Nucleic Acids Research, Vol. 20, No. 14 3665–3669

A new strategy useful for rapid identification of microsatellites from DNA libraries with large size inserts

Bruno Baron, Christophe Poirier, Dominique Simon-Chazottes, Cécile Barnier and Jean-Louis Guénet
Unité de Génétique des Mammifères de l'Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France

Received March 30, 1992; Revised and Accepted June 26, 1992 EMBL accession nos X66850–X66854 (incl.)

ABSTRACT

Microsatellites are new powerful polymorphic markers used for gene mapping. Their characterization requires that all the sequence surrounding the repeat be known in order to be able to design primers for PCR amplification. However, when using DNA libraries with large cloned inserts, this sequence characterization is not immediately practicable. In this paper, we describe a new strategy, based both on the use of a microsatellite specific probing and on the creation of nested deleted clones with the Exonuclease III, in order to position microsatellites in a range allowing direct sequencing. This method was applied to the screening of a mouse chromosome 19 DNA specific library. In this way, thirteen clones were identified by specific probing and seven were submitted to the nested deletion strategy. Five of them presented microsatellite sequences in specific deleted subclones which were selected and sequenced. Primers were designed for each of them and polymorphism between the genomes of several inbred strain of mouse have been determined. These microsatellites were mapped, three of them to chromosome 19 and two to chromosome 11.

INTRODUCTION

Within the last few years the use of new strategies for gene mapping based on the analysis of polymorphisms at the DNA level (RFLPs) allowed the establishment of high resolution maps of the entire genome in several species. These techniques make it possible to map molecularly identified markers with a resolution of 0.5 to 1.0 cM (200–400 markers per chromosome) (DNA reports HGM 10 and 10.5, 1). However, in spite of their great value for the localization of recently cloned DNA sequences, these techniques are too slow and expensive to be suitable for random characterization of genetic markers.

Some of the drawbacks reported above have been partially overcome by the discovery of new kinds of polymorphic markers which it is possible to identify by PCR analysis. Among these markers some are made up of tandemly repeated short nucleotidic sequences referred to as microsatellites (2–4).

The origin and nature of these polymorphic sequences is not established but they may result from errors of the polymerase during DNA replication and/or from slightly unequal recombinations between homologous chromatids during meiosis. Two important features make these repeated sequences interesting markers. The first is that being untranslated, they may not be counterselected and thus constantly accumulate with time. The second is that originating probably from accidental process, they must occur at random in the genome. In fact, these microsatellites have been proved to be very useful markers for genetic mapping as they are highly polymorphic (5), and very common (between 10^5 and 10^6 per genome (2, 5)).

Characterization of microsatellites requires that the region flanking the repeat be sequenced to allow the designing of primers for PCR amplification. For this purpose, laboratories which are searching for totaly 'anonymous' microsatellites are generally using total DNA libraries with small sized cloned fragments, allowing direct sequencing of inserted DNAs. However, libraries with large size fragments may be notably advantageous as, for example, when constructing specific DNA libraries from chromosomes purified by flow cytometry (9, 10). With such libraries, direct sequencing is not possible and it is generally necessary to subclone and rescreen the subclones to be sequenced.

In this paper we report an alternative method for direct characterization of microsatellites from clones of libraries with large sized inserts. This method is based on the combination of dinucleotide repeat specific probing and on the production of nested deleted clones using exonuclease III in order to detect microsatellite containing clones and to allow the direct sequencing of these microsatellites. This method was successfully applied to the screening of (CA) tandem repeats contained in a mouse chromosome 19 specific DNA library (11).

* To whom correspondence should be addressed

MATERIAL AND METHODS

DNA library

The DNA library was constructed in a λ NM1149 vector (12), from mouse chromosomes 19 purified by flow cytometry (11). The DNA fragments, ranging from 0.5 to 9kb in length, with an average size of 4kb, were cloned in the *Eco*RI site of the λ phage.

$(CA)_n$ specific detection

About 3.10^3 colonies were plated and transfered to colony plaque-screen filters (NEN). The identification of colonies carrying repetitive sequences was performed by hybridization with sonicated mouse DNA radiolabelled by random priming (11, 13). The filters were then dehybridized (0.5N NaOH, 45°C, 30 min) and rehybridized with $(CA)_n$ specific probes. (CA) specific labelling was obtained using a poly $(CA)_n-(GT)_n$ heteropolymer template (Pharmacia), and a $(CA)_3$ primer. For each assay, 100 ng of heat-denaturated template was incubated with 3ng of primer in the presence of [α-^{32}P] dCTP, cold dATP, dGTP and dTTP, and Klenow fragment. Hybridization was performed in 1 mM EDTA, 0.5 M $NaHPO_4$ pH 7.2, 7% $NaDodSO_4$ (14). Blots were washed at 65°C in 0.1× SSC, 0.1% SDS.

Subcloning and production of nested deletions with Exonuclease III

The selected clones were individually amplified and their DNA extracted as previously described (11). After *Eco*RI digestion of the clones, their inserts were purified from agarose gel by selective binding to glass beads (Sephaglas Band prep, Pharmacia) and ligated with dephosphorylated *Eco*RI digest of pT7T3 18U multifunctional phagemid (Pharmacia).

The unidirectional deletion method of Henikoff (15), with Exonuclase III, was performed with the double-stranded Nested Deletion Kit (Pharmacia). Briefly, about 3μg of each plasmid DNA was submitted to double digestion with enzymes corresponding to unique sites of the Multiple Cloning Site and giving linearized molecules having one susceptible site at one end and one resistant site at the other. Deletions were performed with experimental conditions adjusted to obtain a rate of digestion of 100 bp/mn.by adding 100 units of Exonuclease III in a final volume of 40μl of buffer containing 75mM NaCl salt, and incubating at 30°C. 2μl was taken every 3 min, thus corresponding to DNA samples differing by 300 bp in length, and treated with nuclease S1. One half of each aliquot was then loaded on a 0.8% agarose gel and electrophorezed at 50V for 2−3 hours. DNAs were then transfered to nylon filters (Hybond N^+, Amersham) under vacuum as previously decribed (11). Filters were then directly hybridized with (CA) specific probe as described above.

DNA sequencing

The remaining DNA of each clone was recircularized and used to transform competent NM 522 bacteria. Random colonies selected from each transformation were used for the preparation of single strand DNA, according to Blondel and Thillet (16). Single strand DNA was purified by selective binding on glass beads (Sephaglas Phage Prep Kit, Pharmacia). Sequencing was performed on an Automated Laser Fluorescent sequencer A.L.F. (Pharmacia), using fluoresceinated universal or reverse primers of M13 phage.

DNA samples and PCR experiments

DNA samples were prepared from the spleen of different inbred strains or the offspring of an interspecific backcross segregating several specific molecular markers (17).

PCR primers were designed for each clone in order to prevent self-annealing and achieve equivalence of T_m. They were chosen as close as possible to the microsatellite proper, in order to increase the proportion of the tandemly repeated sequence in the PCR product while minimizing its size. PCR was performed by mixing 25 μg of each primer and 0.5 Units of *Taq* polymerase (Promega) with 100ng of DNA, in 10mM Tris-EDTA (pH 8.4); 0.1% Tween, 50mM KCl; 1 to 3mM $MgCl_2$; 0.2mM dNTPs (Pharmacia). PCR procedure was then carried out as previously described (18). PCR products were analyzed on 4% agarose gels (NuSieve agarose (FMC Corp) and Type II agarose (SIGMA) in a ratio of 3:1), in TBE containing 5 μg/ml ethidium bromide, or on 10% non denaturating polyacrylamide gel, stained by incubation in 0.5 μg/ml ethidium bromide in water.

Genetic localization

Size variations of the amplified region were checked using DNA samples prepared from wild-derived inbred strains and laboratory inbred strains. When a suitable length polymorphism was observed in the PCR amplified products between the C57BL/6 laboratory inbred strain and (SPE), a moderately inbred strain of the *Mus spretus* species, its segregation pattern was checked on a panel of 75 DNA segregating for 283 molecular markers spanning the whole of the genetic map (17). Data were then added to the laboratory database and analyzed with the computer program GENE-LINK (19). When the polymorphism was detected among laboratory inbred strains, its segregation was checked on DNA samples from the corresponding set of recombinant inbred strains purchased from the Jackson Laboratory (Bar Harbor, Maine USA), and data were studied with the help of the computer program R.I. Manager (20).

RESULTS

Specific detection of (CA) repeats

Recombinant λ clones were screened successively for the presence of total repetitive DNA and of $(CA)_n$ repeat (Fig.1A). Comparing the patterns obtained with the two labelling procedures, we observed 3 types of responses:

— clones that were positive when probed with total repetitive DNA but negative or weakly positive with (CA) probe (20%);

— clones that gave similar intensities with both labelling methods, which may thus contain CA rich sequences, possibly including microsatellites (60%);

— clones that exhibited a stronger signal with the (CA) probe than with the total repetitive probe (20%). These clones, which were presumably made out of (CA) tandem repeats, were preferentially selected for subsequent analysis.

A total of fourteen clones were selected and individually amplified. All inserts were purified on agarose gel and transfered into phagemid vectors. The size range of inserts was 1.3 to 9 kb, with an average of 3.6 kb. For clones with more than one insert (5 out of 14), an aliquot of each fragment was dot-blotted on filters which were then hybridized with the (CA) specific probe. Two types of responses were observed, as presented on Fig.1B. Four clones showed a specific labeling with one insert only, which was consistent with defined (CA) microsatellite

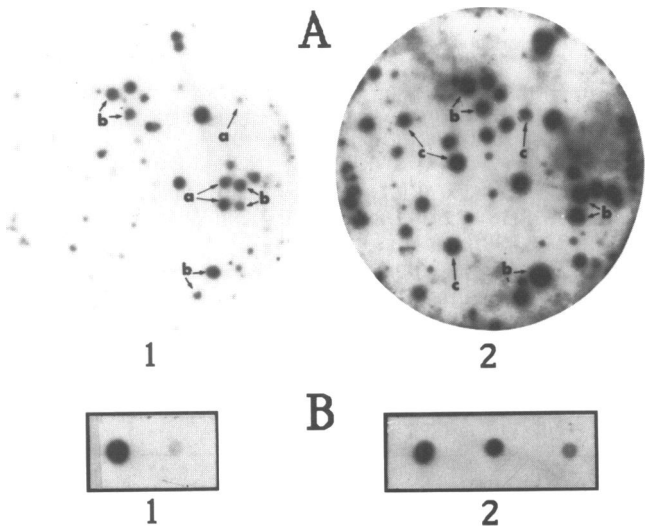

Figure 1. Detection of (CA) repeat containing clones. A: phage colonies transfered on nylon filters and hybridized with (1) total sonicated mouse DNA or (2) CA repeat specific probe. Arrows indicate examples of signals (a): stronger with total repetitive labelling, (b) equivalent with both labelling methods and (c): stronger with CA specific labelling. B: dot-blot hybridization of (CA) specific probe on DNA of clones containing multiple inserts. (1): example of clone presenting a specific labeling with single insert consistent with the presence of a microsatellite; (2): clone presenting more diffuse signal with both inserts, probably corresponding to diffuse CA rich sequence.

sequence (Fig. 1B1). In one clone, the three inserts hybridized with the (CA) probe with different intensities (Fig. 1B2) which seemed to correspond to unlocalized diffuse CA rich sequences more than localized CA tandem repeats. This clone was discarded.

The thirteen remaining clones were digested using two restriction enzymes corresponding to the polylinker of the plasmid and giving respectively one susceptible and one resistant end for digestion with Exonuclease III (15). Seven clones were first selected as lacking internal sites for the two enzymes, thus yielding a single linear fragment available for nested deletion experiment (Fig.2). One half of each sample was loaded on an agarose gel, submitted to electrophoresis (Fig 3A), transferred on nylon filters and hybridized with (CA) specific probe (Fig 3B).

For one clone the signal seemed to decrease according to the successive deletions (Fig.3B). This clone was excluded from the analysis as this could not be attributed to a discrete microsatellite. The six remaining clones gave the same pattern as this shown on Fig 3B1, i.e. a complete loss of the signal when the deletion removed a particular location, as expected for a microsatellite.

For each of the six remaining clones, we selected the samples corresponding to deletions preceding signal extinction. The remaining DNA was circularized and used to transform competent bacteria. For each sample, three candidates exhibiting the expected insert size were selected and sequenced.

Microsatellite sequences were found in at least one of the selected sub-clones for five out of six clones: four of them corresponds to typical (CA) repeats (clones 3−5, Table 1) and one to a more complex microsatellite structure (clone 1, Table 1). However, in one clone, it was not possible to find any (CA) repeat, or any other kind of microsatellite.

Primer sequences (20 mers) were designed for each of the five microsatellites. Four microsatellites exhibited a polymorphism

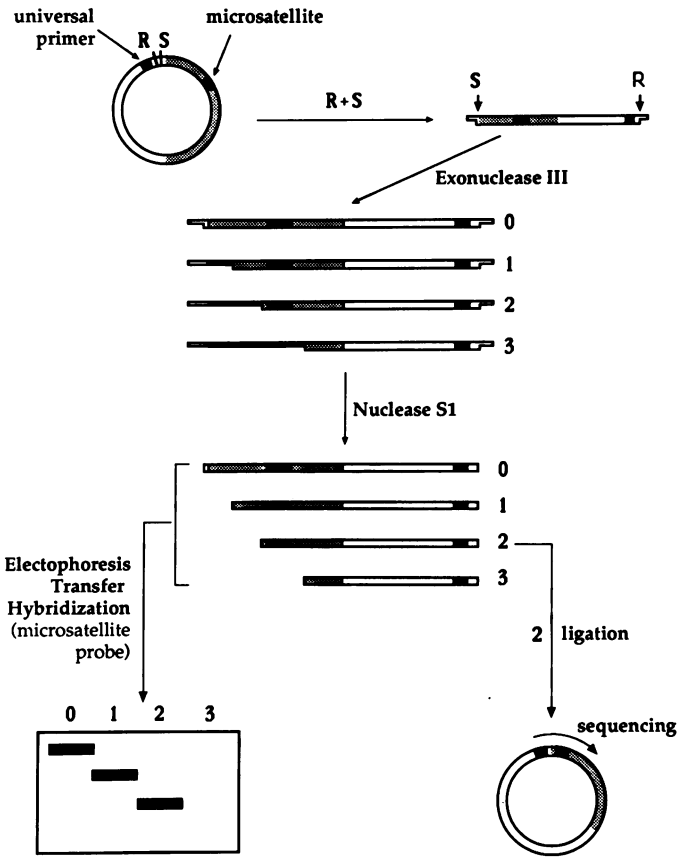

Figure 2. Schematic representation of the strategy used to obtain microsatellites in a place allowing direct sequencing. R and S correspond to the site of restriction enzymes giving ends respectively resistant and sensitive to the action of exonuclease III.

between the C57BL/6 laboratory strain and the moderately inbred strain (SPE), and thus were mapped to a specific chromosome. Three mapped to the Chr 19, and one to Chr 11. The fifth clone exhibited a size polymorphism only between the laboratory inbred strains DBA2 and C57BL/6 but not with *Mus spretus* as frequently observed. This clone was then localized on mouse Chr 11 using the BXD set of recombinant inbred strains (21). Complying with International Nomenclature, these polymorphic DNA segment have been named *D19Pas4* to*6* and *D11Pas5* to *6*, The linear ordering of the three microsatellites localized on the Chr 19 can be presented as follows:

Cen... − ***D19Pas6***−5.3 ± 3.0cM−*Fth* −30,8 ± 3.0cM− *D19Nds1* −11.9 ± 3.0cM.... ....−***D19Pas5*** −9.4 ± 2.7cM−***D19Pas4***

where, *Fth* (ferritin heavy chain locus) and *D19Nds1*(an anonymous probe) represent two markers previously localized on the consensus map of the Chr 19 (22).

The other two microsatellites were localized on Chr 11 by analyzing two different data bases. It is not thus possible to give a linear ordering with both of them. *D11Pas5* was located as follows:

Cen...−*Sparc* −12.1 ± 5.7cM−***D11Pas***5 −11.5 ± 6.3−*Hox-2*

D11Pas6 cosegregated without recombination with *Zfp-3* (Zinc Finger Protein-3) and *Asgr−1* (Asialoglycoprotein Receptor-1),

Table 1

microsatellite designation	sequences of the primers used for PCR amplification	repeat unit	product size in C57BL/6 (bp)	size variation
D19Pas4	ACAGCCTTTTTATTGGTGGT CAGGAGTACCCAGAGAAACC	$(TTTC)_7-(T)_{23}$	135	SEG>(BALB/c=PWK)> (A/J=AKR/J=C3H/He=C57BL/6=C57/L =DBA/2=DDK=SWR=129/Sv=NON=NOD)>NZB
D19Pas5	CCATAGCAACGGGAAAAGAA CGTTGTTCACATGCCACCTGC	$(AC)_{12}-(CA)_9$	148	SPE=PWK>(AKR/J=C3H/He=A/J=BALB/c= C57BL/6=C57/J=DBA/2=SWR=129/SV=DDK= AZB=SJL)
D19Pas6	CTCCCCCATCTGACTTTCTC GTAGAGGGTGAGGGTGTGCG	$(CA)_9$	110	A/J=BALB/c=C57/L=DDK=AKR/J= SWR=C3H/He=DBA/2=129/Sv= C57BL/6=NON=NOD=SJL SPE-
D11Pas5	GGGTAGGCAAGGTGGCTCAG CCCTCCCATTCTTTTCCCCT	$(CA)-(CAAC)-(CA)_{13}$	140	129/Sv>DBA/2>C57BL/6=SPE
D11Pas6	CCAGCCTACCCACATTCATT TTTTCCTAACCAGCTAAGTG	$(CA)_{21}$	300	(C57BL/6=DBA/2=AKR/J=C57/L =BALB/c)>(SPE=DDK)

with a confidence interval of 4.5cM at 95% probability. As these two markers are separated by about 4.5cM on the Chr 11 (22), the localization of *D11Pas6* can be estimated as follows:

Sparc $-$***D11Pas6*** $-$*Asgr*

DISCUSSION

A strategy was developed to allow a rapid characterization and targetted sequencing of microsatellites from large cloned inserts. This procedure has two main original features. The first is a specific and efficient probing of clones containing true (CA) repeats, which allows to significantly reduce the number of clones to be sequenced. The second is the use of the controlled unidirectional deletion methodology of Henikoff (15), coupled with the selective hybridization procedure, to obtain shortened clones allowing direct sequencing.

The (CA) specific labelling was done by priming with short dinucleotide repeats instead of random hexamers as generally used for the random priming (8, 23). In this experiment, we used a poly (CA)$-$(GT) template and $(CA)_3$ hexamers, in order to screen for (CA) repeats which are reported to be the most frequent microsatellites in mammalian genomes (2, 5). The procedure however can be extended in the same way to other polymer/hexamer pairs, for the detection of (CT), (AT) or (GA) repeats. In our experiment, we used a 6/100 ratio in weight between template and primers, allowing to prime 1 hexamer every 100 bp and leading to the synthesis of DNA fragments with an average size of 75 bp. One advantage of this procedure is that with only a slight modification of template/primer ratio, one can easily obtain longer or shorter sized probes when needed.

The specificity of the probe allows the elimination of non specific clones with only (CA) rich regions and the selection of the best candidates to contain true $(CA)_n$ repeats. Indeed, the comparison between the hybridization signals with a total repetitive probe and a (CA) specific probe allowed fourteen clones to be selected, presenting enhanced hybridization patterns with (CA) probe. The specificity was also useful for the elimination of clones presenting dispersed (CA) rich regions instead of clearly defined (CA) repeat (Fig 1B and 3B).

In order to characterize rapidly the microsatellites, it was necessary to obtain the (CA) repeat within a size range compatible with direct sequencing. Therefore, we performed the deletions in order to produce nested deleted clones with a 300 bp difference

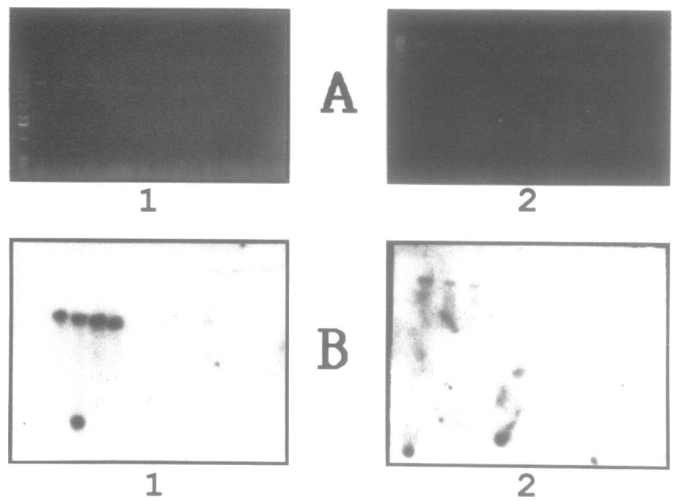

Figure 3. Characterization of nested deleted clones containing microsatellites. (A): electrophoretic separation of nested deleted clones, and (B): hybridization with (CA) specific probe of corresponding DNA transferred on nylon filters. (1): signal obtained with a clone presenting a localized (CA) repeat specific sequence and (2): signal obtained with a clone presenting a diffuse signal decreasing with the deletions.

in size between each deletion step. Indeed, considering that with standard methods one can unambiguously sequence 300 pb, by selecting two successive deletions with an expected overlap, it was possible to sequence a total of 500 bp or more. An alternative strategy could be to perform random digestion of inserts, subcloning and screening of subclones. However, if longer sequences are required, the nested deletion strategy presents an important advantage as sample deletions representing flanking regions are immediatly available, without the need of further screening and walking over subclones.

Four clones out of six which were sequenced showed specific (CA) repeats. For the two remaining clones, one presented a complex (CT) repeat (D19Pas6, Table 1), and the other contained no tandem repeat in the 500 bp which were finally sequenced. Considering that these two clones gave an hybridization signal similar to that presented in Figure 3A, we can expect that a (C-A) repeat is close to the sequenced portion, and that isolation and sequencing of other subclones would have made their characterization possible. In other published works, repeats were characterized in only 1/4 of selected clones or less (8). Therefore,

the use of a fast and precise hybridization procedure allows one to reduce significantly the number of clones to be sequenced.

Three microsatellites out of five were found localized on Chr 19 and two were localized on Chr 11. Adding these results with those obtained during the characterization of the library (11), we conclude that more than 70% of the DNA sequences of our library are indeed specific of mouse Chr 19. As the number of microsatellites on mammalian genomes was estimated at least at 10^5 (2, 5), Chr 19 (2,73% of mouse haploòd genome) must contain about 2,73%, i.e. more than 200 of these repetitive markers. This methodology, applied to the screening of the mouse Chr 19 specific DNA library, will allow the initiation of an extensive search and characterization of microsatellites specific from this chromosome.

ACKNOWLEDGEMENTS

This work was supported in part by a grant from the EEC (EUROGEM) and in a part by a grant from the Association Française contre les Myopathies. The authors thank Patricia Baldacci and Jean Weissenbach for reviewing the manuscript.

REFERENCES

1. Hillyard, A.L., D.P. Doolittle and M.T. Davisson (1991) *Mouse Genome*, **89**, 16–30.
2. Hamada, H., M.G. Petrino and T. Kakunaga (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6465–6469.
3. Weber, J.L. and P.E. May (1989) *Am. J. Human. Genet.*, **44**, 388–396.
4. Litt, M. and J.A. Luty (1989) *Am. J. Human. Genet.*, **44**, 397–401.
5. Weber, J.L. (1990) *Genomics*, **7**, 524–530.
6. Love, J.M., A.M. Knight, M.A. McAleer and J. Todd (1990) *Nucleic Acid Res.*, **18**, 4123–4130.
7. Hamada, H., M.G. Petrino, T. Kakunaga, M. Seidman and B.D. Stollar (1984) *Mol. Cell. Biol.*, **4**, 2610–2621.
8. Cornall, R.J., T.J. Aitman, C.M. Hearne and J.A. Todd (1991) *Genomics*, **10**, 874–881.
9. Van Dilla, M.A. and L.L. Deaven (1990) *Cytometry*, **11**, 208–218.
10. Baron, B., P. Métézeau, D. Hatat, C. Roberts, M.E. Goldberg and Bishop C. (1986) *Somat. Cell Molec. Genet.*, **12**, 289–295.
11. Baron, B., P. Métézeau, H. Kiefer-Gachelin and M.E. Goldberg (1990) *Biol. Cell*, **69**, 1–8.
12. Murray, M. (1983) In Laboratory, C.S.H. (ed.), Lambda II, Cold Spring Harbor: New York, pp.295.
13. Benton, W.D. and R.W. Davis (1977) *Science*, **196**, 180–182.
14. Church, G.M. and W. Gilbert (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.
15. Henikoff, S. (1984) *Gene*, **28**, 351–359.
16. Blondel, A. and J. Thillet (1991) *Nucl. Acids Res.*, **19**, 181.
17. Guénet, J.-L. (1986) (ed.), Current Topics in Microbiology and Immunology., Springer Verlag: Berlin Heidelberg, Vol.127, pp.109–113.
18. Serikawa, T., Kuramoto, T., Hilbert, P., Mori, M., Yamada, J., Dubay, C.J., Lindpainter, K., Ganten, D., Guénet, J.-L., Lathrop, G.M. and Beckmann, J.S. (1992) *Genetics*, (inPress)
19. Montagutelli, X. (1990) *J Heredity*, **81**, 490–491.
20. Manly, K. and E. R.W. (1991) *Mammalian Genome*, **1**, 123–126.
21. Taylor, B. (1989) In Lyon, M.F. and Searle, A.G. (ed.), Genetic Variants and Strains of the Laboratory Mouse, Oxford University Press: Oxford, pp.773–796.
22. Guénet, J.-L. (1991) *Mammalian, Genome.*, **1**, S306-S317.
23. Hearne, C.M., M.A. McAleer, J. Love, T.J. Aitman, R.J. Cornall, S. Gosh, A. Knight, J.-B. Prins and J. Todd (1991) *Mammalian Genome*, **1**, 273–282.