194–199 Nucleic Acids Research, 1994, Vol. 22, No. 2

A rapid and simple method to isolate and characterize highly polymorphic markers from the centromeric regions of the human chromosomes

A.M.Laurent, B.Marçais, M.Muleris[1] and G.Roizès*
UPR 9008 CNRS and U 249 INSERM, Montpellier and [1]URA 620 CNRS, Paris, France

Received October 18, 1993; Revised and Accepted December 15, 1993

ABSTRACT

Using oligonucleotide primers complementary to the 3′ ends of either the Alu or the L1Hs consensus sequences in conjunction with a primer complementary to alpha satellite subsets specific to different human chromosomes, it was possible to detect and characterize polymorphisms originating from the microsatellites which are often present downstream these repetitive elements. The methodology does not require cloning, sequencing or synthesis of specific primers. Centromeric location was confirmed by linkage analysis, *in situ* hybridization and sequencing. The method is proposed for the generation of polymorphic markers from all centromeric regions.

INTRODUCTION

The construction of genetic maps has made important progress in the recent years and all human chromosomes are covered with more and more highly polymorphic markers, in particular with microsatellites of the $(CA)_n$ type (1,2). However, as seen from the recently published maps (3–5), the centromeres so far have not been amenable to such an approach . This is for two main reasons. First, centromeric regions are supposed not to recombine frequently (6), the consequence being that a larger number of meioses are necessary to estimate genetic distances. Second, one would need rapid and simple methods to handle such a great number of analyses, and these techniques have not yet been developed. Centromeric regions contain polymorphic markers, but when they are easily detectable, they are biallelic with allelic frequencies which are often too low to be usable. Alternatively, when they are highly polymorphic, Pulse Field Gel Electrophoresis (PFGE) is necessary, but too sophisticated a technique to be used routinely. Both types of marker originate from alpha satellite DNA sequences which are abundant on each human chromosome (7). A biallelic marker was found on chromosome 21 (8), but when analysed on the Ceph panel, it was found in only two families (9). As an example of the polymorphic markers, the alpha satellite DNA blocks present on chromosome 21, and on other chromosomes, are highly variable in length as evidenced by PFGE (10).

There have been few reports that retrotransposable elements are inserted within the centromeric alpha satellite DNA blocks. An Alu element has been found in the African Green Monkey (11) and L1 sequences in Man (12–14). It is therefore highly likely that such elements are to be found in all the human centromeres. At their 3′ ends, the two types of repeats possess more or less pure polydA runs of variable length (15). Moreover, at least in the Alu sequences, microsatellites, tri, tetra or pentanucleotides, are often found (16). The same is expected in the L1 elements although there has been no formal report published so far.

We have recently taken advantage of this property to develop a method we called '3′ Alu PCR' to detect and characterize these 3′ Alu microsatellites (17). For that purpose, one needs an oligonucleotide primer complementary to the Alu consensus sequence, the second one being any of those which are already available in the laboratory. With this method, a number of non-assigned highly polymorphic markers can be generated at once without any cloning, sequencing, or synthesis of specific primers. Their localisation can be determined by linkage analysis.

In this paper, we have applied this approach to generate polymorphic markers from the centromeric regions of the human chromosomes.

RESULTS AND DISCUSSION

Principle of the method

Alpha satellite DNA sequences are found at the centromere of all human chromosomes. However, they exhibit reduced sequence homology among non-homologous chromosomes so that almost every chromosome is characterized by one (or several) subset(s) of satellite DNA (18).This important property has been used to specifically recognize several specific centromeric alpha satellite DNA sequences. This also holds for PCR experiments since it has been shown that it is possible to amplify centromere-specific sequences with alpha satellite oligonucleotide primers (19,20).

A few reports have described the presence of Alu elements within the alpha satellite DNA blocks of the African green monkey centromeres. As they were discovered by chance, and

*To whom correspondence should be addressed

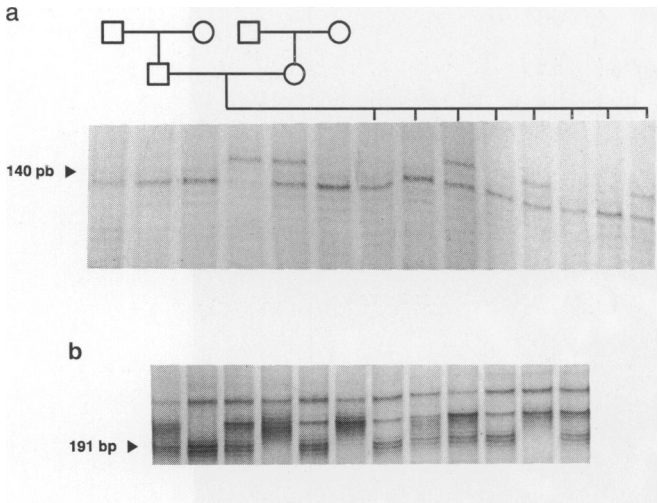

Figure 1. Polymorphic markers detected by the oligonucleotide primers Alu2 and 21A: two such polymorphic markers are shown in two different Ceph families, (**a**) and (**b**). The 191 bp alleles differ by 4 nucleotides whilst the 140 bp ones by 2 nucleotides.

not because they were looked for, we expect that a significant number of such Alu elements could have transposed to within the centromeric sequences of human chromosomes as well as the rest of the genome during the course of evolution.

In principle, therefore, an oligonucleotide primer complementary to the alpha satellite DNA sequences has to be used in conjunction with a second one complementary to the Alu or the L1 consensus sequence. The PCR products can then be resolved on sequencing gels and their polymorphic character determined by comparing banding patterns obtained with unrelated individuals. The chromosome specificity of the alpha satellite DNA sequences can be used to generate markers from different chromosomes. Once detected, the polymorphisms have to be checked as being of centromeric origin both by *in situ* hybridization and by linkage analysis. This methodology could be applicable to all human chromosomes by varying the sequence of the alphoid oligonucleotide primer and/or the satellite target in which transposable elements might have been inserted during evolution.

Centromeric polymorphic markers can be generated at once from several chromosomes

We designed oligonucleotide primers capable of initiating PCR amplification at the extremities of the Alu elements expected to be found within the centromeric alpha satellite DNA blocks.

When the centromere 21-specific oligonucleotide 21A was used as a primer, a series of DNA bands were detected in a sequencing gel, some being monomorphic, others polymorphic. Figure 1 shows those bands which exhibit length polymorphism when they are compared between different unrelated individuals. Two were clearly resolved in this sequencing gel. They correspond to labeled PCR products approximately 140 and 191 bp in length and were shown to be tri and biallelic respectively in the few unrelated individuals tested. Two others were not distinctly resolved in this gel, but they are probably both triallelic as deduced from analyses made in the same panel of unrelated individuals (results not shown). Linkage analysis was done on two informative Ceph families for the 140 and 191 bp bands. In these analyses, we were able to compare them to the most

Table 1. Oligonucleotide primers used in the study

NAME	NUCLEOTIDE SEQUENCE
Alu 2	5' TTGCAGTGAGCCGAGATCGCGCC 3'
L1Hs1	5' CATGGCACATGTATACATATGTAACATAACC 3'
Roalpha21	5' ACCCAGCCAAAGGAG 3'
21 A	5' TGATGTGTGTACCCAGCC 3'
α sat cons	5' TTGTGATGTGTGCGTTCAACT 3'
13 B	5' GCTATCCAAATATCCACT 3'

Cen proximal markers available at this time: D21S13, D21S1, D21S11 and D21S8. When compared to each other, the markers corresponding to the 140 and 210 bp bands were found to be distant by 0.2 cM and to be close to D21S13. These data should be interpreted with caution as they were obtained with only two informative Ceph families. However, *in situ* hybridization partly confirmed this conclusion since centromeres of chromosomes 1, 9, 13, 15,16 and 21 were stained when total PCR products were used as probe (not shown).

Polymorphic markers can be generated from all human centromeres

This last result is not surprising because oligonucleotide 21A was synthesized according to the consensus sequence of the alpha satellite DNA of chromosome 21 and it differs by only one nucleotide from that of chromosome 13 (21). However, the conditions of PCR amplification cannot be controlled to the point of preventing any other satellite domain of other chromosomes having a partial homology to initiate amplification, provided an Alu repeat is present in its vicinity. This is actually an advantage of the method, since one can therefore expect to generate polymorphic centromeric markers from all human chromosomes without the necessity of using a too large number of oligonucleotide primers.

When another oligonucleotide primer (Roalpha21), also specific for the chromosome 21 alphoïd sequences, was used instead of 21A, a totally different banding pattern was obtained in a sequencing gel (not shown). Again, several centromeres were stained by *in situ* hybridization, mainly on chromosomes 1, 9, 15 and 16 (Figure 2). Chromosome 21 centromere was, however, not detected. This might be due to the absence of Alu repeats close to the sequence corresponding to the primer used, or because the DNA band products originating from chromosome 21 are under-represented in this experiment in comparison with those corresponding to other chromosomes. Without further analysis we cannot conclude on this point. Several polymorphisms were, however, detectable in this experiment.

When an oligonucleotide primer representing the alpha satellite DNA consensus sequence was used (AlphaCons), again a different pattern was obtained showing a series of eight polymorphic bands in the same range of sizes as those of Figure 1 when 12 unrelated individuals were compared (not shown). *In situ* hybridization experiments showed signals at the centromeres of several chromosomes which are, by decreasing order of intensity, chromosomes 1,7, 6, 12, 19 and 4 (data not shown).

Finally, we used an oligonucleotide primer homologous to the alphoïd component of chromosome 13 (13B of Table 1). It appeared to generate also a multibanded pattern, three of the

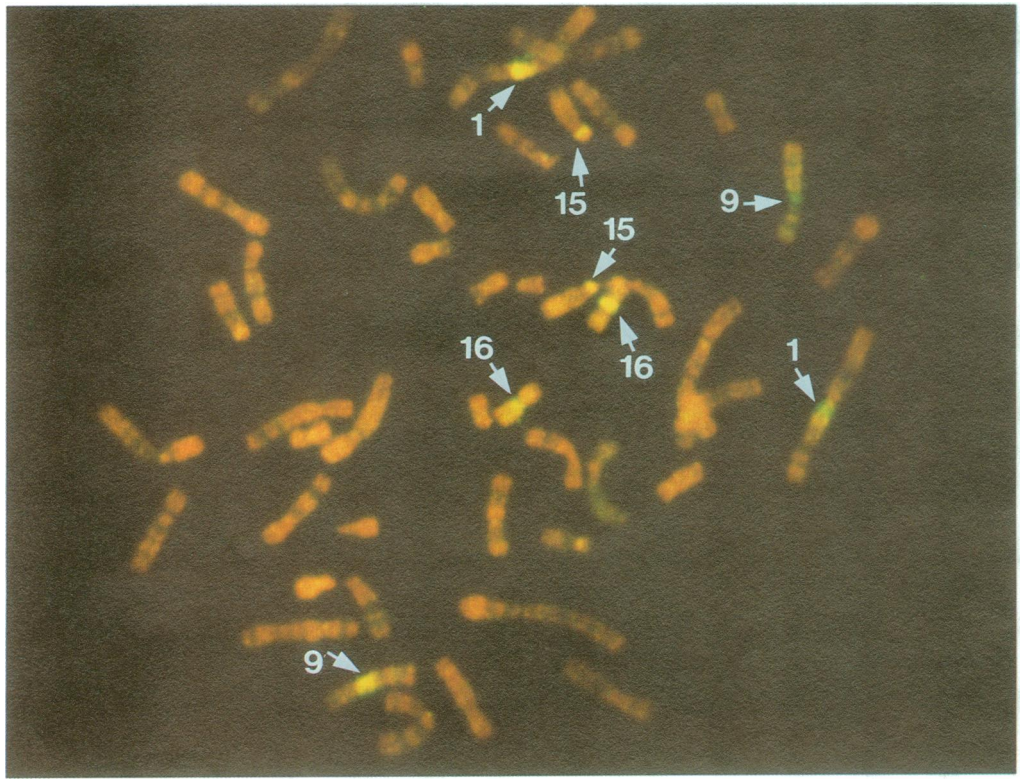

Figure 2. *In situ* hybridization to normal metaphase spreads performed with PCR product using oligonucleotide Roalpha 21 as primer, showing signals on the centromeres of chromosomes 1, 9, 15, 16.

Table 2. Allele frequencies of three of the centromeric markers

MARKER GENERATED BY	NUMBER OF ALLELES	NUMBER OF INDIVIDUALS TESTED	ALLELE FREQUENCIES							
			0	1	2	3	4	5	6	7
Alu 2+13 B	2	34		12	56					
Alu 2+13 B	8	34	13	2	8	6	17	2	18	7
L1HS1+ Roalpha21	4	36		4	19	47	2			

bands being clearly polymorphic (Figure 3). One band exhibited eight alleles (Table 2), one of which represented a zero allele (absence of the Alu element at the locus). For two of these, the DNA bands of certain different alleles were extracted from the sequencing gel and the sequence determined after PCR amplification (Table 3).

Centromeric L1 elements also possess polymorphic markers at their 3′ ends

As L1Hs elements have also been shown to have transposed within alpha satellite DNA sequences in the human genome, we have tested the possibility that the same approach could be used to detect the same type of polymorphisms. Figure 4 shows that this is indeed the case, a four alleles polymorphism being detected with the first couple of oligonucleotide primers tested in this paper: an oligonucleotide complementary to part of L1Hs consensus DNA sequence in conjunction with Roalpha21 (Table 1). A Linkage analysis localized it at the centromere of chromosome 2.

The method used does not need previous detailed knowledge of the loci to be analysed

In order to try to fill in the gaps which exist in the centromeric regions of the genetic maps produced so far, we developed a methodology for which the purpose has to detect and characterize polymorphic markers in these chromosomal regions.

It is based on an approach which was designed by us to detect the polymorphic microsatellites which are often present at the

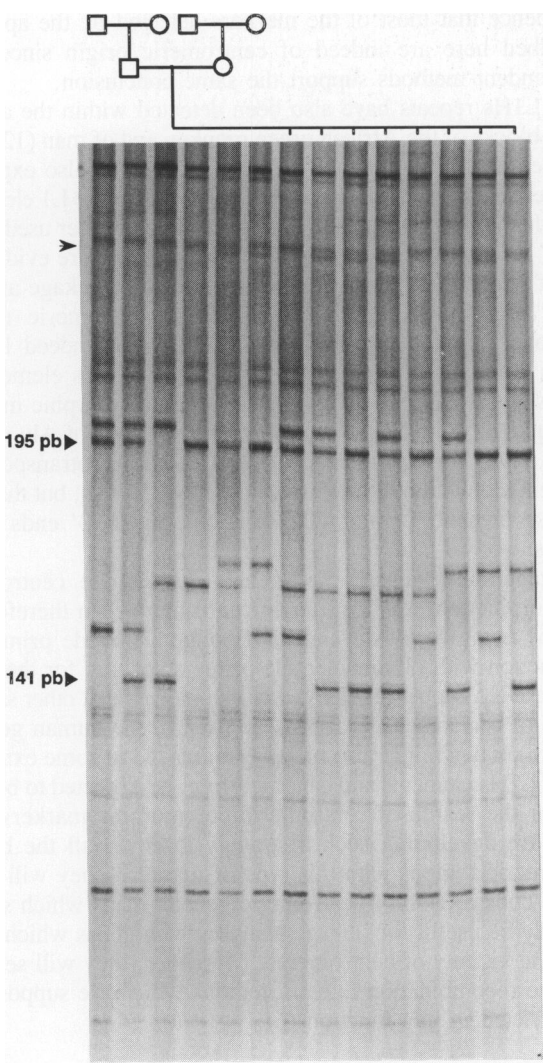

Figure 3. Polymorphic markers detected by the oligonucleotide primers Alu2 and 13B: The markers corresponding to 141 bp and 195 bp were analysed in details (sequences are shown in Table 3). A third non characterized marker is indicated by an arrow.

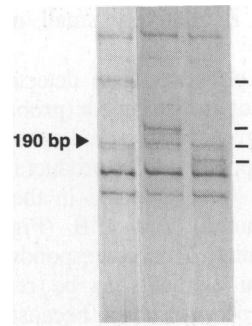

Figure 4. Polymorphic markers detected by the oligonucleotide primers L1Hs1 and RoAlpha21: The four alleles (1 to 4) corresponding to the 190 bp polymorphic bands are shown.

Table 3. Partial sequence analysis of three centromeric markers

Alu 2 →	CCCGGCAGCAAGAGATGAGATTCCATCTCAAA*(TAAA)*$_n$**GATAAATN**
	CTAAAGGAGCAGGCAAATACGC ← 13B
Alu 2 →	TGGGAACAGAGCAAGACTCTTGCT*(CAAA)*$_n$**GGGTACCAAGTGCTGG**
	AGTAGGGAAATAGAAGGTAATATTTGAGCAAAAACNGAAAAAGGT
	GAGTGGTAACTAGCTAA ← 13B
L1Hs1 →	CACATGTACCCTAGAACTTAAAGTA*(TAA)*$_n$**AGAAAACAAATACCACA**
	TGTGGGGGAAAAAAAAAAAAGGAAAAGGAAAGCAGAGATTCACT
	TTAAACACTGAGAACGTGCAATCGCT ← RoAlpha 21

The oligonucleotide primers used are indicated on both sides of each sequence. The microsatellites detected in each case have been determined as $(TAAA)_n$, $(CAAA)_n$ and $(TAA)_n$ in (a), (b) and (c) respectively. The DNA sequences, of presumably alphoïd type, are shown in thick letters.

3′ end of a number of Alu element (16), an approach called '3′ Alu PCR' (17). The application of this methodology to the Alu elements presumably present within the alphoid DNA sequences of the centromere of each human chromosome was tested by the polymerase chain reaction. Couples of oligonucleotide primers representing chromosome-specific alphoid sequences were used, each in association with a common primer complementary to a sequence close to the 3′ end of the consensus Alu repeat (Alu2 in Table 1).

This approach, like the original 3′ Alu PCR method, does not require any cloning or sequencing of microsatellites prior to testing their polymorphic character which is, as usual, performed on a series of unrelated individuals. It has the additional advantage of avoiding oligonucleotide synthesis for each locus to be tested as is the case for the classical detection of microsatellites of the $(CA)_n$ type (5). Preliminary results of this method are presented in this paper.

Several chromosome-specific alphoid oligonucleotides were used. As in these experiments the Alu oligonucleotide primer was labeled, two types of labeled amplification products were expected to occur, but the molecular proportion of 1 to 10 in favor of the alphoid specific oligonucleotide allowed to avoid those which would have been generated between two Alu repeats regardless of their position in comparison with the alphoid DNA blocks. When analysed on sequencing gels, the banding patterns, presumably originating from Alu sequences present within alphoid sequences, were totally different in different experiments. A number of polymorphic bands were clearly visible on the autoradiographs of the gels illustrated in Figures 1, 3 and 4 as well as in other examples not presented here.

An interesting feature of this approach is that several polymorphic markers are detected at one time in each experiment. The lengths of the DNA bands which correspond to these markers are spread over a large range and appear at different places in the same sequencing gel. It is, therefore, easy to characterize them by changing the conditions of migration if they appear to be difficult to be clearly characterized because of their large sizes. As judged from the few examples shown here, the minimum number of alleles is variable, ranging from two to eight, with frequencies in the populations which render them practically usable in linkage analyses since most of them were only detected

as polymorphic on a relatively small number of unrelated individuals.

The first type of polymorphism detected originates by the presence or absence of an Alu repeat (probably an Alu insertion polymorphism) at a given locus as judged by the presence or the absence of the corresponding PCR product in the sequencing gel. Two such examples were detected in the present study with RoAlpha21 (not shown) and 13B (Figure 4) as alphoid oligonucleotide primers. This corresponds to the well known property of the Alu elements to be retrotransposable, the consequence being a polymorphism because the element has not yet been fixed in the population. A number of such cases have been described in other regions of the human genome (17,22). As it is difficult to distinguish between heterozygotes and homozygotes when a 3′ Alu PCR product is detected, this type of polymorphism is of low interest in practice.

In the second type, the polymorphism is presumably the consequence of a variation of the number of dA residues at the 3′ end of the Alu element as judged from the fuzzy band detected in the gel. Such an example is given in Figure 1 by the 191 bp long band. However, most of these polymorphisms will not be usable if the difference in length between the different alleles is by an increment of one or two nucleotides only as in most cases (unpublished results, this laboratory).

Centromeric location is confirmed by three independent approaches

The third type a polymorphism which was the most commonly detected here, corresponds to the expected presence of microsatellites at the 3′ ends of a number of Alu repeats (16). The fact that they are true microsatellites (di, tri, tetra and even penta-nucleotide repeats) is shown by the sequencing gels which clearly indicate that the length polymorphisms of the different alleles found vary by increments of two, three, four and five nucleotides. Moreover, this has been confirmed in three examples by direct sequencing of the DNA bands purified from the sequencing gels (Table 3).Further, sequencing showed that the corresponding amplification products were indeed originating from Alu sequences localised within alphoid sequences as judged from the homologies found (Table 3). Homology with the Alu consensus sequence was up to 90%, whilst with alphoid sequences extracted from the GenBank Database, significant homologies were found with matches ranging from 65 to 80%.

Two other ways of showing the centromeric origin of these polymorphic markers were used in this study. The first one is *in situ* hybridization. This shows that most of the PCR products thus obtained are indeed of centromeric origin, which indicates that, most probably, this holds for those detected as polymorphic in the sequencing gels. But the main feature of these *in situ* analyses is that several centromeres are stained at once and differ from one oligonucleotide primer to the other even if they are specific for the same chromosome (21A and RoAlpha21 for instance. This reflects the high flexibility of the method which is expected, therefore, to yield polymorphic markers from all human centromeres if a series of alphoid specific oligonucleotide primers are to be used.

The exact location of each of the polymorphic markers detected can be made by using a few families of the Ceph panel. However, at this stage, the data presented here have to be taken with caution as the number of informative families was too small. A more systematic analysis is needed to confirm the locations found in all cases. Nevertheless, on the whole, it can be said with confidence that most of the markers detected by the approach described here are indeed of centromeric origin since three independent methods support the same conclusion.

As L1Hs repeats have also been detected within the alphoid DNA blocks of the African green monkey and of man (12−14), we used a similar approach as the 3′ Alu PCR, also expecting the presence of microsatellites at the 3′ end of some L1 elements. With the only alphoid-specific oligonucleotide primer used in this study, RoAlpha21, such polymorphic markers were evidenced, one of which being tetra-allelic and shown by linkage analysis to be originating from chromosome 2 centromeric region. Moreover, sequencing established that it was indeed located within alphoid DNA sequences close to an L1Hs element.

This last result shows that centromeric polymorphic markers are expected to be present not only at the 3′ ends of Alu repeats which have, during the course of evolution, retrotransposed at different loci within the alpha satellite DNA blocks, but that they are also present, with similar locations, at the 3′ ends of L1 repeats.

A general strategy to detect and characterize centromeric polymorphic markers for human centromeres can therefore be built by using both Alu and L1Hs oligonucleotide primers in conjunction with oligonucleotide primers specific for the alpha satellite DNA sequences. It must be emphasized that other satellite DNAs of centromeric locations do exist in the human genome (23−25). They might also have been invaded to some extent by Alu or L1Hs elements and can, therefore, be expected to be used also for the search of centromeric polymorphic markers.

We are developing such a strategy to cover all the human centromeric regions with this type of marker. They will serve for the construction of centromeric genetic maps which should certainly be useful for linkage analyses with genes which often lie in the vicinity of centromeres. Moreover, they will serve to estimate recombination rates in regions which are supposed to be restricted in this function (6).

METHODS

DNA samples

The DNA samples used in this study were obtained from Ceph and used directly for PCR amplification.

Oligonucleotide primers

The oligonucleotides used in this study are listed in Table 1.

PCR amplification

PCR experiments were performed in a PREM thermocycler using the buffer supplied by Promega, 50 μg^{-1} BSA (Bovine serum albumin), 0.2 mM each dNTP, 2.5 units Taq Polymerase (Promega) in a total volume of 50 μl. Samples were initially denatured at 92°C for 5 minutes, then at 92°C for 5 sec during each of the 30 amplification cycles. Annealing was performed for 30 sec at different temperatures as indicated in the text for each couple of oligonucleotides. Extension was at 72°C for 1 minute during the cycles, with a final extension step of 10 minutes. 2 picomoles of the Alu 2 or L1Hs oligonucleotides were added in combination with 20 picomoles of the oligonucleotides priming in the alpha satellite DNA sequences.

Detection and characterization of the polymorphic markers

Radiolabelling was done by 5′ end labelling the Alu 2 or L1Hs primer with T4 Polynucleotide kinase (Boehringer) and (gamma

S^{35}) ATP (Amersham). The PCR products were analysed on sequencing gels (5% acrylamide; 7M urea) using M13mp8 as size marker.

Sequence analysis

PCR products were directly sequenced according to a procedure developed by Casanova *et al.* (26).

In situ hybridization

PCR products were labeled with biotin-14-dATP using a nick-translation kit according to the suppliers' instructions (Bionick, Gibco BRL). *In situ* hybridization was performed as described by Lemieux *et al.* (27) with the following modifications: 50 to 200 ng of denatured probe DNA in hybridization buffer was added to the hybridization mixture. Hybridization was carried out overnight at 42°C. Post hybridization washes were performed twice in 50% formamide, 2×SSC and twice in 2×SSC at 45°C for 5 min each. Slides were then incubated with goat antibiotin antibody (Vector) diluted at 1:100 for 45 min at 37°C, washed once and incubated with fluorescein-conjugated rabbit anti-goat antibody (Biosys, Compiègne, France) diluted at 1:100 for 45 min at 37°C.

ACKNOWLEDGEMENTS

The authors are indebted to Christophe Person (Ceph) for performing linkage analyses with oligonucleotide 21A and to Jean Morissette (Généthon) for performing those with L1Hs. They also thank Jean Paul Charlieu for discussions. This work was supported by grants from AFM, ARC, GREG and regulat funding from CNRS and INSERM.

REFERENCES

1. Litt MP & Luty JA (1989) Am. J. Hum. Genet., 44: 397–401
2. Weber JL & MAY PE (1989) Am. J. Hum. Genet., 44: 388–396
3. (1992) Science, 258:148–162
4. Guyer MA (1992) Science, 258:67–86
5. Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vayssex G & Lathrop M (1992) Nature, 359:794–801
6. Charlesworth B, Langley C & Stephan W (1975) Genetics, 122: 47–49
7. Manuelidis L (1978) Chromosoma, 66: 1–21
8. Marçais B, Gérard A, Bellis M & Roizès G (1991) Hum. Genet., 86: 307–310
9. Jabs EW, Warren AC, Taylor EW, Colyer CR, Meters DA & Antonarakis SE (1991) Genomics, 9: 141–146
10. Marçais B, Bellis M, Gérard A, Pagès M, Boublik Y & Roizès G (1991) Hum. Genet., 86: 311–316
11. Grimaldi G & Singer MF (1982) Proc Natl Acad Sci USA, 79:1497–1500
12. Grimadi G & Singer MF (1983) Nucleic Acids Res., 11: 321–338
13. Potter SS (1984) Proc. Natl. Acad. Sci. USA, 81: 1012–1016
14. Higgins MJ, Wang H, Shtromas I, Haliotis T, Roder JC, Holden JJA & White BN (1985) Chromosoma, 93: 77–86
15. Orita *et al.* (1990) Genomics, 8: 271–278
16. Beckmann JS & Weber JL (1992) Genomics, 12: 627–631
17. Charlieu JP, Laurent AM, Carter DA, Bellis M & Roizès G (1992b) Nucleic Acids Res., 20: 1333–1337
18. Choo KH, Vissel B, Nagy A, Earle E & Kalitsis P (1991) Nucleic Acids Res., 19:1179–1182
19. Warburton PE & Willard HF (1992) Nucleic Acids Res., 20:6033–6042
20. Charlieu JP, Murgue B, Marçais B, Bellis M & Roizès G (1992a) Genomics, 14: 515–516
21. Jorgensen AL, Bostock C & Bak AL (1987) Proc Natl Acad Sci USA, 84:1075–1079
22. Perna NT, Batzer MA, Deininger PL & Stoneking M (1992) . Human biology, 64: 641–648
23. Prosser J, Frommer M, Paul C & Vincent VC (1986) J Mol Biol, 187:145–155
24. Menervi R, Agresti A, Della Valle G, Talarico D, Siccardi AG & Ginelli E (1985) J Mol Biol, 186:483–489
25. Waye JS & Willard HF (1989) Proc Natl Acad Sci USA, 86:6250–6254
26. Casanova JL, Pannetier C, Jaulin C & Kourilsky P (1990) Nucleic Acids Res.,13: 4028
27. Lemieux N, Dutrillaux B & Viégas-Péquignot E (1992) Cytogenet. Cell Genet., 59: 311–312