Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 9523–9527, November 1991
Genetics

A genomic scanning method for higher organisms using restriction sites as landmarks

(two-dimensional electrophoresis/dideoxy α-thio analogue of nucleotide/human/mouse)

IZUHO HATADA, YOSHIHIDE HAYASHIZAKI*, SHINJI HIROTSUNE, HIDEYUKI KOMATSUBARA, AND TSUNEHIRO MUKAI

Department of Bioscience, National Cardiovascular Center Research Institute, 5-7-1, Fujishiro-dai, Suita, Osaka 565, Japan

Communicated by Takashi Sugimura, July 19, 1991 (received for review March 2, 1991)

ABSTRACT **We have developed a powerful genomic scanning method, termed "restriction landmark genomic scanning," that is useful for analysis of the genomic DNA of higher organisms using restriction sites as landmarks. Genomic DNA is radioactively labeled at cleavage sites specific for a rare cleaving restriction enzyme and then size-fractionated in one dimension. The fractionated DNA is further digested with another more frequently occurring enzyme and separated in the second dimension. This procedure gives a two-dimensional pattern with thousands of scattered spots corresponding to sites for the first enzyme, indicating that the genome of mammals can be scanned at ≈1-megabase intervals. The position and intensity of a spot reflect its locus and the copy number of the corresponding restriction site, respectively, based on the nature of the end-labeling system. Therefore, this method is widely applicable to genome mapping or detection of alterations in a genome.**

Genomic DNA analysis is essential for clarifying the characteristics of mammalian DNA. However, the genomes of these organisms are very large. For example, the mammalian genome is about 3×10^9 base pairs (bp), which is 1000 times that of *Escherichia coli*. The first step in analyzing large genomes requires the scanning of many landmarks. Southern blotting has been used for this purpose (1); however, usually only one locus on the genome can be detected with a single probe. Therefore, when this method is applied to scanning a genome, it must be repeated many times with many probes. Several Southern blot repetitions might be theoretically sufficient for some purposes, if a repetitive sequence is used as a probe (2, 3). However, bands (spots) corresponding to loci cannot be separated in these systems.

In this paper, we introduce a concept termed "restriction landmark," in which each restriction enzyme recognition site can be used as a landmark. Based on this concept, we developed a restriction landmark genomic scanning (RLGS) method, which employs (*i*) direct end labeling of the genomic DNA digested with a restriction enzyme and (*ii*) high-resolution, two-dimensional electrophoresis. Using this method, we simultaneously separated and detected thousands of signals (spots) derived from restriction sites. Thus, we could locate landmarks on mammalian genomes at intervals averaging ≈1 megabase pair (Mbp).

MATERIALS AND METHODS

DNA Preparation. Genomic DNA was extracted from each sample as described (4). The extracted DNA was electrophoresed through a 0.5% agarose gel in 1× TBE buffer (50 mM Tris-borate, pH 8.3/1 mM EDTA) to determine the average fragment size and RNA contamination.

Procedure for Two-Dimensional Gel Electrophoresis. Two different procedures were used depending on whether the DNA end to be labeled was the 5′ or 3′ protruding end. The following steps were followed for the 5′ protruding end. (*i*) Ten micrograms of genomic DNA was allowed to react for 30 min at 37°C with 10 units of DNA polymerase I in 50 μl of 50 mM Tris·HCl (pH 7.4), 100 mM NaCl, 10 mM dithiothreitol, 0.33 μM 2′-deoxyribonucleoside 5′-[α-thio]triphosphate {dXTP[αS] (Amersham), which can be incorporated into the cleavage site of restriction enzyme A by fill-in reaction; for *Mlu* I digestion, dGTP[αS] and dCTP[αS] were used} and 33 μM 2′,3′-dideoxyribonucleoside 5′-[α-thio]triphosphate {ddYTP[αS], which cannot be incorporated into the cleavage site of restriction enzyme A by fill-in reaction; for *Mlu* I digestion, ddATP[αS] and ddTTP[αS] were used; these reagents were synthesized for this experiment by Toyobo (Osaka)}. Thereafter, the enzyme was inactivated at 65°C for 30 min. (*ii*) The treated DNA was then digested with 100 units of restriction enzyme A for 1 hr in 85 μl of the reaction buffer appropriate for restriction enzyme A. (*iii*) The cleavage ends were filled in with 20 units of Sequenase version 2.0 (United States Biochemical) in the presence of 0.33 μM 2′-deoxyribonucleoside [α-^{32}P]triphosphate ([α-^{32}P]dXTP, which can be incorporated into the cleavage site of restriction enzyme A with Sequenase version 2.0) (3000–6000 Ci/mmol; 1 Ci = 37 GBq) for 30 min at 37°C in 100 μl of 50 mM Tris·HCl (pH 7.4), 100 mM NaCl, 10 mM dithiothreitol, 0.16 μM dXTP[αS], and 33 μM ddYTP[αS]. To inactivate the enzyme, this reaction mixture was incubated at 65°C for 1 hr. When the average size of the DNA fragments was more than several hundred kilobase pairs, an additional digestion was performed using restriction enzyme B. (*iv*) One microgram of the DNA from step *iii* was fractionated on a 50 × 20 × 0.1 cm agarose gel (0.8–1% Seakem GTG agarose; FMC) and then electrophoresed in 1× TAM buffer (50 mM Tris-acetate, pH 7.5/0.7 mM magnesium acetate) at 4.5 V/cm for 12 hr. (*v*) The DNA-containing portion of the gel was excised as a strip and soaked for 30 min in the reaction buffer appropriate for restriction enzyme C. Thereafter, DNA was digested in the gel with 1500 units of restriction enzyme C for 2 hr. (*vi*) The gel was fused with a 50 × 50 × 0.1 cm polyacrylamide gel (5–6% polyacrylamide to acrylamide/bisacrylamide, 29:1) by adding melted agarose to fill up the gap. Second-dimensional electrophoresis was carried out in 1× TBE buffer at 8 V/cm

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: RLGS, restriction landmark genomic scanning; Mbp, megabase pair; dXTP[αS], 2′-deoxyribonucleoside 5′-[α-thio]-triphosphate; ddYTP[αS], 2′,3′-dideoxyribonucleoside 5′-[α-thio]triphosphate; dXTP, 2′-deoxyribonucleoside triphosphate; ddTTP[αS], 2′,3′-dideoxythymidine 5′-[α-thio]triphosphate; I_{ns}, normalized spot intensity(ies). For the use of dXTP, dXTP[αS], and ddYTP[αS], see text.

*To whom reprint requests should be addressed.

for 6 hr. (*vii*) The gel was dried. An area 35 × 43 cm of the original gel was then excised and autoradiographed for 3–10 days on a film (XAR-5; Kodak) at −70°C using an intensifying screen (Quanta III; DuPont). The same procedures were used for the 3′ protruding or blunt end labeling, except for steps *i* and *iii*. (*i*) The DNA was allowed to react with 25 units of terminal deoxynucleotidyltransferase (Toyobo) in the presence of 10 μM 2′,3′-dideoxythymidine 5′-[α-thio]triphosphate (ddTTP[αS], preprepared for this experiment by Toyobo) for 30 min at 37°C. (*iii*) The cleavage ends were labeled by reaction with 25 units of terminal deoxynucleotidyltransferase in the presence of 0.8 μM [α-^{32}P]ddATP (5000 Ci/mmol) for 30 min at 37°C. In these steps, the reactions were carried out in 140 mM sodium cacodylate (pH 7.0), 1 mM $CoCl_2$, 50 μg of bovine serum albumin per ml, and 0.1 mM dithiothreitol.

Quantification of the Spot Intensity on the Two-Dimensional Profile. Spot intensities in the two-dimensional profile were measured by PDQUEST (Protein Database, New York). The calibration of the spot density to dpm/mm^2 was performed using a set of calstrips with serially diluted radioactivities. The spot intensity was calculated according to the following formula: $I = \pi h \sigma_x \sigma_y$, where I is the spot intensity, h is the maximum dpm/mm^2, and σ_x and σ_y are the half-widths of the spot. [Details of the measurement and calculation of the spot intensity have been described by Garrels (5).] The normalized spot intensities (I_{ns}) were calculated from $I_{ns} = I_s/I_c$, where I_s is the intensity of the target spot and I_c is the intensity of the control spot (for example, spot C in Fig. 3*B* and spot MP153 in Fig. 4*B*), which is considered to be a single copy spot.

RESULTS AND DISCUSSION

Strategy for Two-Dimensional Gel Electrophoresis by RLGS. This is comprised of seven steps (Fig. 1). (*i*) Blocking. Generally, genomic DNA is cleaved in the preparation step and thus has nonspecific cleaved ends, nicks, and/or gaps. This leads to high background caused by the incorporation of radioisotopes into sites damaged in the labeling process. Such incorporation can be avoided by blocking those sites with enzymatically incorporated new nucleotide analogues (ddYTP[αS]), because such analogues prevent exonucleolysis and/or the additional incorporation of the nucleotide at blocked ends. (*ii*) Landmark cleavage (with restriction enzyme A). Blocked DNA is cleaved with a rare cleaving restriction enzyme (restriction enzyme A; average fragment size, >100 kbp). (*iii*) Labeling. Cleavage ends of the genomic DNA are labeled with the radioisotope. Depending on the average size of the DNA fragments generated by restriction enzyme A, this step is followed by fragmentation with restriction enzyme B, which gives a higher cleavage frequency than that with restriction enzyme A (restriction enzyme B; average fragment size, from several to scores of kilobase pairs). (*iv*) First fractionation. DNA restriction fragments are fractionated in one dimension by thin-layer agarose gel electrophoresis. (*v*) Fragmentation of labeled DNA with restriction enzyme C. Fractionated genomic DNA fragments are cleaved in the gel with restriction enzyme C (average fragment size, <10 kbp). (*vi*) Second fractionation. DNA restriction fragments in the agarose gel are fractionated in the second dimension by polyacrylamide gel electrophoresis. (*vii*) Autoradiography. The genomic DNA fragments are detected.

Application of RLGS to Higher Organisms. Fig. 2 shows spot profiles of human DNA obtained by RLGS. This profile contains about 2500 completely separated spots, indicating that the human genome was scanned at mean intervals of 1.2 Mbp, because the size of the human genome is 3 × 10^9 bp. The position and intensity of the spot were reproducible,

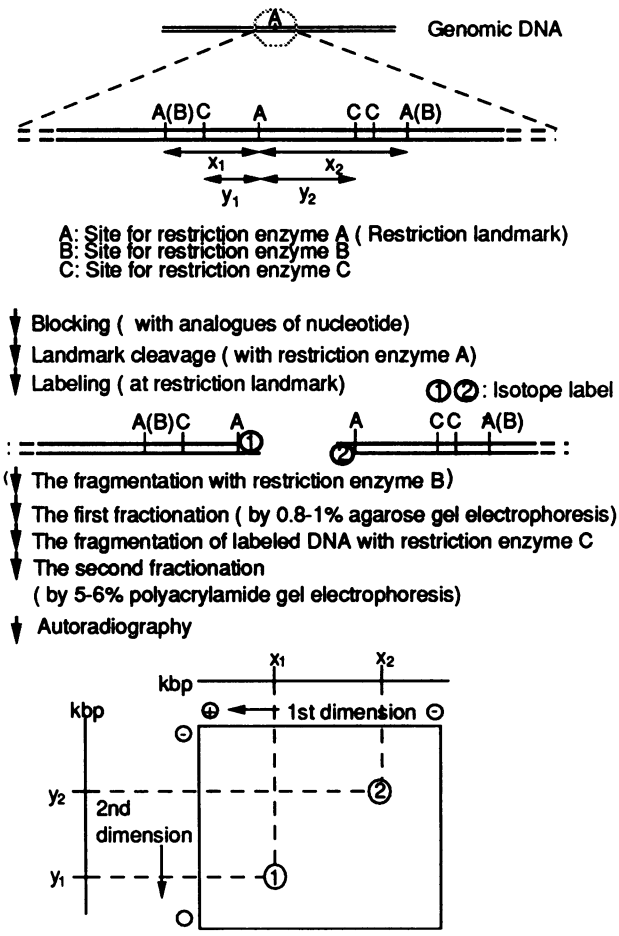

FIG. 1. Procedure for genome scanning by two-dimensional gel electrophoresis. RLGS consists of seven steps: (*i*) blocking, (*ii*) landmark cleavage (with restriction enzyme A), (*iii*) labeling (and the fragmentation of labeled DNA with restriction enzyme B), (*iv*) first fractionation, (*v*) fragmentation of labeled DNA with restriction enzyme C, (*vi*) second fractionation, and (*vii*) autoradiography. x_1 or x_2 and y_1 or y_2 represent the distance from a restriction landmark to the neighboring site for restriction enzyme A or B and that to the site for restriction enzyme C, respectively. kbp, Kilobase pairs.

except when partial digests were subjected to RLGS. We confirmed the complete digestion, since a 10- to 20-fold excess of the enzyme for digestion gave the same RLGS pattern. Similarly, *Drosophila* and mouse genomes were also subjected to RLGS, as shown in Figs. 3 and 4, respectively. The *Drosophila* genome gave 1000 spots with *Mlu* I and a mouse genome gave 2000 spots with *Pac* I, indicating that these genomes were scanned at mean intervals of 150 kbp and 1.5 Mbp, respectively.

Restriction enzymes are sensitive to site-specific methylation. However, to use the spots as the signal of landmarks on the genome, the detection of restriction landmarks should not be affected by DNA modification, such as methylation. This problem can be overcome by using a restriction enzyme insensitive to methylation. For example, as the only known site for the methylation is position 5 of the cytosine residue in vertebrates (6), a rare cleaving enzyme not containing the cytosine residue, such as *Pac* I (TTAATTAA) or *Swa* I (ATTTAAAT), or enzymes insensitive to 5-methylcytosine should be used.

Each Spot Represents the Locus and Its Copy Number on the Genome. We examined whether the position and density of a spot correspond to its locus and copy number, respectively. The white (*w*) gene of *Drosophila melanogaster* is a single-copy gene responsible for the deposition and distribution of

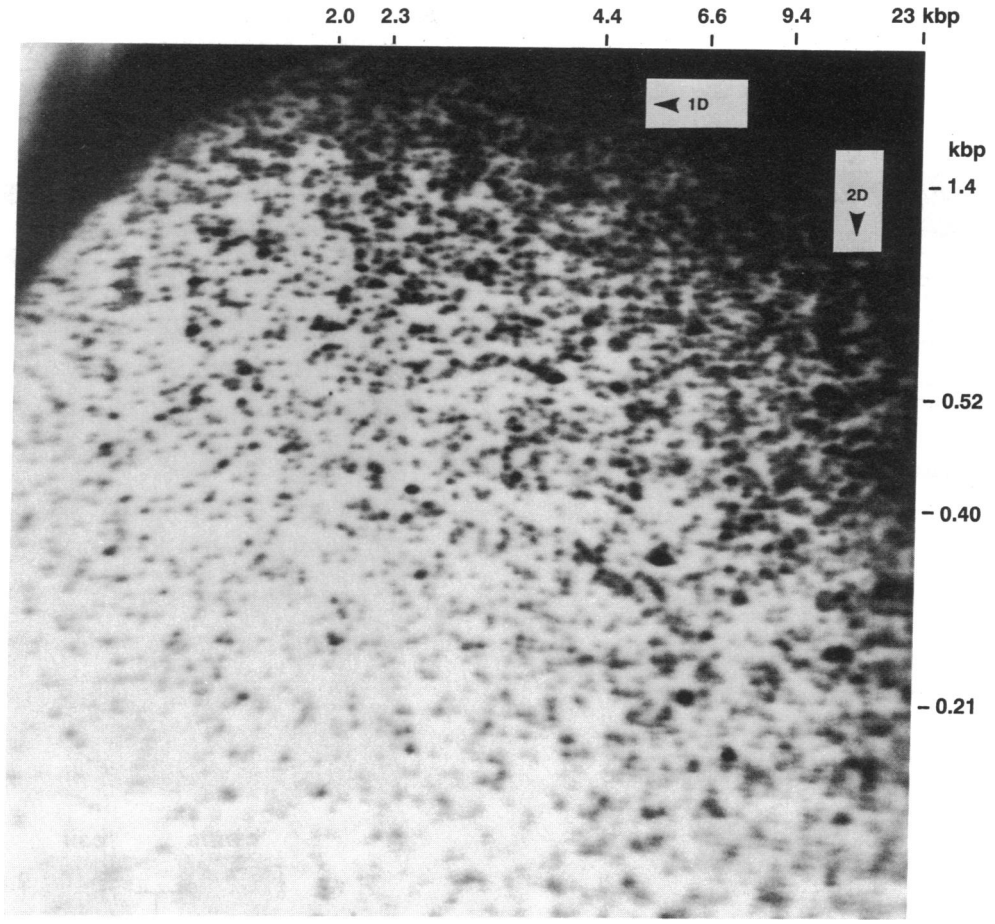

FIG. 2. Autoradiographic profile of RLGS analysis for human genomic DNA. Human genomic DNA was treated with 25 units of terminal deoxynucleotidyltransferase in the presence of 10 μM ddTTP[αS] for 30 min. *Pac* I was then used as restriction enzyme A. The cleavage ends were labeled by reaction with 25 units of terminal deoxynucleotidyltransferase in the presence of 0.8 μM [α-^{32}P]ddATP (5000 Ci/mmol) for 30 min. In this case, the cleavage by restriction enzyme B was not done because the *Pac* I fragments were short. *Eco*RV was used as restriction enzyme C. About 2500 spots were obtained. These spots correspond to 4% of all *Pac* I sites, assuming that *Pac* I sites exist at a frequency of 62,500 in a human genome. The remainder of a *Pac* I site is retained in the dark area of the upper side of this figure. The first dimension (1D) is right to left and the second dimension (2D) is top to bottom. Size markers are shown in kbp. The photograph covers the 12- to 45-cm region of the original gel in the first dimension and the 8- to 40-cm region in the second dimension.

eye pigment (7). A deletion mutation (w^-) results in the absence of eye pigment. A wild type (+/+), a heterozygous mutant ($w^-/+$), and a homozygous mutant (w^-/w^-) had 1, 0.5, and 0 copies per haploid genome, respectively, as confirmed by Southern blot analysis (Fig. 3*D*). As the restriction site for *Mlu* I is located on the deleted region of the w^- allele, we used the *Mlu* I recognition site as a landmark. To identify the spot corresponding to this landmark, *Drosophila* genomic DNA mixed with several copies of cloned *w* gene DNA was subjected to RLGS. As shown in Fig. 3*C*, this clone gave a higher intensity spot in the position identical to that seen in the wild-type genomic DNA. On the other hand, this spot was not found in the profile of homozygous mutant (w^-/w^-) DNA. These findings revealed the one-to-one correspondence of this spot and the locus. The I_{ns} of the wild type (+/+), the heterozygote ($w^-/+$), and the homozygote (w^-/w^-) were 1.04 $\pm$ 0.02 (n = 6), 0.48 $\pm$ 0.03 (n = 6), and 0 (no spot was detected by PDQUEST in this area; n = 6) (mean $\pm$ SEM), respectively, which are close to 1, 0.5, and 0 (see *Materials and Methods* for the calculation of I_{ns}). This indicates that the I_{ns} reflects the copy number. Thus, RLGS can detect heterozygously or homozygously deleted regions on the genomic DNA. This advantage is brought about by the end-labeling procedure, in which the radioisotope should be incorporated in proportion to the end number of the DNA fragment.

To further confirm the correspondence of a spot and a locus, we examined whether a spot would follow Mendelian law in the same manner as observed in a locus. We analyzed the transmission pattern of the spot in a mouse pedigree originating from two inbred strains, C57B/6 and C3H/HeJ. Fig. 4 shows the two-dimensional profiles of RLGS and the transmission pattern of the C3H-specific polymorphic spot (spot P), which is a single copy per haploid genome. Spot P is thought to be transmitted from F_0 through F_1 by Mendelian law, because its I_{ns} in F_1 is always half that in F_0, depending on the copy number (F_1, 0.46 $\pm$ 0.01, n = 6; F_0, 1.02 $\pm$ 0.02, n = 6). In F_2, this spot exhibited three types of I_{ns} in proportion to its copy number—namely, 0 (n = 3), 0.46 $\pm$ 0.01 (n = 5), and 1.02 $\pm$ 0.02 (n = 5). Thus, we proved that the spots on the two-dimensional profile segregate according to Mendelian rules in the same manner as observed in their corresponding loci on the genome.

Advantages and Potential Use of RLGS Analysis. RLGS has the following advantages. (*i*) It can be applied to studies on higher organisms, because of its high sensitivity and resolution. (*ii*) Thousands of restriction landmarks can be scanned simultaneously. In addition, the scanning field can be extended by the use of different kinds of landmarks (restriction enzyme A) in an additional series of electrophoresis. In addition to *Mlu* I and *Pac* I, other 8-bp and rare cleaving enzymes, such as *Not* I, *Bss*HII, *Asc* I, and *Sse*8387I, can be

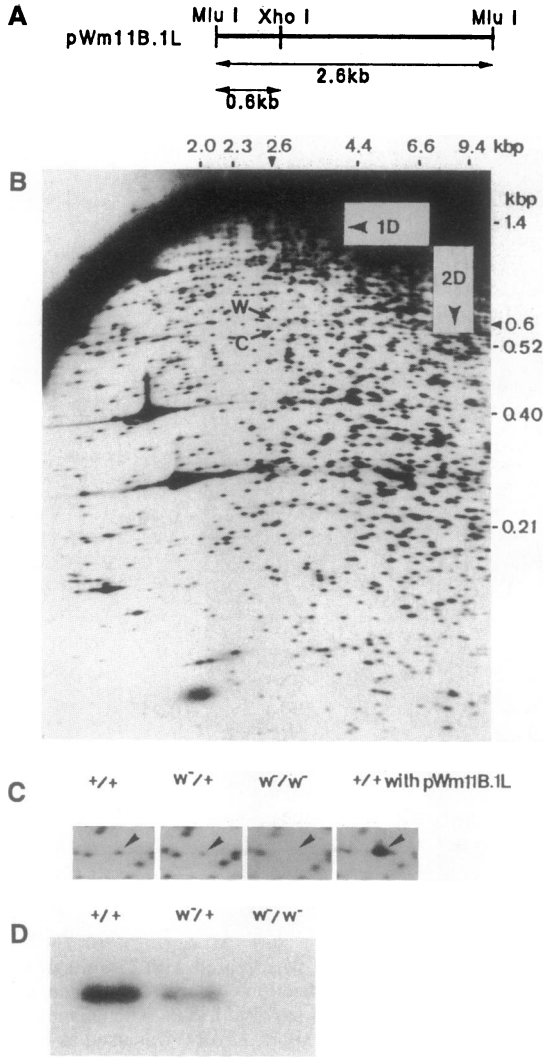

FIG. 3. One-to-one correspondence of a spot and a locus and the correlation of spot intensity with the number of copies. (*A*) Restriction map near the white (*w*) locus. (*B*) Autoradiographic profile of RLGS analysis of the genomic DNA of wild-type *D. melanogaster* (Canton S). *Mlu* I and *Xho* I were used as restriction enzymes A and C, respectively. W and C indicate the spot of the *w* gene and a control spot, respectively. The control spot was used to normalize the spot intensity. The photograph covers the 15- to 43-cm region of the original gel in the first dimension (1D) and the 9- to 45-cm region in the second dimension (2D). (*C*) Details around the spot corresponding to the *w* gene. Arrowheads indicate spots corresponding to the *w* gene [+, wild type; w^-, null allele of *w* locus; +/+ with pWm11B.1L, five copies per haploid genome of cloned *w* gene (pWm11B.1L) was mixed with wild-type genomic DNA and subjected to RLGS]. The intensities of spots on the RLGS profile and those of the bands obtained from Southern blot analysis were measured by PDQUEST. The actual size of each photograph is 2.5 × 1.6 cm. (*D*) Southern blot analysis of *Drosophila* DNA used here. The DNA was digested with *Mlu* I and *Xho* I and probed with pWm11B.1L.

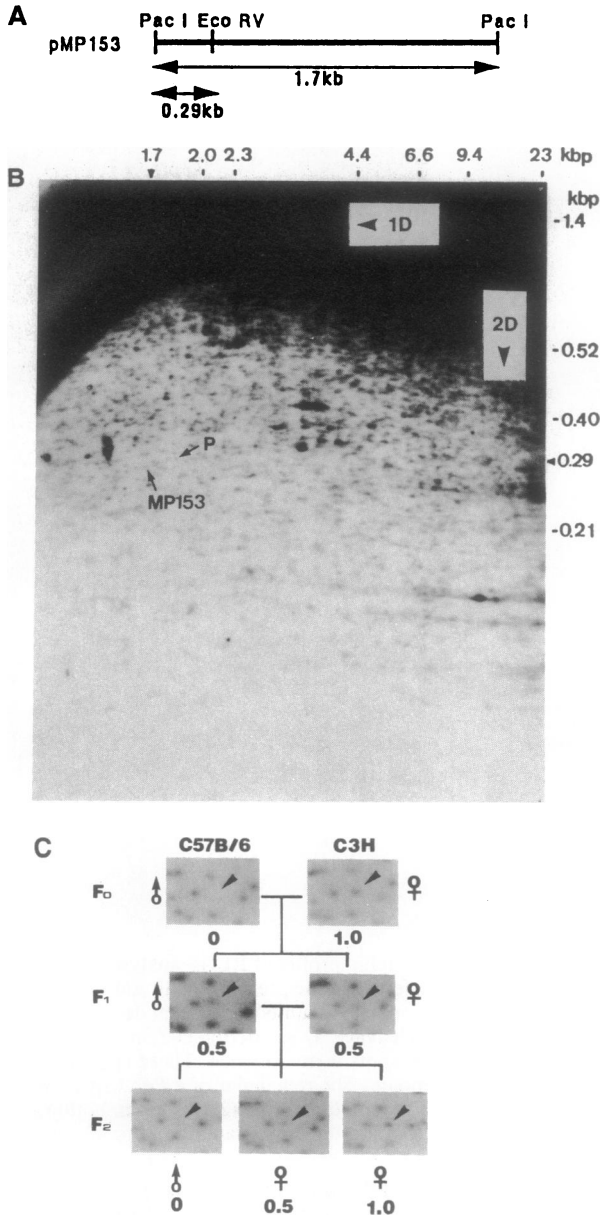

FIG. 4. Hereditary transmission of a restriction landmark. (*A*) Restriction map of DNA clone pMP153. (*B*) Autoradiographic profile of RLGS analysis of the genomic DNA of a female inbred strain mouse (C3J/HeJ). *Pac* I and *Eco*RV were used as restriction enzymes A and C, respectively. To identify the spot intensity of the single copy number per haploid genome, we isolated a DNA clone, pMP153, from the *Pac* I library and this clone was estimated to be a single copy by Southern blot analysis (data not shown). This clone corresponds to the spot MP153, and the spot derived from several copies of the cloned DNA was identical to the MP153 spot. Spot P is a polymorphic spot specific for C3H (*C*). Its intensity corresponding to the single copy was identical to that for MP153. The photograph covers the 11- to 43-cm region of the original gel in the first dimension (1D) and the 10- to 49-cm region in the second dimension (2D). (*C*) Spot profiles of F_0 through F_2 in the pedigree made by a cross between C57B/6 (B6) and C3H/HeJ (C3H). The transmission pattern and the intensity of spot P follow the Mendelian rule. Arrowheads indicate spot P. The intensities of the spots from RLGS were measured by PDQUEST. The copy number of spot P from F_0 through F_2 is represented below each photograph. The actual size of each photograph is 2.1 × 1.5 cm.

used as restriction enzyme A. (*iii*) Spot intensity reflects the copy number of the restriction landmark on the genome; thus haploid and diploid genomic DNAs can be discriminated by this method, as shown in Figs. 3 and 4. Since some enzymes possess a sequence preference in their cleavage pattern, analysis of the spot intensities might sometimes be compromised. However, it seems that most spots were not affected by cleavage, as seen in Fig. 4, because these intensities are relatively uniform and correspond to a single copy. (*iv*) This method can be applied to all organisms because no DNA probes are needed, unlike hybridization methods (1–3). Although some efforts have been made to provide many landmarks by hybridization using repetitive DNA (3), satisfactory resolution was not achieved because the repetitive DNA sequence probe did not hybridize completely to target DNAs

due to the difference in homology. In addition, analysis by the repetitive sequence is limited to the area where that sequence exists.

RLGS is applicable to genome mapping, the estimation of evolutional distance, and the study of many biological problems, such as animal mutants, cancer, and genetic diseases. In other words, the applications can be classified into two categories. The first is the detection of physical differences such as deletion and amplification. This includes deletion mapping of the recessive oncogene in cancer and the gene encoding mutant loci (8). The second is genetic analyses using polymorphic spots. RLGS may be useful for genome mapping or mapping of the specific mutant locus by use of interspecies crosses between *Mus spretus* and laboratory mice (9), in which at least 50% of the spots are polymorphic (T. Shiroishi, K. Moriwaki, S.H., I.H., and Y.H., unpublished results).

We thank Drs. H. Ryo and T. Todo for providing the fly samples and plasmid pWm11B.1L. We also thank Drs. K. Matsubara and Y. Takasaki for their advice and Toyobo for providing ddYTP[αS]. This work was supported by a Research Grant for Aging from the Ministry of Health and Welfare and a grant from the Ministry of Education and Culture.

1. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
2. Brilliant, M. H., Gondo, Y. & Eicher, E. M. (1991) *Science* **252,** 566–569.
3. Uitterlinden, A. G., Slagboom, P. E., Knook, D. L. & Vijg, J. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 2742–2746.
4. Blin, N. & Stafford, D. W. (1976) *Nucleic Acids Res.* **3,** 2303–2308.
5. Garrels, J. I. (1989) *J. Biol. Chem.* **264,** 5269–5282.
6. Vanyushin, B. F., Tkachera, S. G. & Belozersky, A. N. (1970) *Nature (London)* **225,** 948–949.
7. Bingham, P. M., Levis, R. & Rubin, G. M. (1981) *Cell* **25,** 693–704.
8. Russell, L. B., Hunsicker, P. R., Cacheiro, N. L. A., Bangham, J. W., Russell, W. L. & Shelby, M. D. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 3704–3708.
9. Avner, P., Amer, L., Dandole, L. & Guenet, L. (1988) *Trends Genet.* **4,** 18–23.