A Computer Support for Genotyping by PCR-Multiplex

Pierre Nicodème*
INRIA and LIX École Polytechnique (Email: Pierre.Nicodeme@inria.fr)
and
Jean-Marc Steyaert
LIX École Polytechnique, 91128 Palaiseau, France (Email: Steyaert@lix.polytechnique.fr)

February 27, 1997

Abstract

We investigate the problem of multiplexing PCR for medical applications. We show that the problem is NP-complete by transformation to the Multiple Choice Matching problem and give an efficient approximation algorithm. We developed this algorithm in a computer program that predicts which genomic regions may be simultaneously amplified by PCR. Practical use of the software shows that the method can treat 250 non-polymorphic *loci* with less than 5 simultaneous experiments.
Keywords. PCR-Multiplexing, diagnostic, urn models, NP-completeness.

1 Introduction

Introduced in the mid-1980s, the Polymerase Chain Reaction, PCR for short, is able to amplify segments of DNA a million times. The method is sensible to very small amounts of DNA, and is considerably faster than other methods. However, in most PCR experiments performed by biologists, the amplification of each target fragment of DNA requires a separate and costly PCR experiment, which involves difficult manipulations and monopolises an automat [8] for its exclusive use. PCR has numerous applications. For instance, it is known that diseases may result of long deletions on one or several genes. Such pathological genes may be characterised by the incorrect amplification of *loci* where deletions have occurred. Genotyping requires PCR amplifications of many different *loci*; whenever it is possible to group these PCR amplifications in a same experiment, called PCR multiplex, time and money can be saved. We study in this paper the conditions required for PCR multiplex and propose an algorithm that performs an almost optimal choice of PCR primers, with the aim of minimising the number of PCR multiplex operations.

*Pierre Nicodème, INRIA-Rocquencourt, BP105, 78153 - Le Chesnay Cedex - France. Tel: 33-(0)139635443 - Fax: 33-(0)139635659

Besides genotyping applications, PCR multiplexing has been used to detect multigene families [5]. Multigene families are composed of many highly homologous genes and biologists often want to discover new genes belonging to the family. More recently, PCR multiplex has also been used to accelerate the ordering of contigs (contiguous segments of DNA) in DNA physical mapping [12]. The algorithmic problems corresponding to these two applications are described in [4, 9].

This paper is organised as follows. In section 2, we present the main concepts of the Polymerase Chain Reaction. In section 3, we present the constraints of PCR multiplex, and describe our algorithm. In section 4, we give experimental results. In section 5, we show that our problem is NP-complete and in section 6, we give a probability analysis of our problem.

2 Biological overview of the Polymerase Chain Reaction

We present in this section a short overview of the PCR subject. We refer to J. D. Watson and al., *Recombinant DNA* [13], for a detailed introduction to this subject.

PCR exploits certain features of DNA replication. Single-stranded DNA is used as a template for the synthesis of a complementary new strand. These single stranded DNA templates can be produced by simply heating double-stranded DNA to near boiling temperature. Then a small fragment of double stranded DNA is required to initiate ("prime") synthesis.

The starting point for DNA synthesis can be specified by supplying an oligonucleotide primer that anneals to the template at this point. Both DNA strands can serve as templates for synthesis provided an oligonucleotide primer is supplied for each strand.

Each cycle of PCR duplicates the segments under amplification. Starting from one segment, n cycles of PCR produce therefore 2^n segments.

Figure 1 shows the synthesis initiated by the forward primer `5'-ACACA...AGCAA-3'` on the 3'-5' strand of a segment of DNA.

```
5' ..CTGACACAACTGTGTTCACTAGCAA......AAGGTGAACGTGGATGAAGTTGGTG.. 3'
                              3'<<-TTCCACTTGCACCTACTTCAAC 5'
                                   reverse primer
       forward primer
     5'ACACAACTGTGTTCACTAGCAA->> 3'
3' ..GACTGTGTTGACACAAGTGATCGTT......TTCCACTTGCACCTACTTCAACCAC.. 5'
```

Figure 1: Primers for DNA polymerase

Primers cannot be chosen at will inside a *locus* (a portion) of a gene: they must respect conditions permitting a correct amplification by PCR; the temperature of hybridisation at which the polymerase synthesises the new DNA strand is one of these conditions. This temperature depends on the composition of the primer, and more specifically on the respective percentage of the bases A and T, versus the bases G and C it contains. When choosing a pair of primers, the hybridising temperatures of the two primers should be about the same. It is important to prevent the two primers to hybridise to each other, or identical copies of a self-homologous primer to hybridise together. We discuss these conditions in next section.

3 Multiplexing the Polymerase Chain Reaction

Applying the method described in the preceding section to n *loci* and n pairs of primers, we want the conditions mentioned above (valid for a pair of primers) to still hold for the $\binom{2n}{2}$ combinations of two among the $2n$ primers.

3.1 Conditions for PCR Multiplex

We detail in this section a model of compatibility between primers and its requirements for classical PCR; we leave to future work extension of the model to Long Accurate PCR (LA-PCR) [1, 2] which permits amplification of very long segments of several thousand bases.

We speak of *locus* amplification when considering the amplification of a single segment. Only one amplification is allowed inside a given *locus*, and to each *locus* amplification corresponds a *forward* and a *reverse* primer. We define a *subprimer* as a subsequence of length σ of a primer and we consider in the following that σ has the same value for all subprimers of a multiplexing experiment. In practical experimentations, σ will have values 4 or 5. We define a *3'-subprimer* as the subprimer ending a primer at its 3' extremity (primers are always read in the direction 5' $\Rightarrow$ 3').

The morphism $\mathcal{C}(A \to T, T \to A, C \to G, G \to C)$, corresponding to DNA base pairing, of the alphabet $\Sigma = \{A, C, G, T\}$ over itself is naturally extended to the "complementation" morphism $\mathcal{C}^*$ of Σ^* over itself. The "reverse" morphism $\mathcal{R}^*$ of Σ^* over itself transforms a string $s_1 \in \Sigma^*$ into a string s_2 by simply reading backwards s_1. Both $\mathcal{C}^*$ and $\mathcal{R}^*$ are involutive and $\mathcal{C}^* \circ \mathcal{R}^* = \mathcal{R}^* \circ \mathcal{C}^*$. Two strings s_1 and s_2 are said *reverse complementary* if $s_2 = \mathcal{C}^* \circ \mathcal{R}^*(s_1) = \mathcal{R}^* \circ \mathcal{C}^*(s_1)$.

The requirements are the following:

1. **Locus** amplification **requirements**:
    (a) *Pairing-distance.* The distance between a forward primer and a reverse primer (pairing-distance) must be in the range of 150-450 bases (the minimum and maximum values given here are indicative).
    (b) *Non-palindromicity.* The primers satisfy the conditions of non-palindromicity preventing self-homology.
    (c) *Reverse-complementarity.* The 3'-subprimers must be reverse complementary to none of the subprimers (subprimers as 3'-subprimers are assumed to be of length σ bases).
2. **Multi-locus** amplification or **experiment requirements**:
    (a) *Reverse-complementarity.* Any 3'-subprimer of an experiment is not reverse complementary to any subprimer, including itself, of any primer of the

experiment; this would initiate hybridisation of the primers themselves. An example for this condition is given in subsection 3.2. Note that subprimers (including the 3'-subprimers) may be identical between different *loci*, or inside a *locus*.

(b) *GT-AC composition.* The temperatures of denaturation, or the GC/AT percentage in the primers of a multi-*locus* PCR amplification must belong to a limited range of values (for instance, $48\% - 52\%$).

(c) *Electrophoresis distance.* The difference of lengths between any two segments amplified in the same multi-*locus* PCR amplification must be greater than δ bases; this is necessary to allow a correct differentiation of the amplified segments after electrophoresis[1]. This distance supposes that the *loci* are not polymorphic; in case of polymorphic loci, the problem of differentiating the amplified segments has to be handled in a different way.

3.2 An urn model to solve the problem of compatibility between primers

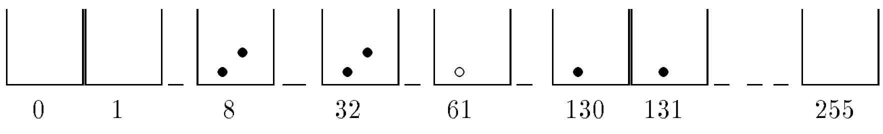
We give here a constructive example of our algorithm.

- We associate to each base a number between 0 and 3
- We associate to each subprimer of length $\sigma = 4$ a number between 0 and $4^4 = 256$, and to each subprimer of length $\sigma = 5$ a number between 0 and $4^5 = 1024$, as follows:

  A $\Rightarrow$ 0, C $\Rightarrow$ 1, G $\Rightarrow$ 2, T $\Rightarrow$ 3, giving for example TTA $\Rightarrow 330_4 \Rightarrow 60_{10}$

We can then consider a model of 256 urns, when $\sigma = 4$, or a model of 1024 urns, when $\sigma = 5$. The compatibility constraint is then transformed into a compatibility rule over an urn model for which the following is an example (for $\sigma = 4$):

Primer			Urn number	
3' CTTA 5'	$\stackrel{\bowtie}{\rightarrow}$ 5' *ATTC* 3'	$\rightarrow$	*61*	○
↕↕↕↕				
5' AAGAAGAAT 3'				
AAGA		$\rightarrow$	8	•
AGAA		$\rightarrow$	32	•
GAAG		$\rightarrow$	130	•
AAGA		$\rightarrow$	8	•
AGAA		$\rightarrow$	32	•
GAAT		$\rightarrow$	131	•



---

[1]Electrophoresis is a migration method which allows short segments to move faster than the long one; this method allows the differentiation of segments of different length, from a mixture of them, but it has a limited precision corresponding to our parameter δ.

The compatibility rule implies that an urn can never contain black and white balls simultaneously.

3.3 An algorithm deriving from the urn model: MULTIPCR

We propose in this section an approximate algorithm with high efficiency in practical computations; this algorithm is likely to be almost optimal.

Our algorithm is as follows.

- We sort our set of *loci* in increasing order of the number of candidate pairs of primers;.
- We process our set of ordered *loci*, *locus* after *locus*.
- For each *locus*, we finally try each possible pair of primers with respect to the conditions given above, including the distance condition (requirement 1(a)).

  For each pair, we "throw white and black balls in the urns", with respect to the model described above; we eliminate the pairs which cause "black and white" collisions; among the acceptable pairs of primers, we select the pair of primers which minimises, in the following order:

  1. the number of urns containing white balls;
  2. the number of urns containing black balls, whenever the number of "white urns" is identical for two pairs.

  The "white and black balls" corresponding to pairs of primers already selected remain in the urns when processing a new *locus*.

The *loci* providing no compatible pair with the pairs of the *loci* already selected for the current experiment are left apart and processed in a next experiment.

Complexity. The complexity C of our algorithm is $C = \mathcal{O}(Knl)$ with $K \approx E\tau_f k_r \approx 6E$, where E is the number of experiments, n is the number of *loci*, l is the average length of a *locus*, $\tau_f \approx 1/7$ is the probability of getting a forward primer at a position, and $k_r \approx 40$ is a constant giving the approximate number of acceptable reverse primers for a forward primer. (See section 4 for the numerical values of τ_f and k_r).

3.4 Software Implementation

The *MULTIPCR* program implements the algorithm described in the preceding section.

Different software programs are available predicting which pair of primers to choose inside a given *locus*.

We use the program *PRIMER*, of S.E. Lincoln, M.J. Daly, and E.S. Lander [6] as a preparation step in our program. *PRIMER* is a two-step program; *step-1* selects candidates for forward and reverse primers; *step-2* chooses a best pair of one forward and one reverse primer among all the possible pairs of candidates. *MULTIPCR* takes as input the output of *PRIMER step-1*, and chooses for each *locus* a forward and a reverse primer compatible with the primers chosen for the other *loci*, whenever this is possible.

Both *PRIMER* and *MULTIPCR* are written in the C language and implemented on the *Unix System V* system. The package "*PRIMER+MULTIPCR*" is available by anonymous ftp at `ftp://ftp.infobiogen.fr/pub/logiciels/unix/bio` and portable on SUN, DEC and Silicon Graphics workstations.

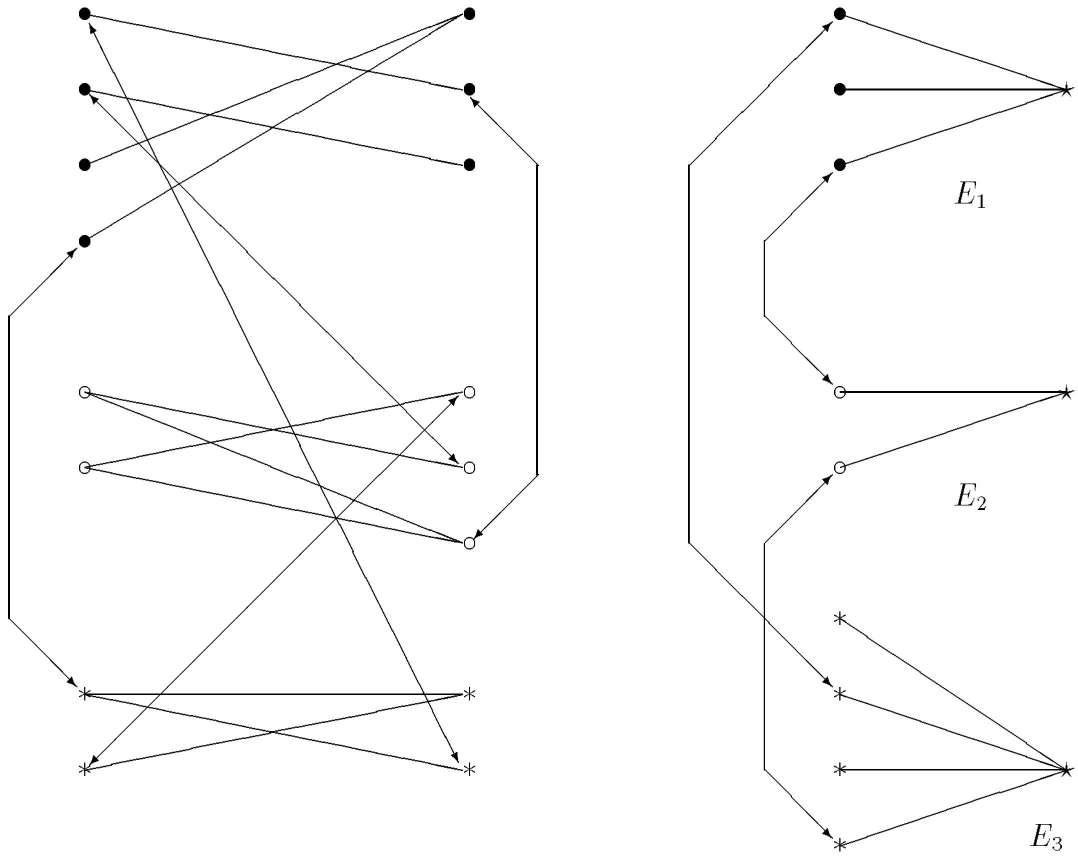
checking length σ	electro-phoresis distance δ	experiment number	number of *loci* amplified	number of white urns	number of black urns
4	1	1	214	34	222
		2	32	27	223
		3	2	4	48
	2	1	133	34	222
		2	113	26	230
		3	2	4	51
	3	1	84	31	224
		2	81	30	226
		3	73	34	222
		4	10	17	168
	4	1	61	33	222
		2	60	30	223
		3	61	32	222
		4	59	31	218
		5	7	10	128
	5	1	48	31	223
		2	45	29	223
		3	49	32	218
		4	49	30	219
		5	47	31	219
		6	10	19	171
no possible amplification			0		
number of *loci* processed			248		

Table 1: Submitting 248 *loci* to *MULTIPCR*, compatibility check length is 4 bases (model of 256 urns)

4 Experimental results

Table 1 shows the results obtained in a few seconds on a SUN 3-60 when processing 248 *loci* of Genbank; as a typical *locus*, the *locus* HUMDYSGP of the gene of dystrophy is 2202 base pairs long. The program *Primer* produces for this gene 343 forward and 339 reverse primers. Therefore, in the average, a primer starts each seventh position, and the pairing-distance condition implies that there are approximately 40 admissible reverse primers for each forward primer.

It is clear that simultaneous amplification of 214 *loci*, as proposed in Table 1, is biologically unrealistic, but a multiplex of about 10 to 20 PCR looks reasonable.



(a) A representation of the compatible primer problem

(b) A transformation of the multiple choice matching problem

Figure 2: Transformation of any graph to a set of bipartite subgraphs modelling the compatible primers problem

5 Determining the pairs of primers which maximise the number of *loci* in a single experiment is a NP-complete problem

We show in this section that our problem is NP-complete, justifying the use of an approximation algorithm. We mention here that Gabriel Robins and *al.* [10] demonstrate the NP-completeness of the minimisation problem for the multigene primer selection; they reduce their problem (an approximate super-string problem) into the *Minimum Set Cover* problem. We describe a transformation of our compatibility problem into the *Multiple Choice Matching* problem which proves NP-completeness of our problem. The demonstration of Robins and *al.* and our demonstration [7] have been made independently and about simultaneously. Moreover, our problem does not seem to be transformable simply to their problem, and this implies the necessity of two different demonstrations.

We first model our problem, representing it as a set of bipartite subgraphs with additional edges (Figure 2 (a)). In this graph, each primer is represented by a vertex and the set of vertices is partitioned by *locus*, each *locus* corresponding to a bipartite subgraph. In our example, vertices belonging to the same *locus* are represented by the same character ($\bullet$ for *locus* 1, $\circ$ for *locus* 2, $*$ for *locus* 3). We distinguish also forward vertices, corresponding to forward primers, that are represented on the left part of the figure (Figure 2 (a)), and reverse vertices, corresponding to reverse primers, that are represented on the right part of the figure (Figure 2 (a)).

We define two kind of edges:

- *acceptance edges*, inside the bipartite subgraph restricted to a single *locus*; such an undirected edge indicates that the primers associated to the forward vertex and the reverse vertex that it links together are compatible.

- *incompatibility edges*, joining a vertex of a *locus* to a vertex of a different *locus*; these directed edges indicate incompatibility of the primers associated with the vertices (the primers do not satisfy condition 2(a) of section 3.1.).

  Remark: the incompatibility edges are not represented inside a *locus*; if they were added, we would have a complete bipartite graph for each *locus*. They are "complementary" to the acceptance edges in each bipartite subgraph.

Our "*Compatible Primers Problem*", in short *CPP*, has the following description:

Instance of the problem: a graph composed of a set of bipartite graphs $B_1, B_2, ...,$ B_J (the edges of these graphs constitute a set of acceptance edges A); a set of incompatibility edges, which join pairs of vertices that do not belong to the same bipartite subgraphs; an integer K.

Question: is it possible to choose a subset of acceptance edges $A' \subseteq A$ with $|A'| \geq K$ such that A' contains at most one edge from each B_i, $1 \leq i \leq J$, and such that no two vertices belonging to these edges are extremities of an incompatibility edge?

CPP is in NP: a non-deterministic Turing Machine may write an arbitrary sequence of symbols representing edges and check in polynomial time if the corresponding set is a subset of acceptance edges A' that answers "yes" to the preceding question.

We now make use of the *"Multiple Choice Matching Problem"*, in short *MCMP*, which is known to be NP-complete [3]. The problem is given as follows:

Instance of the problem: a graph $G = (V, E)$, a partition of E into disjoint sets $E_1, E_2, ..., E_J$, a positive integer K.

Question: is there a subset $E' \subseteq E$ with $|E'| \geq K$ such that no two edges in E' share a common vertex and such that E' contains at most one edge from each E_i, $1 \leq i \leq J$?

Comment: the problem remains NP-complete even if G is bipartite.

Proof of the NP-completeness of CPP. We transform in polynomial time any graph to a graph of the *Compatibility Primer Problem* form (Figure 2), showing that if we could solve this last problem in polynomial time, we would also be able to solve the *Multiple Choice Matching Problem* in polynomial time, which would contradict the NP-completeness of *MCMP*.

Consider an instance *I(MCMP)* of the *Multiple Choice Matching Problem*. We transform any edge e_{ij} of a set E_i of *I(MCMP)* into a vertex f_{ij} of our *I(CPP)* instance of the *Compatible Primers Problem*. We now transform any vertex $v_{ij,kl}$, with $i \neq k$, which is the intersection of two edges e_{ij} and e_{kl} of different subsets E_i and E_k of *I(MCMP)* into an incompatibility edge $i_{ij,kl}$, joining vertices f_{ij} and f_{kl} in *I(CPP)*. We associate to each set E_i of *I(MCMP)* a dummy vertex d_i in *I(CPP)*, and we join this vertex to each vertex f_{ij} by an acceptance edge a_{ij} (remember that subscripts notations ij and kl refer here to edges of sets E_i and E_k labelled j and k respectively and do not refer to nodes indexing; edges e_{ij} and e_{kl} may therefore intersect). We do not transform the vertices of *I(MCMP)* which are intersections of edges of a same subset E_j because these intersections have no effect in the *Multiple Choice Matching Problem*.

Figure 2 (b) illustrates such a transformation, with three subsets E_1, E_2, E_3, and dummy vertices represented by a $\star$.

We have therefore transformed any *MCMP*-graph into a *CPP*-graph; the transform itself is polynomial. At this point, we see that solving the question of the *CPP* problem with the integer K would also solve the *MCMP* problem with the same integer K, because any such solution would provide a set A' of acceptances edges, with $|A'| \geq K$, and the non dummy vertices of these edges clearly correspond to a set E' of edges of the initial general graph, with $\mid E' \mid = \mid A' \mid \geq K$. Therefore, the *Compatible Primer Problem* is NP-complete.

6 Analysis of the MULTIPCR algorithm

Evaluating the limit probability of rejection of a *locus*. This section presents a simplified analysis of the urn model, when a steady state is reached; it gives insights into the optimality of our algorithm.

We make some empirical observations on Table 1: in this table, we see that, with an electrophoresis distance $\delta = 5$, the number of urns filled with black or white balls (subprimers and 3'-subprimers scores respectively) is almost equal to the total number U of urns (with $U = 256$), with less than about 50 *loci* in an experiment. If we consider the case $\delta = 1$, this means that, after less than 50 *loci*, we reach a "stationary" state where there are no more empty urns, and in which the composition in white (w) and black (b) urns stays constant till the end of the experiment. If s is the number of subprimers inside the primers (in practice, if $\sigma = 4$, we have $s = 17$ for primers of length 20), noting $\pi_{1,w,b}$ the probability that a primer is accepted by a system of U urns either black or white ($U = w + b$), we have

$$\pi_{1,w,b} = \frac{w}{U}\left(1 - \frac{w}{U}\right)^s \quad \text{and} \quad \pi_{1,30,226} \approx 0.014 \tag{1}$$

The probability π_{11} of compatibility of two primers (thrown inside an empty system of urns), is

$$\pi_{11} = \frac{1}{U}\left(1 - \frac{1}{U}\right)^{2s} + \left(1 - \frac{1}{U}\right)\left(1 - \frac{2}{U}\right)^{2s} \approx 0.766 \tag{2}$$

The *MULTIPCR* algorithm considers a number F of candidate forward primers for a *locus*, and, for each forward primer, a number R of candidate reverse primers at an acceptable pairing-distance (in the range 150-450 bp.) of this forward primer. Depending on the *loci* length, F is in the range 100-500, while R stays close to 50.

When considering a *locus*, the asymptotical distributions of the number of accepted forward primers f and accepted reverse primers r are both binomials; the small value of $\pi_{1,30,226} \approx 0.014$ allows us to apply the Poisson approximation to these binomial distributions, with respective parameters $\phi = F\pi_{1,w,b}$ and $\rho = R\pi_{11}\pi_{1,w,b}$ (with $w+b = U$). The probability Π of rejection of a *locus*, at the stationary state, is then

$$\Pi(\phi(F), \rho(R)) = \sum_{i=0}^{\infty}(Pr\{f = i\} \times (Pr\{r = 0\})^i) = e^{-\phi+\phi e^{-\rho}}, \tag{3}$$

probability for which some values are given for $R = 50$ in the table below:

V	250	300	350	400	450
$\Pi(\phi(V), \rho(50))$	0.230	0.171	0.128	0.095	0.071

Considering the results obtained for 248 *loci* (Table 1), with an average number of 300 to 350 forward primers per *locus*, we see that our algorithm is quasi-optimal.

Remark It is interesting to have an idea of the dimension of the *space* in which we search the *solutions* to our problem; given $n = m \times k$ *loci*, with m the number of experiments and k the average number of loci in one experiment, let $p_{n,k}$ be the number of partitions of n elements into sets of size k (a set corresponds in this model to a PCR-multiplex experiment); there could be many such partitions. We give here an idea of the value of $p_{n,k}$ for values of n and k obtained in biological applications.

Let $P(z) = \frac{\sum p_{n,k}}{n!} z^n$ be the exponential generating function of the $p_{n,k}$. We count the number of partitions by considering sets $\mathcal{S}_k$ of k elements, and sets of such sets. Noting $S_k(z)$ the exponential generating function for the $\mathcal{S}_k$, standard symbolic methods for counting labelled objects [11] gives

$$S_k(z) = \frac{z^k}{k!} \quad \text{and} \quad P(z) = e^{S_k(z)} = e^{\frac{z^k}{k!}}. \tag{4}$$

We therefore get

$$p_{mk,k} = \frac{(mk)!}{m!(k!)^m}$$

We noticed that in biological applications, it is unreasonable to amplify simultaneously more than about 20 *loci*. Considering $n = 200$ *loci*, $m = 10$ and $k = 20$, we get

$$p_{200,20} = \frac{200!}{10!(20!)^{10}} \approx 3 \times 10^{185}.$$

Our approximation algorithm drastically reduces the size of the space in which we effectively search for solutions.

7 Conclusion

We showed that the theoretical problem of grouping a maximum number of *loci* that satisfy constraints allowing PCR multiplex is NP-complete. We explored a model of compatibility between primers that allows multiplexing and provided the biologists with an efficient and simple tool to choose the pairs of primers in the general case. This model considers that the crucial point is to avoid reverse complementarity of the 3' end of the primers to subsequences of any primer in the mixture used for the multiplex PCR experiment. Considering the traditional PCR method, where the length of the amplified segments is limited to some 500 bp., our *MULTIPCR* approach encounters serious limitations when applied to highly polymorphic *loci*. The method is also not applicable when only one pair of primer is available for each *locus*, which is the case when considering *primers dictionary*. However, the recent emergence of *Long Accurate PCR* [2, 1], which allows amplification of long segments (comprising as much as 35-kb) offers excellent perspectives to the application of our method to genotyping and to DNA physical mapping.

Acknowledgements. We are grateful to Gilles Thomas of Institut Curie, Paris and to Philippe Dessen of Infobiogen, Villejuif, for frequent discussions and advice concerning the biological subjects addressed in this paper. We also thank Marie-France Sagot for reading this paper.

References

[1] Barnes, W. M. Efective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci. USA 91* (June 1994), 5695–5699.

[2] Barnes, W. M. PCR amplification of up to 35-kb DNA with high fidelity and high yield from λ bacteriophage templates. *Proc. Natl. Acad. Sci. USA 91* (March 1994), 2216–2220.

[3] Garey, M. R., and Johnson, D. S. *Computers and Intractability, A Guide to the Theory of NP-completeness.* W. H. Freeman and Company, 1979. Appendix.

[4] Grebinski, V., and Kucherov, G. Reconstructing a Hamiltonian circuit by querying the graph: Application to DNA physical mapping. INRIA-Lorraine and CRIN/CNRS, Campus Scientifique, 615 rue du Jardin Botanique, BP 101, 54602 Villers-lès-Nancy, France, 1996.

[5] Harrison, J. K., Pearson, W. R., and Lynch, K. R. Molecular characterization of alpha-1 and alpha-2 adrenoreceptors. *Trends Pharm. Sci. 12* (1991), 62–67.

[6] Lincoln, S. E., Daly, M. J., and Lander, E. S. *PRIMER: A Computer Program for Automatically Selecting PCR Primers.* MIT Center for Genome Research and Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, 1991.

[7] Nicodeme, P. A computer support for genotyping by multiplex PCR. Tech. Rep. LIX/RR/93/09, LIX, École Polytechnique, 91128, Palaiseau Cedex, France, 1993.

[8] Olschwang, S., Delaitre, O., Melot, T., Peter, M., Schmitt, A., Frelat, G., and Thomas, G. Description and use of a simple laboratory-made automat for *in vitro* DNA amplification. *Methods in Molecular and Cellular Biology 1*, 3 (May/June 1989), 121–127.

[9] Pearson, W. R., Robins, G., Wrege, D. E., and Zhang, T. A new approach to primer selection in polymerase chain reaction experiments. In *Third International Conference on Intelligent Systems for Molecular Biology* (1995), AAAI Press, pp. 285–291.

[10] Robins, G., Wrege, D. E., Zhang, T., and Pearson, W. R. On the primer selection problem in polymerase chain reaction experiments. Tech. Rep. CS-93-68, Department of Computer Science, University of Virginia, Charlottesville, Virginia 22903, U.S.A., 1993.

[11] Sedgewick, R., and Flajolet, P. *An Introduction to Analysis of Algorithm.* Addison-Wesley, 1996.

[12] Sorokin, A., Lapidus, A., Capuano, V., Galleron, N., Putjic, P., and Ehrlich, S. D. A new approach using multiplex long accurate PCR and yeast artificial chromosomes for bacterial mapping and sequencing. *Genome Research 6* (1996), 448–453.

[13] Watson, J. D., Witkowski, J., Gilman, M., and Zoller, M. *Recombinant DNA.* Scientific American Books, 1992. second edition.