Distribution of the mammalian-wide interspersed repeats (MIRs) in the isochores of the human genome

Giorgio Matassi[1,a], Damian Labuda[b], Giorgio Bernardi[a,*]

[a] *Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2 Place Jussieu, 75005 Paris, France*
[b] *Centre de Cancérologie Charles Bruneau, Centre de Recherche de l'Hôpital Sainte-Justine, Département de Pédiatrie, Université de Montréal, 3170 Côte Sainte-Catherine, Montréal, Que. H3T 1C5, Canada*

Received 25 June 1998; received in revised form 6 October 1998

Abstract The distribution of MIRs (mammalian-wide interspersed repeats) was investigated in 164 human sequences (≥100 kb), which were assigned, according to their GC level, to isochore families L, H1, H2 and H3. MIR elements, whose total number in the genome was estimated to be about 3.3×10^5, were found to be unevenly distributed in human isochores. The majority of MIRs (55%) were found in the L isochore family. In contrast, MIR density was highest in H2, closely followed by H1, whereas densities in L and H3 were 2- and 3-fold lower than in H2, respectively. For this reason, the assessment of MIR distribution by inter-repeat PCR led to an overestimation of MIR numbers in H2 isochore and an underestimation in L isochores.

Key words: Isochore; Mammalian-wide interspersed repeat; Inter-mammalian-wide interspersed repeat polymerase chain reaction

1. Introduction

Among the repeated sequences present in the human genome, MIRs (mammalian-wide interspersed repeats) represent about 0.4–1.0% of the genome [1,2]. MIR elements are tRNA-derived SINEs [2], and are readily identifiable by a more conserved central core region of about 70 nucleotides [1,3,4].

Found in placental mammals, marsupials and monotremes [1], MIRs are thought to have amplified about 130 million years ago [1], and might, therefore, be considered the most ancient SINE family detected so far in mammals. The amplification of these elements seems to have ceased in the ancestors of placental mammals [1]. Two interesting evolutionary issues concern the spreading of MIRs in the mammalian genome and whether this event can be related to the isochore structure of the latter. Indeed, the human genome, which is a good representative of the majority of mammalian genomes [5], is a mosaic of DNA segments, >300 kb in size, called isochores, belonging to four families (L, H1, H2 and H3), each one of which is characterised by different gene and GC levels [6,7]. GC-poor (L) isochores represent about 63% of the genome, whereas GC-rich (H1, H2 and H3) isochores make up about 24%, 7.5% and 4–5%, respectively, the remaining DNA corresponding to satellite and ribosomal sequences [7]. On the other hand, gene density increases almost 20-fold from L to H3 isochores [7].

Here we have analysed the distribution of MIR elements in human isochores both by computer search on large sequences from GenBank and by inter-MIR PCR on human DNA fractionated according to base composition. We have shown that the distribution of MIR elements in human isochores is not uniform and we discuss the implications of this finding.

2. Materials and methods

2.1. Searching GenBank for MIRs

One hundred and sixty-four sequences ≥100 kb in size, corresponding to about 23.7 Mb of human DNA (Table 1), were extracted from GenBank using the ACNUC retrieval system [8] and were subsequently partitioned into four human isochore classes according to their GC level. Between-isochore boundaries were taken as 41.5% (L/H1), 46.3% (H1/H2) and 50.8% GC (H2/H3) [7].

MIR repeats were searched for in the data set, using the CENSOR program [9]. In order to reach a reliable estimate of the number of MIRs, two CENSOR searching criteria were used. More conserved MIR elements were detected using a stringent searching mode which keeps all hits scoring above 35.0 in the Smith-Waterman local alignment. A larger number of MIRs, comprising less conserved copies, were recovered using a lower stringency searching mode, which also keeps all hits scoring between 22.0 and 35.0, if their ratio of mismatches to transitions is smaller than 2.8.

2.2. Compositional fractionation of human DNA

DNA was prepared from peripheral leukocytes by the SDS/proteinase K extraction procedure [10] and fractionated according to base composition by preparative density gradient centrifugation in Cs_2SO_4/BAMD, as described [11]. BAMD is 3,6-bis-acetate mercuri-methyldioxane. A ligand/nucleotide molar ratio of 0.14 was used. The GC level of each fraction was determined by standard HPLC analysis [12].

2.3. Inter-MIR PCR

Primers were based on a consensus sequence derived from an earlier compilation of the central core sequence of 455 MIR copies [1]. Four PCR primers were used: *mir* (3′- TGGAACTCGTTCAGTGA-5′), *omir* (5′-ACCTTGAGCAAGTCACT-3′), *mil* (5′-GCCTCAGTTT-CCTCATC-3′), *omil* (3′-CGGAGTCAAAGGAGTAG-5′) [1]. PCR amplifications on unfractionated DNA (25 ng) and DNA fractions (7 ng) were carried out in a total volume of 50 ml containing 200 mM of dNTPs, 25 pmol of primer, 2.5 U of Taq DNA polymerase (Gibco BRL) in 20 mM Tris-HCl pH 8.4, 50 mM KCl, 1.5 mM $MgCl_2$. The reactions were performed in a Perkin Elmer Cetus 2400 thermal cycler under the following conditions: 5 min at 95°C (1 cycle); 30 s at 95°C, 30 s at 50°C (*omir*) or 56°C (*mir, mil*) or 58°C (*omil*) and 1 min at 72°C (30 cycles). Reaction products were fractionated by electrophoresis in 2.0% agarose gels, stained with ethidium bromide and visualised under UV light.

3. Results and discussion

3.1. Distribution of MIR elements in human isochores by computer search

The majority of MIRs were found in L isochores, their amount decreasing in H1, in H2 and more so in H3 (Table

*Corresponding author. Fax: (33) (1) 44277977.
E-mail: bernardi@lovelace.citi2.fr

[1]Present address: INSERM U76, 6, rue Alexandre Cabanel, 75015 Paris, France.

PII: S0014-5793(98)01338-6

Table 1
Distribution of MIR elements in human isochore families

	Sequence (Mb)	MIRs in the data set[a]	MIRs in the genome	MIR density per Mb
L	14.9	1101–1408	180 059	85
H1	4.5	474–624	100 796	122
H2	2.9	351–513	37 986	149
H3	1.4	59–85	8 218	52
Total	23.7	1985–2630	327 059	102

[a]Minimum and maximum numbers correspond to two stringency searching criteria (see Section 2).

1). On the other hand, the MIR density was highest in H2, diminished in H1 and L to be lowest in H3.

Assuming a haploid human genome size of 3400 Mb [13], and taking into account the relative amounts of isochore families in the human genome (see Section 1), we assessed the approximate number of MIRs in each isochore family, as well as in the whole genome (Table 1). We estimate 3.3×10^5 ($\pm0.5\times10^5$) MIRs to be present in the human genome, which is very close to the previous figure of 3.0×10^5 copies [2], assessed from only 0.67 Mb of DNA. It should be noted that virtually the same numbers of MIRs are found on each of the two DNA strands.

3.2. Inter-MIR PCR assessment of the distribution of MIRs in human isochores

Human DNA was fractionated according to base composition and the distribution of MIR elements in DNA fractions was assessed by PCR amplification, in which four MIR-specific primers were used one at a time so that PCR products could be obtained from loci flanked by two copies of the same primer arranged in opposite orientations, within amplifiable distance [1] (see Section 2). Fig. 2 shows the distribution in DNA fractions of the amplified DNA segments obtained with the '*mir*' primer. In this experiment, 17 amplified regions were detected and assigned to a specific fraction based on the maximum intensity of the fluorescent signal.

Overall, 139 inter-MIR PCR products were detected with the four MIR-specific primers. Table 2 summarises their distribution in human isochore families: 17 products were found in L, 35 in H1, 78 in H2 and 9 in H3. The number of PCR products detected in the present analysis by each of the four primers was constantly about half of that obtained in a previous inter-MIR PCR analysis, carried out on unfractionated DNA, since the authors' estimates were based on the quantitation of a radioactive signal [1]. In contrast, here we have optimised PCR conditions in order to detect a distinguishable number of PCR products so that they could be easily assigned to DNA fractions.

The results obtained by the inter-MIR PCR approach indicated that the majority of MIR elements reside in the H2 isochore class, with remarkably lower values in H1 and in L and even lower values in H3 (Fig. 1C). Therefore, the simple assessment of MIR densities in different isochores (and therefore in unfractionated DNA as well) by using inter-MIR PCR may be misleading since the relation between the genomic density of repeats and the number of inter-repeat amplification products is logarithmic. Indeed, given that the actual distribution of MIRs in human isochores is that obtained by computer search (Fig. 1A), the number of MIR elements present in H2 isochores appears to be overestimated by this experimental approach (see Fig. 1C) because of their highest density in this isochore family (Fig. 1B).

Table 2
Distribution of inter-MIR PCR products in human isochore families

Primer	L	H1	H2	H3	Total
mir	3	4	7	3	17
omir	6	10	17	4	37
mil	5	8	26	1	40
omil	3	13	28	1	45
Total	17	35	78	9	139

3.3. Conclusions

MIRs, as currently seen in the human genome, represent the result of their retroposition and subsequent evolutionary process. These are most likely the repeats that were stably integrated.

The present investigations have shown that the MIR density increases monotonously by a factor of almost 2 from L to

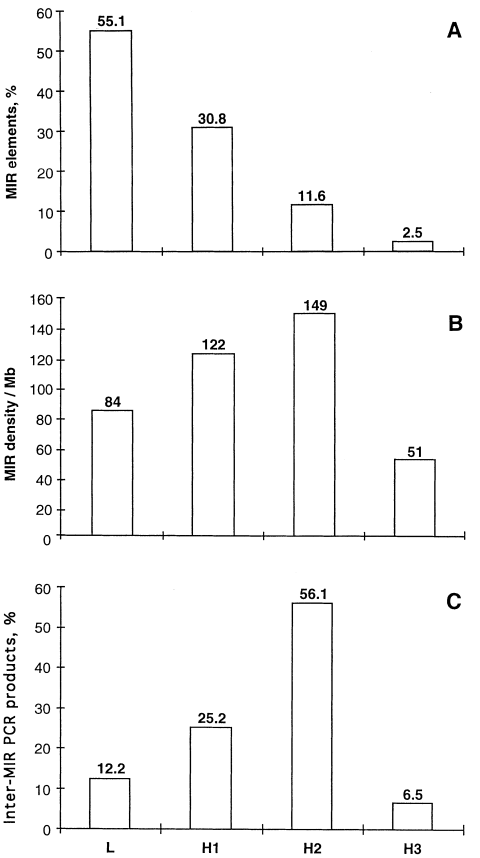

Fig. 1. Distribution of MIRs in human isochore families. A: Relative amounts of MIR elements from 164 large (≥100 kb) human sequences (see Table 1). B: MIR density (i.e. the average number of MIRs divided by the amount of DNA in the corresponding data set, see Table 1). C: Inter-MIR PCR analysis (see Table 2).

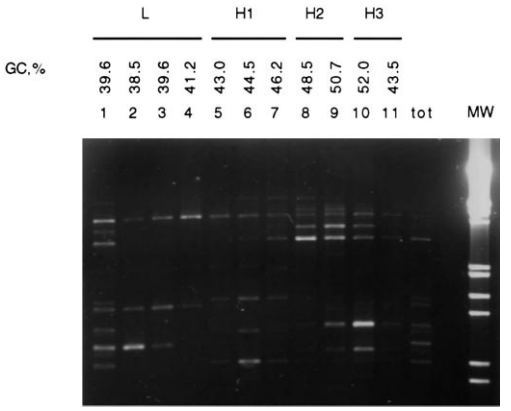

Fig. 2. Distribution of inter-MIR PCR products in human DNA fractions. Seventeen PCR products, obtained with the *mir* primer (see Section 2), were visible on the stained agarose gel under UV light. GC levels of fractions and their pooling into isochore classes are shown (see text). Fraction 1 corresponds to the pelleted material in the fractionation experiment and is slightly contaminated by DNA from other fractions [10]; this explains its slightly higher GC content as compared to fraction 2. The lower GC content of fraction 11 is due to the presence of an AT-rich satellite DNA detected by CsCl analytical ultracentrifugation analysis (not shown). tot, total unfractionated DNA. MW: *Eco*RI/*Hind*III digest of phage λ DNA used as size marker.

H2 isochores, to decrease then by a factor of 3 from H2 to H3. These differences may account for the fact that MIR density is experimentally underestimated by factors of 2 and 1.7 in H3 and L isochores, respectively, whereas it is overestimated by a factor of 1.5 in H2 isochore and is reasonably well assessed in H1 isochores. The low density of MIRs in L and H3 isochores may explain why a number of inter-MIR PCR products are missing. This may, however, be only a partial explanation, since the possibility that the primers used are better targeted towards MIRs from H1 and more so from H2 isochores, rather than towards MIRs located in L and H3, should be taken into consideration. Needless to say, the experimental results obtained with MIR-specific primers in this work should sound a note of caution concerning the use of this approach to analyse other repeated sequences, especially Alus.

The main result presented here, namely the different density of MIR sequences in different isochore families, is a clear indication that the mobility of these repeated sequences in the human genome is restricted, as shown in the most striking way by the low MIR density in H3 isochores. At least two factors may play a role in the poor integration of MIRs in H3 isochores. The first is the fact that the size of intergenic and intronic sequences in these isochores is the smallest in the human genome [7,14]. The second is that intergenic sequences in H3 are represented to a large extent by CpG islands and untranslated 3′ sequences, two sequence elements endowed with regulatory roles. Therefore, integration of repeats in most of these regions tends to be avoided. In contrast, MIRs may be better accepted in L isochores, where genes are very sparse. An intermediate situation is found in H1 and H2 isochores.

Finally, it is worth mentioning that the isochore distribution of MIRs is similar to that of Alus, the most abundant family of human repetitive sequences. In fact, while the majority of Alus are found in L isochores, their density is highest in the H2 family, decreases in H3 and H1, and is lowest in L [14]. These two families of SINEs are presumed to have been amplified in a similar manner by using the retroposition machinery of LINEs. However, Alus are believed to use the reverse transcriptase from LINE1 elements [15], whereas MIRs might use that encoded by LINE2 elements (Gilbert and Labuda, submitted). While the isochore distribution of LINE1 is clearly different from that of Alus, their highest density and number being found in the L family [16–18], that of LINE2 seems much more uniform throughout the genome, with a very slight preference for integration into H2 isochores (our unpublished data; [19]).

References

[1] Jurka, J., Zietkiewicz, E. and Labuda, D. (1995) Nucleic Acids Res. 23, 170–175.
[2] Smit, A.F. and Riggs, A.D. (1995) Nucleic Acids Res. 23, 98–102.
[3] Armour, J.A., Wong, Z., Wilson, V., Royle, N.J. and Jeffreys, A.J. (1989) Nucleic Acids Res. 17, 4925–4935.
[4] Donehower, L.A., Slagle, B.L., Wilde, M., Darlington, G. and Butel, J.S. (1989) Nucleic Acids Res. 17, 699–710.
[5] Sabeur, G., Macaya, G., Kadi, F. and Bernardi, G. (1993) J. Mol. Evol. 37, 93–108.
[6] Bernardi, G. (1995) Annu. Rev. Genet. 29, 445–476.
[7] Zoubak, S., Clay, O. and Bernardi, G. (1996) Gene 174, 95–102.
[8] Gouy, M., Gautier, C., Attimonelli, M., Lanave, C. and di Paola, G. (1985) Comput. Appl. Biosci. 1, 167–172.
[9] Jurka, J., Klonowski, P., Dagman, V. and Pelton, P. (1996) Comput. Chem. 20, 119–121.
[10] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[11] Cortadas, J., Macaya, G. and Bernardi, G. (1977) Eur. J. Biochem. 76, 13–19.
[12] Gehrke, C.W., McCune, R.A., Gama-Sosa, M.A., Ehrlich, M. and Kuo, K.C. (1984) J. Chromatogr. 301, 199–219.
[13] Cavalier-Smith, T. (1985) in: The Evolution of Genome Size (Cavalier-Smith, T., Ed.), pp. 69–103, Wiley, London.
[14] Jabbari, K. and Bernardi, G. (1998) Gene (in press).
[15] Ohshima, K., Hamada, M., Terai, Y. and Okada, N. (1996) Mol. Cell. Biol. 16, 3756–3764.
[16] Meunier-Rotival, M., Soriano, P., Cuny, G., Strauss, F. and Bernardi, G. (1982) Proc. Natl. Acad. Sci. USA 79, 355–359.
[17] Soriano, P., Macaya, G. and Bernardi, G. (1983) Proc. Natl. Acad. Sci. USA 80, 1816–1820.
[18] Zerial, M., Salinas, J., Filipski, J. and Bernardi, G. (1998) Eur. J. Biochem. 160, 479–485.
[19] Smit, A.F.A. (1996) Curr. Opin. Genet. Dev. 6, 743–748.