

DNA FINGERPRINTING INVOLVING FLUORESCENCE-LABELED TERMINI OF ANY ENZYMATICALLY GENERATED FRAGMENTS OF DNA

Xiaoren TANG,¹ Yifei WANG,² Hai-Ou LI,¹ Osamu SAKATSUME,¹
Akinori SARAI,² and Kazushige YOKOYAMA^{1,*}

¹DNA Bank and ²Genetic Information Bank, Gene Bank, Tsukuba Life Science Center,
RIKEN (The Institute of Physical and Chemical Research),
3-1-1 Koyadai, Tsukuba, Ibaraki 305, Japan

Summary We have developed a new fluorescence-based method for DNA fingerprinting that does not require a fluorescent linker or a synthetic oligonucleotide primer, both of which are normally used for labeling of DNA. Cosmid DNAs are digested with appropriate restriction enzymes and the 3' termini of DNA fragments are labeled with the corresponding, fluorescent dye-conjugated dideoxynucleotide triphosphate terminator (dye-ddNTP) by the Klenow fragment of DNA polymerase I from *Escherichia coli*, which has 3'→5' exonuclease and replacement activities as well as its main 5'→3' polymerase activity. Samples are separated on a DNA-sequencing gel and data are analyzed by application of both the Version 0.3.8a mapper program (Applied Biosystem Inc., Foster City, CA) and our Overlap I program that facilitate rapid analysis of the frequency of overlapping of cosmid DNAs. Using this method we have determined the overlap frequency of DNA fragments of each cosmid clone from the mouse MHC class I gene cluster.

Key Words DNA fingerprinting, mapping, cosmid contigs, Overlap I program, MHC Class I gene cluster

INTRODUCTION

Electrophoresis of DNA fragments obtained by digestion with restriction enzymes has been useful for determining the sizes of such DNA fragments (Smith *et al.*, 1987). The DNA fragments of various sizes, derived from individual cloned DNAs, provide a "fingerprint" which can be used to identify other DNA clones

Received May 6, 1994; Revised version accepted August 11, 1994.

* To whom correspondence should be addressed.

that have a partially matching contig pattern (Olson *et al.*, 1986). Although gel electrophoresis of DNA is a simple and versatile method for routine analysis of DNA, it is sometimes difficult to identify the overlapping DNA fragments from yeast artificial chromosome (YAC) or cosmid recombinant clones that contain a large piece of DNA for the following reasons. 1) The distribution of sizes of DNA fragments is logarithmic (Calladine *et al.*, 1991) so the separation of smaller (more rapidly migrating) DNA fragments is greater than that of larger DNA fragments, 2) The DNA in agarose gels is usually stained with ethidium bromide and the mobility of DNA fragments is determined manually with a ruler or by some semi-automated system. 3) One or more lanes on the gel must be sacrificed for molecular-size standards. Clearly, there is a need for the development of new approaches to DNA fingerprinting that are amenable to complete automation.

Recently, the semi-automated detection of fluorescently labeled fragments of DNA has been accomplished by use of a fluorophore that is covalently attached to the oligonucleotide primer used in enzymatic DNA-sequence analysis. The fluorophores and instruments have been designed such that the fluorometer is located at a fixed distance (25 cm) from the origin of the gel, and all fragments of DNA travel the same distance through the gel before they are detected (Smith *et al.*, 1986). Furthermore, a fluorescence-based method for DNA fingerprinting has been developed (Carrano *et al.*, 1989). A single-stranded oligonucleotide primer (18-mer) with a covalently bound fluorochrome serves to create a double-stranded oligonucleotide linker with a 5'-overhang that is complementary to a restriction site. DNA is digested with the appropriate restriction enzyme and ligated to the fluorochrome-labeled linker by T4 ligase. Three labeled and restricted samples of DNA, plus one size-standards, with each of the different fluorescent colors, can be loaded in individual lanes of a polyacrylamide gel on a DNA sequencer (Carrano *et al.*, 1989), allowing the sizes of DNA fragments to be calculated automatically. However, a long time is required for preparation of samples. For example, it takes 8–12 hr for a complete digestion reaction and fluorescent labeling and 1–2 days for the synthesis and purification of oligonucleotide primers. Furthermore, Brenner and Livak (1989) have developed a different way for DNA fingerprinting. 5' overhangs of DNA fragments generated by cleavage with a class IIS restriction enzyme, such as *FokI* or *TthIII* or *AccI*, are labeled with a mixture of succinylfluorescent dideoxynucleotide triphosphates (SF-ddNTPs) using the reverse transcriptase under an appropriate buffer. And then, SF-ddNTP-labeled DNA samples are loaded on a denaturing polyacrylamide gel and analyzed for direct fingerprinting. This method seems to be simple but it is limited within a few of restriction enzymes as described and within 5' overhang of DNA fragments only.

We report here a rapid and simple method for DNA fingerprinting that involves fluorescence-labeled termini of any enzymatically generated fragments of DNA without a need for dye-primers or complementary synthetic oligonucleotides. Moreover, we have also developed a contig program using our own software. Overlap I, which represents a time-saving improvement over the usual software (Michael,

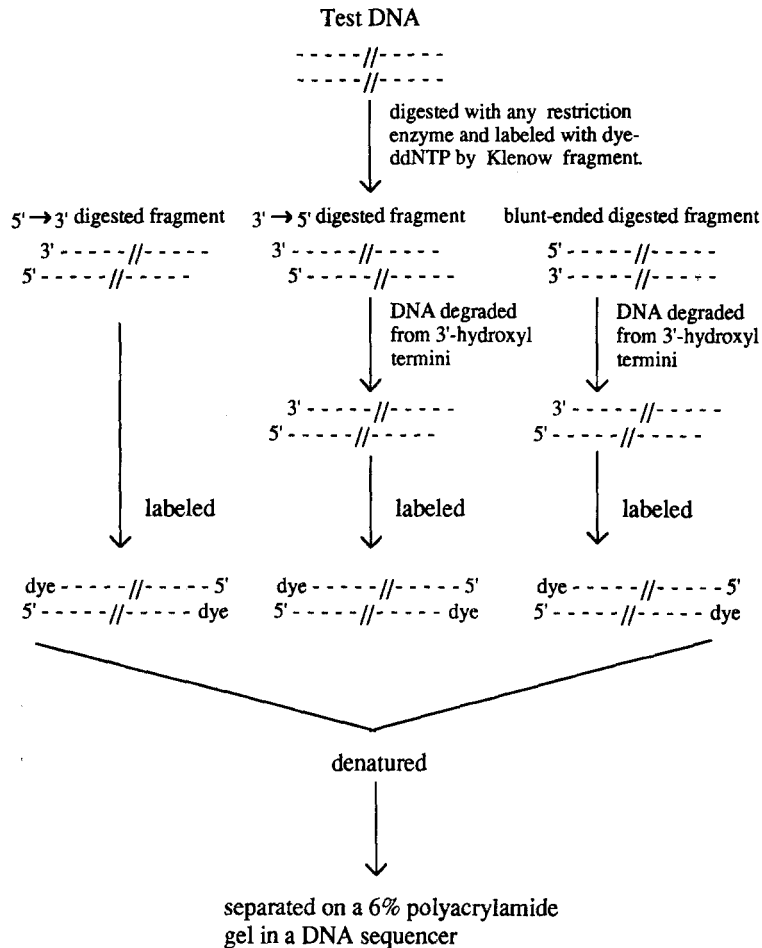


Fig. 1. Preparation of DNA fragments labeled with a dye-ddNTP. Any expected cosmid DNA can be used as "Test DNA." The reaction is initiated when uncut cosmid DNA, any restriction enzyme, the Klenow fragment of DNA polymerase I and dye-ddNTP are added simultaneously. After the reaction is completed, the mixture is purified by Quick Spin™ Column G-50 and denatured and then fractionated on a 6% polyacrylamide gel in a DNA sequencer (model 370A, Applied Biosystems).

1989), for simple contig-analysis of cloned DNAs. An example is presented the application of this method to the generation of a contig of cosmid DNAs from the mouse major histocompatibility complex (MHC) class I gene cluster.

MATERIALS AND METHODS

DNAs. A series of cosmid clones of the mouse MHC class I gene cluster was kindly provided by Dr. R. Flavell (Yale University, CT). The restriction map of the cosmid DNAs had already been determined, as shown in Fig 6 (Weiss *et al.*, 1984). Cosmid DNAs or pBR322 plasmid DNA as controls were prepared and purified by a standard method (Sambrook *et al.*, 1989).

Digestion, labeling, and mapping of DNAs. The details of the new fluorescence-based method for DNA fingerprinting are summarized in Fig. 1. Purified cosmid DNAs are digested with the restriction enzyme *EcoRI* (or *AluI* or *PstI*) and simultaneously and individually labeled with dye-ddATP (or dye-ddGTP or dye-ddCTP) (Applied Biosystems, Foster City, CA) at the 3' termini of the digested DNA fragments by the Klenow fragment of DNA polymerase I (Toyobo, Tokyo). The plasmid pBR322 was digested with *HaeIII* and labeled with dye-ddGTP. Each reaction was performed in a total of 20 μ l of reaction mixture that contained 1 μ g of DNA of the cosmid or plasmid pBR322, 0.5 pmol of dye-ddNTP (dye-ddATP or dye-ddGTP or dye-ddCTP), 2 μ l of enzyme buffer, 5 units of restriction enzyme (*EcoRI* or *HaeIII* or *AluI* or *PstI*) and 5 units of the Klenow fragment of DNA polymerase I from *Escherichia coli*. Reaction mixtures were incubated at 37°C for 1 hr. In some experiment labeled DNA fragments were purified by using Quick Spin™ Columns G-50 (Boehringer Mannheim GmbH, Mannheim, Germany) for exclusion of small molecular labeled oligonucleotides (<60 bp). Then they were denatured and separated on a 6% polyacrylamide gel. In the model experiment, the 15 to 600 bases of each DNA sample were resolved and analyzed on a DNA sequencer (model 370A, Applied Biosystems). The maximum size of DNA fragment for the size determination was dependent on the concentration of polyacrylamide. We estimated about 1.0-kb on a 6% polyacrylamide DNA-sequencing gel. Furthermore, the overlapping fragments of cosmid DNA were analyzed by use of a Version 0.3.8a mapper program (Applied Biosystems) on a Macintosh IIci computer.

Design of Overlap I software. The signal function, sign ($L_{i,j}$, $L_{k,h}$), with its symbols, was defined as follows:

N = the number of clones;

M_i = the number of restriction fragments in the insert of each recombinant DNA clone, where $1 \leq i \leq N$;

$L_{i,j}$ = the length of each restriction DNA fragment, where $1 \leq j \leq M_i$, $1 \leq i \leq N$;

$\xi_{i,k}$ = the number of overlapping DNA fragments between any two matched recombinant DNA clones, clone i and clone k ;

ϵ = a control parameter which provides an artificial compensation for errors in data;

γ = a control parameter, as the threshold value, that indicates the probability of

a contig between two recombinant DNA clones.

$$\text{sign}(L_{i,j}, L_{k,h}) = \begin{cases} 1 & \text{if } L_{i,j} = L_{k,h} \\ 0 & \text{if } L_{i,j} \neq L_{k,h} \end{cases} \quad (1)$$

where $1 \leq j, h \leq M_i, 1 \leq i, k \leq N$. The number of overlapping DNA fragments between clone i and clone k can be scored as follows:

$$\hat{\xi}_{i,k} = \sum_{1 \leq j \leq M_i} \sum_{1 \leq h \leq M_k} \text{sign}(L_{i,j}, L_{k,h}) \quad (2)$$

Moreover,

$$P_{i,k} = \hat{\xi}_{i,k} / M_i. \quad (3)$$

Clearly, if $P_{i,k} > \eta$, DNA fragments from clones i and k should be overlapped with a certain probability. Since the control parameter ε is used to avoid effects of data error, the length $L_{i,j}$ (2) can be placed with $L_{i,j} + \varepsilon$. Finally, Overlap I was encoded in Fortran-77 language (Michael *et al.*, 1989) on a VAX-4000 computer run with a VMS system. The flow chart of the Overlap I program is shown in Fig. 4.

RESULTS AND DISCUSSION

Schematic protocol

The new fluorescence-based method for DNA fingerprinting is shown in Fig. 1. Since it is well known that the termini of enzymatically prepared DNA fragments can be labeled with the corresponding dye-ddNTP by the Klenow fragment of DNA polymerase I (Brenner and Livak, 1989; Klenow and Henningsen, 1970), we postulated that the restriction and labeling reactions could be performed simul-

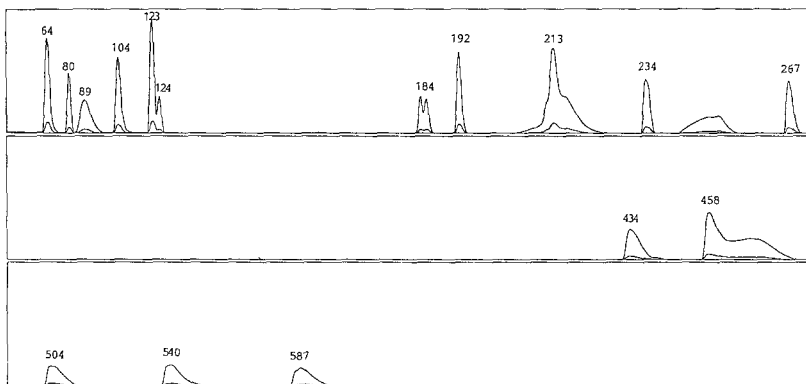


Fig. 2. pBR322 plasmid DNA, as the standard marker, was digested with *Hae*III and labeled with dye-ddGTP and fractionated on a 6% acrylamide gel at 1410 V, 20 mA, 26 W, and 47°C for 12 hr. The exact length (base) of every fragment is given above each peak.

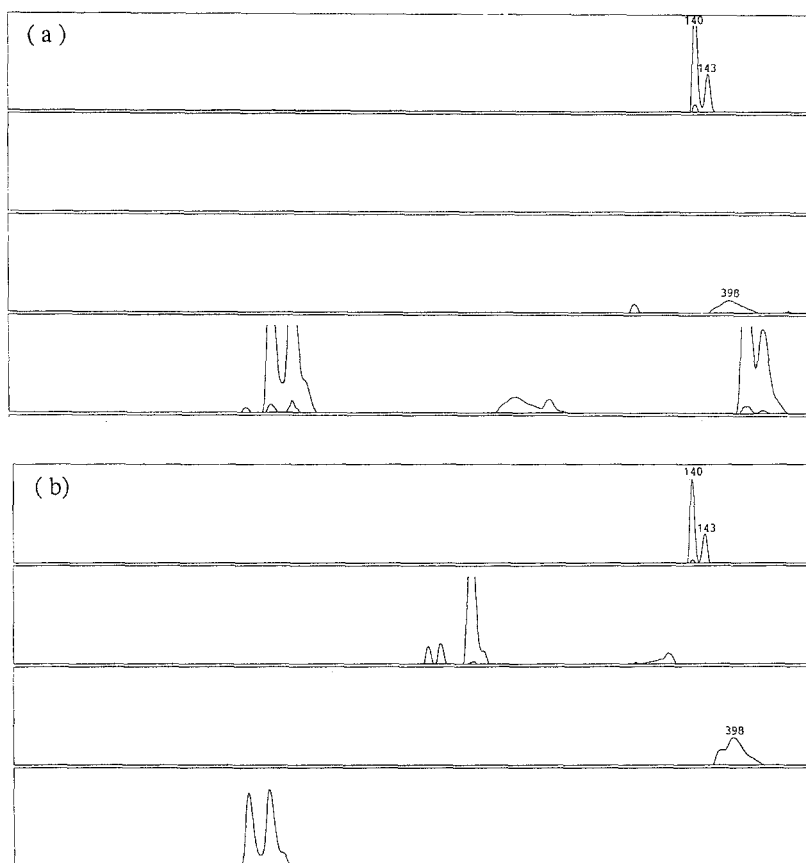
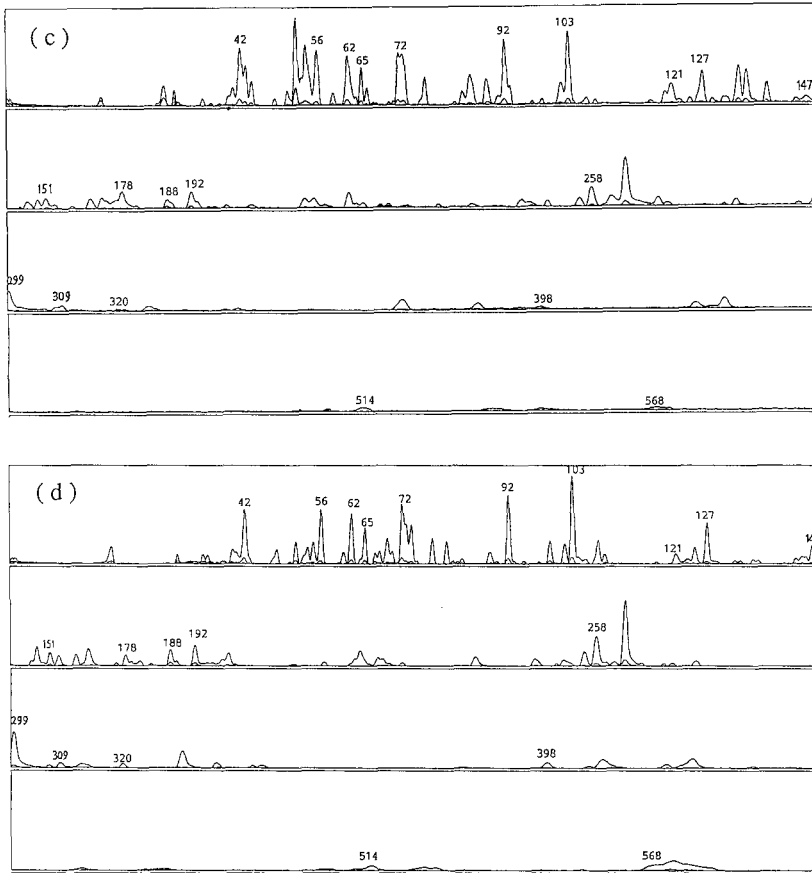


Fig. 3. Processed data after electrophoresis on a polyacrylamide (6%) gel of cosmid DNAs. *Eco*RI-generated DNA fragments of H21 (a) and H24 (b) or *Alu*I-generated DNA fragments of H21 (c) and H24 (d) were labeled with dye-ddATP or dye-ddGTP. Conditions for electrophoresis were 1410 V, 20 mA, 26 W and

taneously because the ligation of dye-ddNTP at the termini of restricted DNA fragments destroys the recognition site of the restriction enzyme. Moreover, the dye-ddNTP-labeled fragments of DNA can be well separated and are detectable with a DNA sequencer.

Mapper software

After the collection of data from the DNA sequencer, the size of each fragment is calculated by the Version 0.3.8a mapper program (Applied Biosystems) and the possibility of overlap between DNA fragments is automatically shown on computer by application of our Overlap I program. In order to accelerate analysis of contigs from a large number of cloned DNAs, we have developed the analytical program, Overlap I, in Fortran-77 programming language (Michael, 1989).



47°C. *Hae*III-generated pBR322 DNA was used as a size standard. Only matching DNA fragments of the same sizes in both H21 and H24 are labeled with their numbers of bases, as determined from the size standards.

The Overlap I program is divided into four main steps: Input, Data processing, Overlap search, and Output (Fig. 4). The experimental data for DNA fragments of different sizes that are collected by the Version 0.3.8a mapper program on the DNA sequencer, are read by Input, and a control parameter is also set by Input. The relative length of every DNA fragment is then determined by the Data processing step. All of the fragments with different, relative lengths in each cloned DNA are lined up from the shortest fragment to the longest one. The line-ups from all clones are converted to a matrix. The step Overlap search is the key to this program. The signals due to overlapping between any two matched DNA fragments are detected. The program for such matching is summarized in Fig. 5. $L_{i,j}$ equals the length of each restricted DNA fragments, where $L_{i,j} \geq L_{i,j+1}$; if $(L_{i,j_0} - L_{i+1, k_0-1}) \times (L_{i,j_0} - L_{i+1, k_0}) \leq 0$, then L_{i,j_0+1} will be matched with the DNA

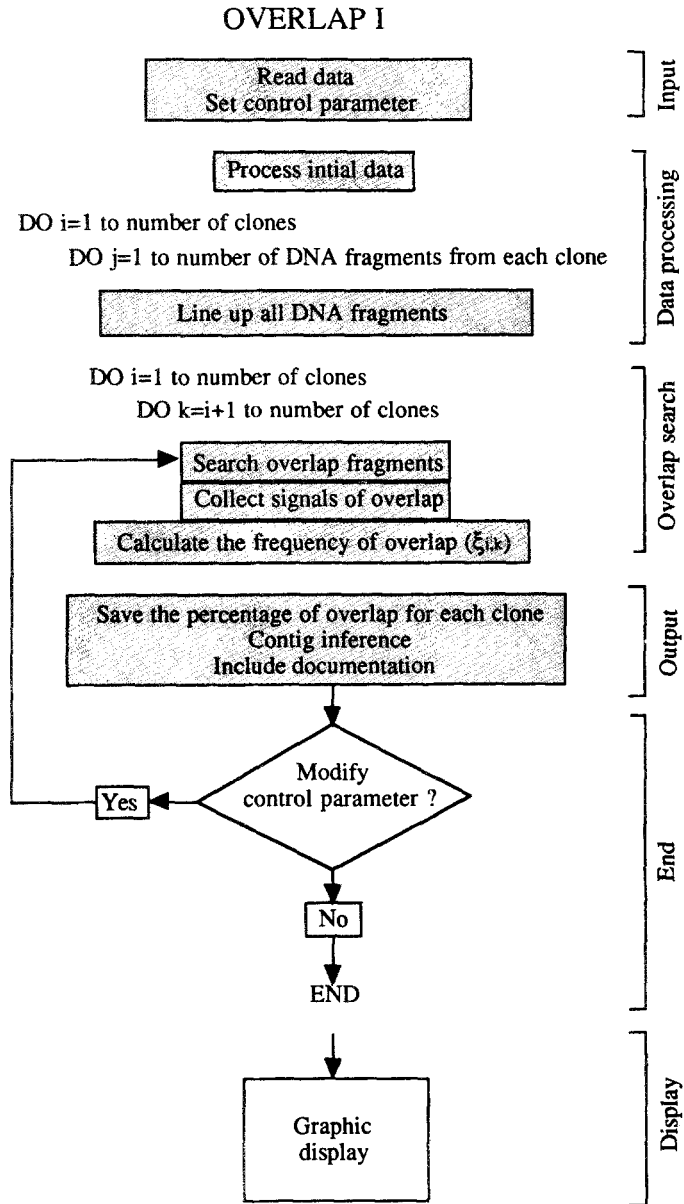


Fig. 4. Flow chart for the Overlap I program.

fragment that contains a longer length of DNA than L_{i+1, k_0-1} . Namely, for matching of both clone i and clone $i+1$ (Fig. 5, as arrow 1), in the case where DNA fragment j_0 of clone i contains the same length of DNA as DNA fragment k_0-1 of clone $i+1$ (as arrow 2), the fragment j_0 of clone i does not match any DNA frag-

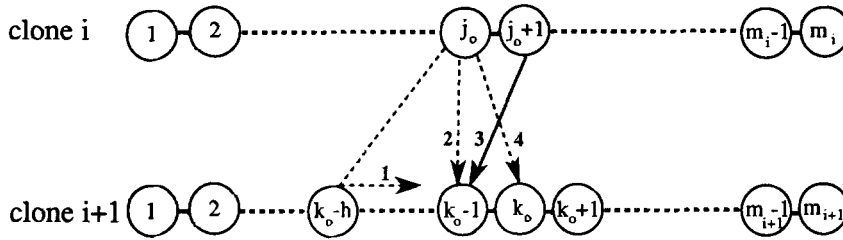


Fig. 5. The mathematical model for matching of DNA fragments between two clones i and $i+1$. All DNA fragments are lined up from the left side to the right according to their relative sizes, such as fragments $1 < 2 < j_0 < j_0+1 < m_i-1 < m_i$ of clone i or $1 < 2 < k_0-h < k_0-1 < k_0 < k_0+1 < m_{i+1}-1 < m_{i+1}$ of clone $i+1$. Thus, fragment 1 is the shortest while m_i and m_{i+1} are longer than all the others. Dotted arrow 1 indicates that a DNA fragment j_0 of clone i matches fragments one by one between k_0-h and m_{i+1} in clone $i+1$. Dotted arrows 2 and 4 indicate that j_0 of clone i is matched to fragments of clone $i+1$ between k_0-1 and k_0 . The solid arrow 3 indicates that j_0+1 which is longer than j_0 does not match fragments shorter than k_0-1 of clone $i+1$ when j_0 is longer than k_0-1 of clone $i+1$.

ments with longer lengths of DNA than the DNA fragment k_0-1 of clone $i+1$. Equally, if the length of j_0 is longer than k_0-1 but shorter than k_0 of another clone $i+1$ (arrows 2 and 4), then j_0 does not match any DNA fragments with longer lengths of DNA than k_0 of clone $i+1$. Naturally, j_0+1 of clone i does not match fragments with shorter lengths of DNA than k_0-1 of clone $i+1$ when $L_{i,j_0} \geq L_{i+1,k_0} \geq L_{i+1,k_0-1}$, as described above. Thus, the matching of DNA fragments among clones can be scanned without irrational matching so that the time required for matching can be efficiently reduced. Finally, the frequencies of matches are automatically calculated by Overlap search step and the Output step is, in fact, the input of other routines that are used to display the results graphically.

Fingerprinting of DNAs

To verify this new method for labeling the blunt termini of DNA restriction fragments with dye-ddNTP by the Klenow enzyme, we used pBR322 plasmid DNA digested with *Hae*III and analyzed its DNA fingerprint. Indeed, *Hae*III-generated DNA fragments of pBR322 were completely labeled by dye-ddGTP after only a one-hour incubation and their fluorescent peaks were well separated with appropriate mobilities on a gel. Clearly, as shown in the case of the blunt ends of DNA fragments created by *Hae*III, the 3'-hydroxyl termini of the DNA fragments can be replaced and labeled with a dye-ddGTP by the Klenow fragment of DNA polymerase I (Fig. 2). Thus, we conclude that any DNA fragments without a fluorescent linker can easily be used for DNA fingerprinting after fluorescent labeling by the Klenow fragment, which has 3'→5' exonuclease activity as well as its main 5'→3' polymerase activity. This experimental protocol can also be adapted for the use of DNA fragments with protruding 3' ends to make a contig map of indicated DNA

clones (Fig. 1) although we sometimes detect the apparent heterogeneity of individual bands, which may be due to ragged nature of exonuclease/polymerase activity.

We examined the validity of our new method using cosmid contig clones derived from the mouse MHC class I gene cluster (Weiss *et al.*, 1984) and compared the relative overlap frequency of each cosmid clone generated by our new method

Table 1. Comparative study of the overlapping frequency of MHC class I cosmid DNAs.

No	Overlapped pairs of cosmid DNAs (name / name)	This paper			Carrano's* method
		labeling **		Frequency of overlapping (%)	Frequency of overlapping (%)
		enzyme	ddNTP		
1	H21 / H24	<i>AluI</i>	ddGTP	30	25
2	H21 / H24	<i>EcoRI</i>	ddATP	30	25
3	B4.15 / B3.2	<i>EcoRI</i>	ddATP	23	17
4	B3.2 / Bm11.25	<i>EcoRI</i>	ddATP	34	33.5
5	Bm11.25 / B1.24	<i>EcoRI</i>	ddATP	11.7	9.6
6	B1.24 / H26	<i>Pst I</i>	ddCTP	15.3	12
7	H26 / B25	<i>Pst I</i>	ddCTP	10.6	12.6
8	B25 / H19	<i>Pst I</i>	ddCTP	14	11
9	H19 / B2.17	<i>Pst I</i>	ddCTP	18	14.8
10	B2.17 / B2.6	<i>EcoRI</i>	ddATP	36	32
11	B26 / B1.4	<i>EcoRI</i>	ddATP	24	26
12	B1.4 / LS AW3	<i>EcoRI</i>	ddATP	5.6	8
13	H4 / H27	<i>AluI</i>	ddGTP	38	36.6
14	H27 / H18	<i>AluI</i>	ddGTP	22.6	18.5
15	H18 / S14	<i>AluI</i>	ddGTP	43.8	42
16	S14 / LSK14	<i>AluI</i>	ddGTP	41.5	41
17	LSK14 / B2.7	<i>AluI</i>	ddGTP	25	26
18	B2.7 / H6	<i>AluI</i>	ddGTP	42	37
19	H24 / B2.6	<i>EcoRI</i>	ddATP	1.1	1.5
20	H4 / B4.15	<i>EcoRI</i>	ddATP	1.3	1.3

* A single-stranded oligonucleotide primer (18-mer, dye-TCCAGTCACGACGTTGT) with a complementary synthetic oligonucleotide (20-mer, GGTCAGTGCTGCAACAT TAA) served to create a double-stranded oligonucleotide linker that was complementary to the *EcoRI* restriction site. The fluorescent linker was ligated to the *EcoRI*-digested cosmid DNAs. All experimental protocols are described elsewhere (Carrano *et al.*, 1989).

** Labeling with dye-ddNTP of termini of enzymatically prepared cosmid DNA was performed by the Klenow fragment of DNA polymerase I.

and by the standard method (Carrano *et al.*, 1989). Results of a representative study are shown in Fig. 3, a and b. Two cosmid DNAs, H21 and H24, which covered the H-2K region of the MHC class I locus, were examined for overlapping. Nine *EcoRI*-generated DNA fragments prepared from each of the H21 and H24 clones gave clear peaks (Fig. 3, a and b). Three DNA fragments of 140 bases,

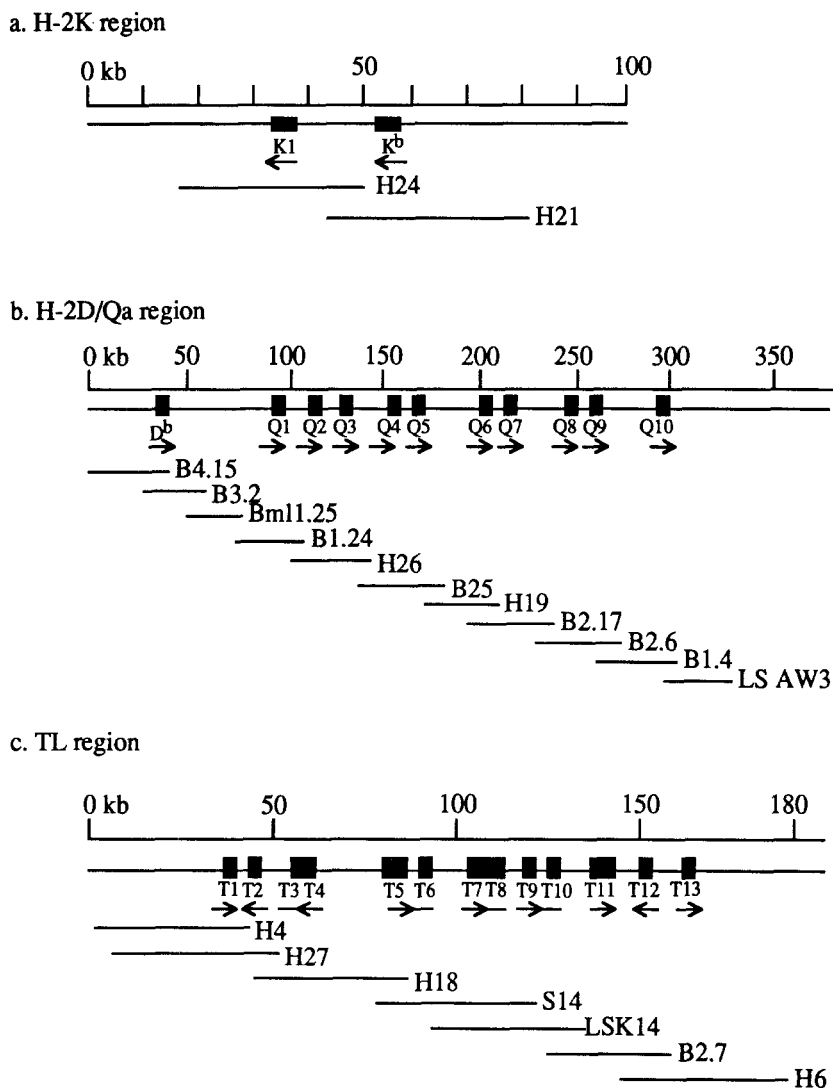


Fig. 6. Schematic representative sets of overlapping cosmids defining the H-2K, H-2D/Qa, and TL regions are shown. The length of each region (kb) is shown on the top line. MHC class I genes are shown as solid boxes, with arrows indicating the direction of transcription (5'→3'). a, H-2K region; b, H-2D/Qa region; c, TL region. Names of cosmids are shown.

143 bases and 398 bases were observed to be overlapping and the frequencies of overlapping between them was 30%. Similarly, 21 of 70 *AluI*-generated DNA fragments of 42, 56, 62, 65, 72, 92, 103, 121, 127, 147, 151, 178, 188, 192, 258, 299, 309, 320, 398, 514, and 568 bases from both H21 and H24 were identical and the frequencies of overlapping were 30% (Figs. 3, c and d). Using our new method, we compared the overlapping frequencies of cosmid clones of the MHC class I gene cluster (Table 1) with those reported by Carrano *et al.* (1989). It is evident that the overlap frequencies of cosmid-contigs generated by both methods are almost identical, the exceptions being some contig-clones such as H4/H27 and B2.7/H6. The very much lower frequency of overlapping of clones H24/B2.6 or H4/B4.15 as negative controls was detected because each did not contain any overlapping DNA fragments from the other (Table 1).

We demonstrated clearly that cosmid DNAs can be easily examined for overlapping DNA fragments by application of our new fluorescence-based method (Fig. 3 and Table 1). Moreover, it appears that not only can we apply this simple method to any enzymatically generated fragments of DNA for DNA fingerprinting, but we can also save a lot of time. There is no need to purchase a commercial fluorochrome-labeled linker or to synthesize an appropriate oligonucleotide primer, and only one hour is required for the complete fluorescent-labeling reaction rather than 8–12 hr (Carrano *et al.*, 1989). Also we can develop the facility of this method to use multiple dyes for processing more DNA samples at once or for confirming pattern matches by mixing DNA samples simultaneously. The data collected from the DNA sequencer can be fed directly to a computer and matching DNA fragments can be analyzed simply and automatically with our Overlap I program.

Acknowledgments We thank Drs. A. V. Carrano, M. Raff, and M. W. Hunkapiller for helpful discussions and useful advice. This work was supported by the Special Coordination Funds of the Science and Technology Agency of the Japanese Government and by the Life Science Research Project of RIKEN.

REFERENCES

- Brenner S, Livak KJ (1989): DNA fingerprinting by sampled sequencing. *Proc Natl Acad Sci USA* **86**: 8902–8906
- Calladine CR, Collis CM, Drew HR, Mott MR (1991): A study of electrophoretic mobility of DNA in agarose and polyacrylamide gels. *J Mol Biol* **221**: 981–1005
- Carrano AV, Lamerdin J, Ashworth LK, Watkins B, Branscomb E, Slezak T, Raff M, deJong PJ, Keith D, Meister S, Kronick MT (1989): A high-resolution fluorescence-based, semiautomated method for DNA fingerprinting. *Genomics* **4**: 129–136
- Klenow H, Henningsen I (1970): Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from *Escherichia coli* B by limited proteolysis. *Proc. Natl Acad Sci USA* **65**: 168–175
- Michael M (1989): *Effective Fortran 77*. Oxford University Press, NY
- Olson MV, Dutchik JE, Graham MY, Brodeur GM, Frank M, MacCollin M, Scheinman R, Frank T (1986): Random-clone strategy for genomic restriction mapping in yeast. *Proc Natl Acad Sci USA* **83**: 7826–7830

- Sambrook J, Fritsch EF, Maniatis T (1989): Molecular cloning: A Laboratory manual, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Smith CL, Econome JG, Schutt A, Kico S, Cantor CR (1987): Physical map of the *Escherichia coli* K12 genome. *Science* **236**: 1448-1453
- Smith LM, Sanders JZ, Kaiser RJ, Hughers P, Dodd C, Connell CR, Heine C, Kent SBH, Hood L (1986): Fluorescence detection in automated DNA sequence analysis. *Nature* **321**: 674-679
- Sulston J, Mallett F, Staden R, Durbin R, Horsnell T, Coulson A (1988): Software for genome mapping by fingerprinting techniques. *Cabios* **4**: 125-132
- Weiss EH, Golden L, Farner K, Mellor AL, Devlin JJ, Bulman H, Tiddens H, Bud H, Flavell RA (1984): Organization and evolution of the class I gene family in the major histocompatibility of the C57BL/10 mouse. *Nature* **310**: 650-655