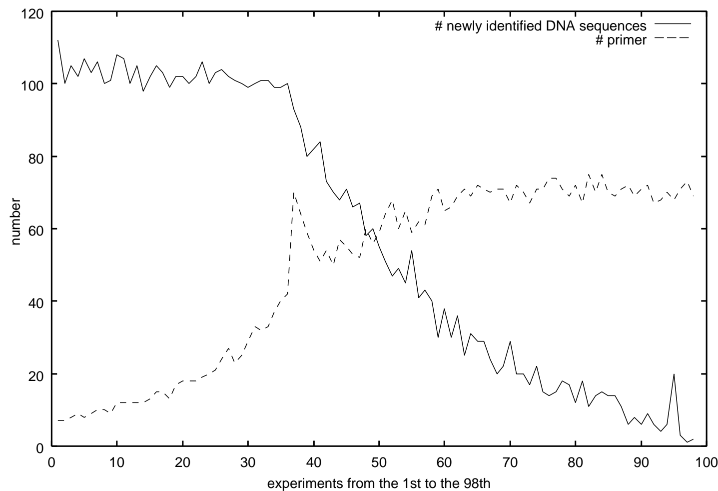

Figure 4: Number of primers and number of newly identified DNA sequences for each amplification.

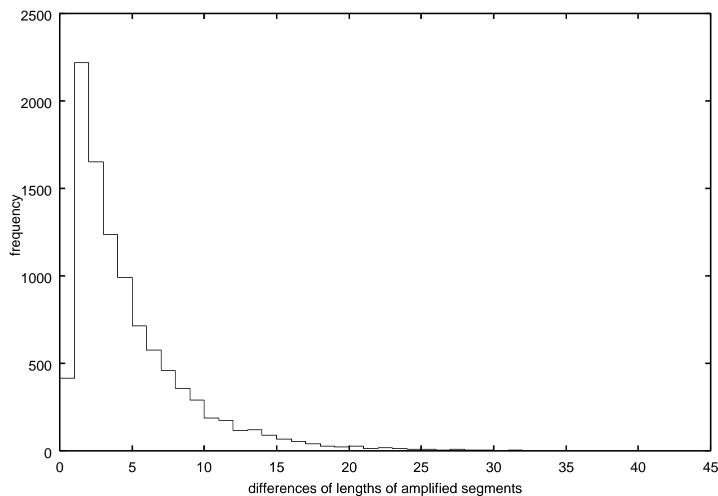

Figure 5: Histogram of differences in the lengths of amplified segments.

sequences. A simple algorithm can improve this solution. This algorithm removes primers that do not amplify newly identified DNA sequences from the set of primers. We have omitted details of this algorithm and the experimental results for this version because of space limitations.

This paper considers the case when the difference in the lengths of the amplified segments is 1. However, it is preferable that the difference is more than 5. This would mean that 90 segments between 50 and 500 bases long could be amplified. Fig. 5 is a histogram showing the differences in the lengths of segments between 50 and 500 amplified for the results on the right side of Table 2. For already identified DNA sequences, the difference can be 0. The greatest number of segments is produced when the difference is 1. It is preferable that the frequency of cases with a difference of 1 is small. We must improve the algorithm dealing with differences.

5 Conclusions and Future Work

We extended the greedy algorithm in [3] to solve a real large-scale problem by controlling the DNA sequences identified by already selected primer pairs and modifying the priorities in the greedy algorithm. We obtained a good result; the number of primers was 85% the number of identified DNA sequences, which represents more than 90% of all the target DNA sequences. Nevertheless, some DNA

sequences still cannot be distinguished. This is partially the result of the strict biological constraints. When the primer length is 8 and the GC content is 45% to 55%, multiplex PCR will not amplify about 100 DNA sequences. We plan to devise an algorithm to decide suitable biological constraints. We also consider it necessary to incorporate another algorithm, to obtain a solution that distinguishes all DNA sequences. This paper deals with the case when the difference in the length of the amplified segments is 1. However, a difference of 1 is not preferable, and we plan to examine the case for larger differences.

The approach in this paper may be regarded as non-hierarchical. We may adopt a hierarchical approach if we have suitably well clustered DNA sequences. We are now investigating the problem of finding good clusters.

ATAC PCR (adaptor-tagged competitive PCR) is a new technique for determining the quantity of gene expression. We also plan to analyze problems and construct algorithms related to ATAC PCR.

Acknowledgments

Part of the work by the second author was supported by a Grant-in-Aid for the Priority Area 'Genome Science' from the Ministry of Education, Science, Sports, and Culture of Japan.

References

[1] Chamberlain, J.S., Gibbs, R.A., Ranier, J.E., Nguyen, P.N., and Caskey, C.T., Deletion screening of the duchenne muscular dystrophy locus via multiplex DNA amplification, *Nucleic Acids Research*, 16:11141–11156, 1988.

[2] Doi, K., Complexity Analyses and Experiments about Minimizing the Number of Primers in PCR Experiments, Master Thesis, Department of Information Science, Faculty of Science, University of Tokyo, February 1998.

[3] Doi, K. and Imai, H., Greedy algorithms for finding a small set of primers satisfying cover and length resolution conditions in PCR experiments, *Genome Informatics 1997*, 43–52, 1997.

[4] Ito, T. and Sakaki, Y., Fluorescent differential display method for high-speed scanning of tissue or cell-specific transcripts, *Methods Mol. Genet.*, 8:229–245, 1996.

[5] Mitsuhashi, M., Technical Report: Part 2. Basic Requirements for Designing Optimal PCR Primers, *Journal of Clinical Laboratory Analysis*, 10:285–293, 1996.

[6] Mitsuhashi, M., Cooper, A., Ogura, M., Shinagawa, T., Yano, K., and Hosokawa, T., Oligonucleotide probe design – a wew approach, *Nature*, 367:759–761, 1994.

[7] Nicodème, P. and Steyaert, J., Selecting optimal oligonucleotide primers for multiplex PCR, *Proc. Fifth International Conference on Intelligent Systems for Molecular Biology*, AAAI press, 210–213, 1997.

[8] Pearson, W.R., Robins, G., Wrege, D.E. and Zhang, T., On the primer selection problem in polymerase chain reaction experiments, *Discrete Appl. Math.*, 71:231–246, 1996.

[9] Rowe, T., Sharefkin, J., Yan, S.Q., and Diffenbach, C.W., A computer program for selection of oligonucleotide primers for polymerase chain reactions, *Nucleic Acids Research*, 18:1757–1761, 1990.

[10] Welsh, J. and McClelland, M., Fingerprinting genomes using PCR with arbitrary primers, *Nucleic Acids Research*, 18:7213–7218, 1990.