Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 355–359, January 1982
Biochemistry

Sequence organization and genomic distribution of the major family of interspersed repeats of mouse DNA

(interspersed repeated sequences/genome organization/recombinant DNA)

MICHÈLE MEUNIER-ROTIVAL, PHILIPPE SORIANO, GÉRARD CUNY, FRANÇOIS STRAUSS, AND GIORGIO BERNARDI

Laboratoire de Génétique Moléculaire, Institut de Recherche en Biologie Moléculaire, 2 Place Jussieu, 75005 Paris, France

Communicated by Gary Felsenfeld, September 30, 1981

ABSTRACT We have investigated the organization and the distribution of a family of interspersed DNA repeats in the mouse genome. The repeats are at least 5600 base pairs (bp) in size and contain two contiguous *Bam*HI endonuclease fragments, 4000 and 540 bp in size, the larger of which includes a 1350-bp *Eco*RI fragment studied by previous authors. The repeats are polymorphic in their restriction maps, and represent the major family of interspersed repeats in the mouse genome. The repeats are present almost exclusively in the two light major components of mouse DNA, and the base composition of their large *Bam*HI fragments matches that of those components. The genomic distribution of the repeats is different from that of structural genes, which are present not only in the two light components but also in the two heavy components of mouse DNA. This distribution indicates that the repeats are not involved, at least in any simple way, in the regulation of gene expression.

Degradation of mouse main-band DNA with restriction enzymes *Eco*RI, *Hin*dIII, *Hin*dII + III and *Hae* III yields discrete fragments that appear in gel electrophoresis as distinct bands over a continuous background smear (1, 2). In particular, *Eco*RI yields 1.3-kilobase (kb) fragments, representing 1.6% of main-band DNA; the degree of repetition of these fragments (30,000 copies per haploid genome) puts them in the intermediate repetitive kinetic class (1). Further studies on the 1.3-kb *Eco*RI fragments (3–6) have confirmed their interspersion in the main-band DNA and their lack of correlation with satellite DNA and have shown that these fragments are heterogeneous in some of their internal restriction sites and are included in larger repeats 3 kb in size.

In the present work, we have studied the sequence organization and the genome distribution of the interspersed repeats of mouse DNA. The main conclusion reached is that the major family of interspersed repeats of the mouse genome consists of sequences at least 5600 bp in size, polymorphic in their restriction maps, and present almost exclusively in the two major light components of mouse DNA. Such genomic distribution is different from that of structural genes, which are present on all four major components of mouse DNA (7–14); this indicates that the major family of interspersed repeats is not involved, at least in any simple way, in the regulation of gene expression.

MATERIALS AND METHODS

BALB/c mouse DNA was prepared from thymus or liver as described (11, 15). Its major components were isolated as reported (10, 11); the average molecular weights of the two preparations used in the experiments of Figs. 1–3 and 4 were 20 kb and 100 kb, respectively.

Restriction endonucleases were either purchased from commercial suppliers or preparéd by L. Mallet and A. Meier of our Institute. Vertical gels of agarose (type II, Sigma) or polyacrylamide (Bio-Rad) were used for electrophoresis.

Extraction of the large [4000-base pair (bp)] *Bam*HI fragment and of the (1350-bp) *Eco*RI fragment from agarose were performed by dissolving the gel in sodium perchlorate at 65°C and by chromatographing the mixture at 65°C on hydroxyapatite (16). Extraction of the small (540-bp) *Bam*HI fragment from polyacrylamide was done by diffusion. *Eco*RI and *Bam*HI fragments were inserted, using phage T4 ligase (Bethesda Research Laboratories) at the corresponding sites of plasmids pSF2124 and pBR322, respectively. These plasmids were used to transform competent *Escherichia coli* HVC 45 and HB101 cells, respectively. Recombinant clones were identified by the colicin test (17) in the first case or by plating the bacterial cells on L-agar supplemented with tetracycline at 10 μg/ml or ampicillin at 40 μg/ml in the second one.

DNA fragments were transferred from agarose gels to nitrocellulose filters (BA 85, Schleicher & Schuell) by blotting. Radioactive labeling of plasmid DNA was done by nick-translation. Prehybridization and hybridization were done as described (18). Low stringency washing conditions (150 mM NaCl/15 mM sodium citrate, 65°C) were used with filters loaded with genomic DNA; more stringent conditions (15 mM NaCl/1.5 mM sodium citrate, 65°C) were used with filters loaded with recombinant plasmid DNA. Autoradiography was done at −80°C on preflashed Kodax X-Omat or Fuji Rx films, using Du Pont Cronex Lightning Plus intensifying screens.

RESULTS

Repeated Fragments in Digests of Mouse DNA and of Its Major Components. When digested with *Eco*RI, *Bam*HI, *Kpn* I, *Hin*cII, and *Hae* III and subjected to gel electrophoresis, mouse DNA exhibits several bands over a smear (Figs. 1 *A* and *C*, 2 *A* and *A*′, 3*A*, and 4*A*). The lengths, in bp, of the fragments corresponding to the bands are the following: *Eco*RI, 1350; *Bam*HI, 4000, 1350, 970, 840, and 540; *Kpn* I, 2500 and 820; *Hin*cII, 2500; *Hae* III, 3200, 1960, 1860, 1180, 1120, 1000, 950, 850, 350, and 180.

When restriction digests of isolated major components of mouse DNA (11) were examined, it was observed (Figs. 1 *A* and *C* and 4*A*) that the bands mentioned above were exclusively (or almost so in the case of the 1350-bp *Eco*RI and 540-bp *Bam*HI bands) present in the two light components (1.699 and 1.701 g/cm^3); the buoyant density of the 4000-bp *Bam*HI fragment, as isolated from plasmid pMRB1.1, was found to be 1.698 g/cm^3. In contrast, the 1350-bp *Bam*HI band was present in all four major components.

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: kb, kilobase(s); bp, base pair(s).

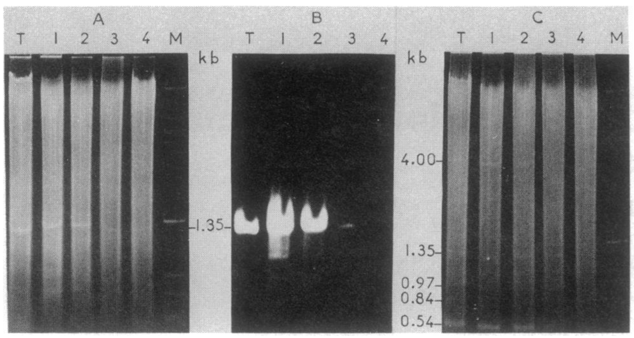

FIG. 1. Repeated *Eco*RI and *Bam*HI fragments in mouse DNA and its major components. Thirty micrograms of mouse DNA (T) and 25 μg each of its major components of buoyant density 1.699 (lanes 1), 1.701 (lanes 2), 1.704 (lanes 3), and 1.708 (lanes 4) g/cm^3 were digested by *Eco*RI (*A*) or *Bam*HI (*C*), electrophoresed on 1% agarose, and stained with ethidium bromide. (*B*) Autoradiogram obtained by hybridizing the pMRE1 probe with *Eco*RI fragments. Molecular weight markers (M) were λ phage DNA digested by *Eco*RI and simian virus 40 DNA digested by *Hae* III.

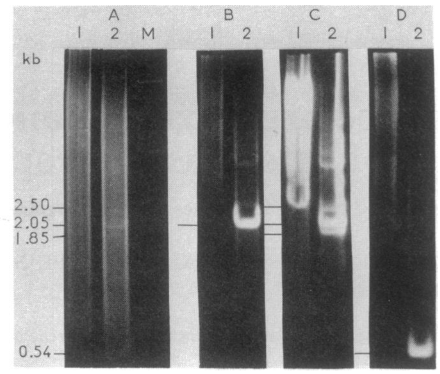

FIG. 3. Repeated *Hinc*II fragments in mouse DNA. Mouse DNA major component 1.699 g/cm^3 was digested by *Hinc*II (lanes 1) and *Hinc*II plus *Bam*HI (lanes 2), electrophoresed on 1% agarose, and stained with ethidium bromide (*A*). (*B, C,* and *D*) Autoradiograms obtained after hybridizing labeled pMRE1, pMRB1.1, and pMRB5, respectively, with the transferred digests. Molecular weight markers (M) are as in Fig. 1.

Restriction Mapping of Cloned Repeated Fragments. Fig. 5 shows the restriction maps of the repeated fragments carried by recombinant plasmids pMRE1, pMRB1.1–3, and pMRB5, namely one 1350-bp *Eco*RI fragment, three 4000-bp *Bam*HI fragments, and one 540-bp *Bam*HI fragment. The maps were established on the basis of single, double, and triple digests and of cross-hybridization experiments which showed: (*i*) that the 4000-bp *Bam*HI fragment of pMRB1.1 is homologous in sequence with the 4000-bp fragments of pMRB1.2 and pMRB1.3, (*ii*) that the 1350-bp *Eco*RI fragment of pMRE1 is homologous in sequence with the 1350-bp *Eco*RI fragment contained in the inserts of pMRB1.1, pMRB1.2, and pMRB1.3, (*iii*) that the 540-bp *Bam*HI fragment does not exhibit any sequence homology with the other cloned fragments. It should be noted that the three cloned 4000-bp *Bam*HI fragments exhibit very similar, yet not identical, maps and that the cloned 1350-bp *Eco*RI fragment has the same size and a map similar to, yet not identical with, the map of the *Eco*RI fragments contained in the 4000-bp *Bam*HI fragments.

Restriction Mapping of Genomic Repeated Fragments. Cloned repeated *Eco*RI and *Bam*HI fragments were found to hybridize with genomic fragments of identical size (Table 1 and Fig. 6), under conditions in which unique sequences would show no hybridization. In addition to such a band hybridization, a weak smear hybridization was also detected, indicating the presence of fragments showing sequence homology with the probes but having *Eco*RI and *Bam*HI sites located differently.

Hybridization of recombinant plasmids pMRE1, pMRB1.1, and pMRB5 on other single and double restriction digests (Table 1) allowed the construction of a restriction map of genomic repeated fragments (Fig. 5). This map rests on the hybridization of the probes with genomic fragments obtained with *Eco*RI, *Bam*HI, *Kpn* I, *Hinc*II, and combinations of these enzymes. Some fragments (2900-bp *Eco*RI, 2500-bp *Kpn* I, 2500-bp *Hinc*II) were absent from the probes but could easily be placed on the restriction map of genomic fragments. In contrast, another genomic fragment also absent from the probes (820-bp *Kpn* I) could not be put on the map of Fig. 5; such a fragment was probably due to the polymorphism of the restriction map of the genomic repeated fragments (see *Discussion*); in all likelihood, for the same reason, the 1720-bp *Hinc*II fragment of pMRB1.1 (Fig. 5) was not found in the genomic fragments. It should also be noted that at least one *Hae* III fragment (1960

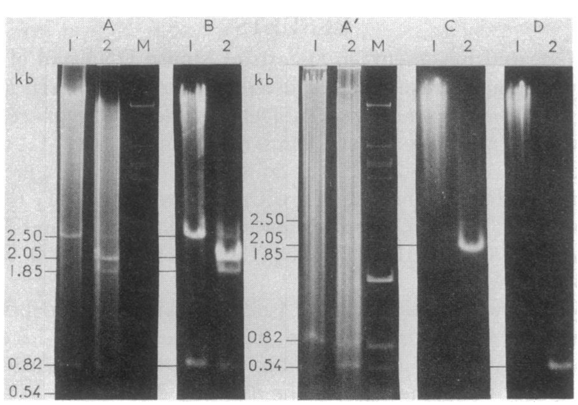

FIG. 2. Repeated *Kpn* I fragments in mouse DNA. Mouse DNA major component 1.699 g/cm^3 was digested by *Kpn* I (lanes 1) and *Kpn* I plus *Bam* HI (lanes 2), electrophoresed on 1% agarose, and stained with ethidium bromide (*A* and *A′*). (*B, C,* and *D*) Autoradiograms obtained after hybridizing labeled pMRB1.1, pMRE1, and pMRB5, respectively, with the transferred digests. Molecular weight markers (M) are as in Fig. 1.

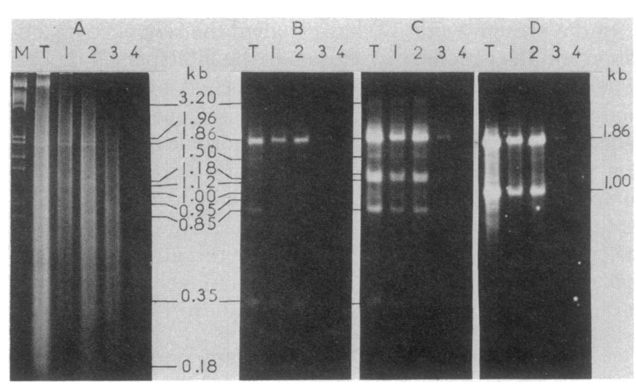

FIG. 4. Repeated *Hae* III fragments in mouse DNA and its major components. Mouse total DNA (T) and its major components of buoyant density 1.699 (lanes 1), 1.701 (lanes 2), 1.704 (lanes 3), and 1.708 (lanes 4) g/cm^3 (in quantities proportional to their amounts in total DNA; see ref. 13) were digested by *Hae* III, electrophoresed on 2.5% agarose, and stained with ethidium bromide (*A*). (*B, C,* and *D*) Autoradiograms obtained after hybridizing labeled pMRE1, pMRB1.1, and pMRB5, respectively, with the transferred digests. Molecular weight markers (M) were λ phage DNA digested by *Eco*RI plus *Hind*III.

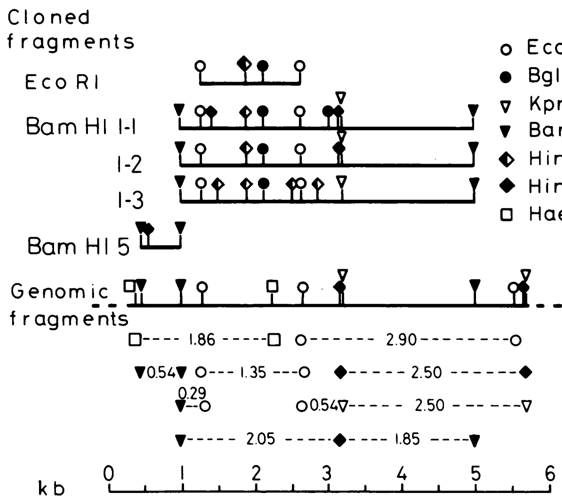

FIG. 5. Restriction maps of cloned and genomic repeated restriction fragments of mouse DNA. In addition to the sites indicated on the maps, the following information is available on the cloned fragments: pMRE1 contains one *Hae* III site (whose cleavage produces two fragments, 1000 and 350 bp) but no *Bam*HI, *Hin*cII, or *Kpn* I sites; pMRB1.1 produces six fragments upon digestion with *Hae* III: 1230, 1120, 880, 350, 315, and 115 bp (the 1230- and 315-bp fragments are *Bam*HI/*Hae* III fragments); pMRB5 contains two *Alu* I sites, one *Hpa* II site, and no *Bgl* II, *Eco*RI, *Hae* III, *Hin*dIII, *Kpn* I, or *Sau*3A sites. Concerning the genomic map, *Bgl* II and *Hin*dIII sites were not mapped, and only two *Hae* III sites were mapped in order to show that the 4000-bp and 540-bp *Bam*HI fragments are linked.

bp) visible after ethidium bromide staining did not hybridize with any of the probes; on the other hand, one *Hae* III fragment (1500 bp) not visible after staining hybridized with pMRB1.1.

The common hybridization of the three probes on the 1860-bp *Hae* III genomic fragment (Fig. 7, lanes 2) revealed that the 540-bp *Bam*HI fragment is contiguous, or nearly contiguous, to the 4000-bp *Bam*HI fragment. Moreover, because a *Kpn* I site can also be mapped at about 1000 bp from the end of the 4000-bp *Bam*HI fragment (see above and Fig. 5), the *Bam*HI genomic fragments belong in segments at least 5600 bp long.

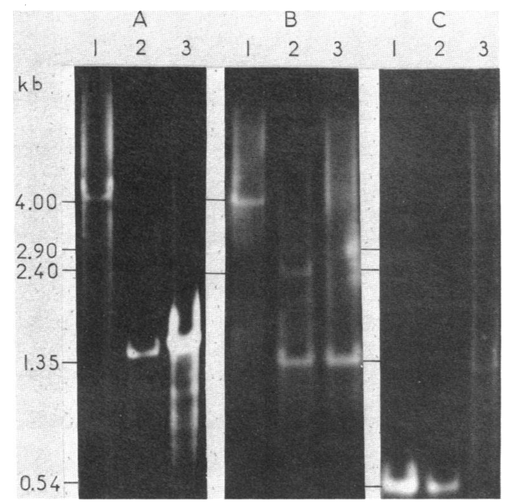

FIG. 6. Restriction mapping of genomic interspersed repeated fragments of mouse DNA. Mouse total DNA was digested with *Bam*HI (lanes 1), *Bam*HI plus *Eco*RI (lanes 2), or *Eco*RI (lanes 3), and electrophoresed on 1% agarose. (*A*, *B*, and *C*) Autoradiograms obtained after hybridizing labeled pMRE1, pMRB1.1, and pMRB5, respectively, with the transferred digests.

DISCUSSION

Distribution of Repeated Fragments in the Major Components of Mouse DNA. The results of Figs. 1 and 4 indicate that, except for the 1350-bp *Bam*HI fragment that appears to be present in all four components (Fig. 1*C*), all the fragments are present, at about the same extent, only in the two light components. Bands corresponding to the 1350-bp *Eco*RI fragment and to some *Hae* III fragments appear to be also present in the 1.704 g/cm^3 component, but hybridization with cloned probes was barely detectable on those bands (Figs. 1*B* and 4 *B–D*). Because there is independent evidence for a small extent of contamination of the 1.704 g/cm^3 component by the lighter ones (11), the conclusion should be drawn that the family of repeated sequences studied here is present almost exclusively in the light components of the mouse genome. Such presence is not due

Table 1. Hybridization of cloned repeated fragments to restriction digests of mouse DNA

		Digest									
Probe		*Bam*HI	*Eco*RI	*Bam*HI + *Eco*RI	*Kpn* I	*Kpn* I + *Bam*HI	*Kpn* I + *Eco*RI	*Hin*cII	*Hin*cII + *Bam*HI	*Hin*cII +*Eco*RI	*Hae* III
pMRE1		4000	1350	1350	Smear	2050	1350	Smear	2050	1350	1860
		Smear	Smear			Smear	Smear		Smear	Smear	1500
											350
	Figs.	6*A*1	1*B*, 6*A*3	6*A*2	2*C*1	2*C*2	NS	3*B*1	3*B*2	NS	4*B*
pMRB1.1		4000	1350	2400	2500	2050	2500	2500	2050	2500	3200
		Smear	2900	1350	820	1850	1350	Smear	1850	1350	1500
			Smear	290 (NS)	Smear	820	820		Smear	540	1180
							540			Smear	1120
											1000
											950
											850
											350
	Figs.	6*B*1	6*B*3	6*B*2	2*B*1	2*B*2	NS	3*C*1	3*C*2	NS	4*C*
pMRB5		540	Smear	540	Smear	540	Smear	Smear	540	Smear	1860
											1000
	Figs.	6*C*1	6*C*3	6*C*2	2*D*1	2*D*2	NS	3*D*1	3*D*2	NS	4*D*

NS, not shown.

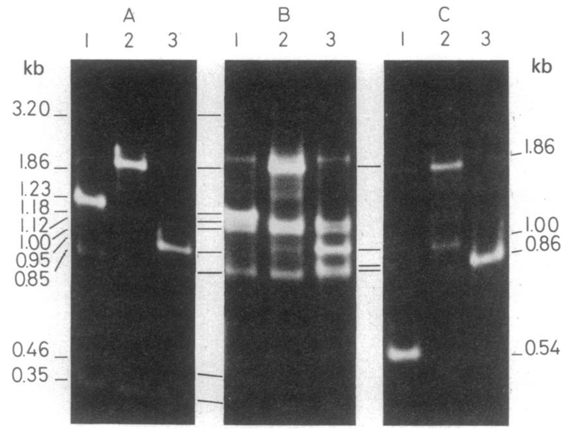

FIG. 7. **Restriction mapping of genomic interspersed repeated fragments of mouse DNA with additional enzymes. Mouse total DNA was digested with *Hae* III plus *Bam*HI (lanes 1), *Hae* III (lanes 2), and *Hae* III plus *Eco*RI (lanes 3), and electrophoresed on 2.5% agarose. (*A*, *B*, and *C*) Autoradiograms obtained after hybridizing labeled pMRE1, pMRB1.1, and pMRB5, respectively, with the transferred digests.**

to the fact that the 4000-bp *Bam*HI fragments have a buoyant density, 1.698 g/cm^3, close to those of the light components, because they are short relative to the average size of the DNA preparations used. Instead, it indicates that the flanking sequences of several thousand interspersed repeats scattered in the genome have about the same base composition, in keeping with the large compositional homogeneity of the major components of warm-blooded vertebrates (11). Finally, the different genomic distribution of the 1350-bp *Bam*HI fragment compared to the other fragments points to the existence of different families of repeated fragments in the mouse genome. Such a conclusion cannot be reached with certainty merely on the basis of a lack of cross-hybridization, as shown by the fact that the 540-bp *Bam*HI fragment does not cross-hybridize with the 4000-bp *Bam*HI fragment and yet is contiguous to it.

Restriction Maps of Cloned Repeated Fragments. The data of Fig. 5 allow two main conclusions to be reached. The first one is that the repeated fragments are polymorphic in their restriction maps, a finding in agreement with previous results (4–6). In spite of this restriction site polymorphism, both the conservation of a number of restriction sites and the hybridization results obtained among cloned fragments and with genomic fragments (see below) strongly indicate a basic homology in the family of repeated sequences as studied here. The second conclusion is that the cloning and the mapping of the 4000-bp *Bam*HI fragments has allowed us to observe that the 1350-bp *Eco*RI fragment is a part of a larger repeat.

Restriction Maps of Genomic Repeated Fragments. The use of labeled cloned repeated fragments as probes allowed us to link together most fragments seen by gel electrophoresis of restriction digests of mouse DNA into repeats over 5600 bp long. In fact, the 540-bp *Bam*HI fragment is contiguous to the 4000-bp *Bam*HI fragment, and the latter appears to extend to a *Kpn* I site located at about 1000 bp from its other end (Fig. 5). The only exceptions found were (*i*) the 1350-bp *Bam*HI fragments; (*ii*) the 840- and 970-bp *Bam*HI fragments (Fig. 1C) and the 1960-bp *Hae* III fragments (Fig. 4). The former have a different genomic distribution and belong, therefore, in a different family of interspersed repeated sequences, and the latter have the same distribution as the 4000- and 540-bp *Bam*HI fragments but do not hybridize with any of the probes and cannot be placed on the genomic map (Fig. 5). They might either represent segments adjacent to the main, 4000- and 540-bp *Bam*HI fragments or correspond to one (or more) additional family (families) of repeated sequences. It appears, however, that some of the repeated sequences discussed here may not be included in larger repeats, because 1350-bp *Eco*RI fragments not included in 4000-bp *Bam*HI fragments have been found close to the β-globin gene cluster (19); interestingly enough, the β-globin gene is also localized in one of the two light components (13) that contain the *Eco*RI fragments.

Estimating the stoichiometry of the repeated restriction fragments contained within the 5.6-kb units is not easy; at the present time it can only be said that the data obtained in the present work are compatible with the idea that the map of Fig. 5 depicts the situation of the majority of the fragments. Equally difficult is to estimate the amount of the fragments belonging in the 5.6-kb units relative to that of the fragments unrelated to them. Because of the fact that the background smear varies in intensity from the top to the bottom of the gel and in different major components of mouse DNA at corresponding molecular weight levels, the intensity of the bands stained by ethidium bromide provides indications that may be misleading as to the relative amounts of DNA fragments contained in the bands, as clearly demonstrated by hybridization experiments (Fig. 1 *A* and *B*). In spite of these problems, the gel electrophoresis results suggest that the family of repeated sequences studied here is the major one in the mouse genome. Two independent arguments are in agreement with this conclusion. The first argument is that the 1350-bp *Eco*RI fragments have been estimated to represent 0.5–3% of mouse DNA (1, 4, 5); if the majority of them belong in larger repeats 5.6 kb in size, then these may correspond to up to 2–12% of mouse DNA. The second argument is that the genomic distribution of the repeated sequences studied here parallels that of interspersed repeated sequences, as estimated by the reassociation kinetics of the major components, in that the latter are very abundant in the light components and very scarce in the heavy ones (12). If such a scarcity is accounted for, at least to a large extent, by the absence of the *Bam*HI repeated sequences in the heavy components, then these should represent a sizable fraction of the mouse genome.

As far as the functional role of the repeated sequences studied here is concerned, an important point is that their genomic distribution does not parallel that of genes, because the latter are present, in roughly equal relative amounts, in all four of the components of the mouse genome (11) [in particular, the α-globin genes are on the 1.708 g/cm^3 component, whereas the β-globin genes are on the 1.701 g/cm^3 component (13)]. This rules out at least any simple correlation between the major family of interspersed repeated sequences and the regulation of gene expression in the mouse genome. In agreement with this conclusion is the finding (12) of very large differences in the distribution of interspersed repeated sequences in the major components of the mouse and human genomes, which contrasts with the conserved distribution of structural genes, because these differences are unlikely to concern sequences involved in the regulation of gene expression. On the other hand, the very existence of distinct families of interspersed repeated sequences, demonstrated here by their different genomic distribution, suggests by itself different functions for such families.

1. Horz, W., Hess, I. & Zachau, H. (1974) *Eur. J. Biochem.* **45**, 501–512.
2. Soriano, P. (1978) Dissertation (Paris 7 Univ., Paris).
3. Manuelidis, L. (1980) *Nucleic Acids Res.* **8**, 3247–3258.
4. Heller, R. & Arnheim, N. (1980) *Nucleic Acids Res.* **8**, 5031–5042.
5. Cheng, S. M. & Schildkraut, C. L. (1980) *Nucleic Acids Res.* **8**, 4075–4090.

6. Brown, S. D. M. & Dover, G. (1981) *J. Mol. Biol.* **150**, 441–466.
7. Filipski, J., Thiery, J.-P. & Bernardi, G. (1973) *J. Mol. Biol.* **80**, 177–197.
8. Thiery, J.-P., Macaya, G. & Bernardi, G. (1977) *J. Mol. Biol.* **108**, 219–235.
9. Macaya, G., Thiery, J.-P. & Bernardi, G. (1977) *J. Mol. Biol.* **108**, 237–254.
10. Cortadas, J., Olofsson, B., Meunier-Rotival, M., Macaya, G. & Bernardi, G. (1979) *Eur. J. Biochem.* **99**, 179–186.
11. Cuny, G., Soriano, P., Macaya, G. & Bernardi, G. (1981) *Eur. J. Biochem.* **115**, 227–233.
12. Soriano, P., Macaya, G. & Bernardi, G. (1981) *Eur. J. Biochem.* **115**, 234–239.
13. Bernardi, G. (1979) in *Recombinant DNA and Genetic Experimentation*, eds. Morgan, J. & Whelan, W. J. (Pergamon, New York), pp. 15–20.
14. Hudson, A., Cuny, G., Cortadas, J., Haschemeyer, A. E. V. & Bernardi, G. (1980) *Eur. J. Biochem.* **112**, 203–210.
15. Blin, N. & Stafford, D. W. (1976) *Nucleic Acids Res.* **3**, 2303–2308.
16. Southern, E. M. (1975) *J. Mol. Biol.* **94**, 51–69.
17. Ozeki, H. (1968) *Methods Virol.* **4**, 565–592.
18. Meunier-Rotival, M., Cortadas, J., Macaya, G. & Bernardi, G. (1979) *Nucleic Acids Res.* **6**, 2109–2123.
19. Haigwood, N. L., Jahn, C. L., Hutchison, C. A., III, & Edgell, M. H. (1981) *Nucleic Acids Res.* **9**, 1133–1150.