Efficient high-resolution genetic mapping of mouse interspersed repetitive sequence PCR products, toward integrated genetic and physical mapping of the mouse genome

LINDA MCCARTHY*, KENT HUNTER†, LEONARD SCHALKWYK*, LAURA RIBA†, SIMON ANSON*, RICHARD MOTT*, WILLIAM NEWELL*, CHARLOTTE BRULEY*, ISABELLE BAR‡, ELANGO RAMU†, DAVID HOUSMAN†, ROGER COX*§, AND HANS LEHRACH*

*Imperial Cancer Research Fund, London WC2A 3PX, United Kingdom; †Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139; ‡Facultés Universitaires Notre Dame de la Paix Medical School, Namur, Belgium; and §The Wellcome Trust Centre for Human Genetics, Oxford OX37BN, United Kingdom

Contributed by David Housman, December 28, 1994

ABSTRACT The ability to carry out high-resolution genetic mapping at high throughput in the mouse is a critical rate-limiting step in the generation of genetically anchored contigs in physical mapping projects and the mapping of genetic loci for complex traits. To address this need, we have developed an efficient, high-resolution, large-scale genome mapping system. This system is based on the identification of polymorphic DNA sites between mouse strains by using interspersed repetitive sequence (IRS) PCR. Individual cloned IRS PCR products are hybridized to a DNA array of IRS PCR products derived from the DNA of individual mice segregating DNA sequences from the two parent strains. Since gel electrophoresis is not required, large numbers of samples can be genotyped in parallel. By using this approach, we have mapped >450 polymorphic probes with filters containing the DNA of up to 517 backcross mice, potentially allowing resolution of 0.14 centimorgan. This approach also carries the potential for a high degree of efficiency in the integration of physical and genetic maps, since pooled DNAs representing libraries of yeast artificial chromosomes or other physical representations of the mouse genome can be addressed by hybridization of filter representations of the IRS PCR products of such libraries.

Mouse genetics is one of the most powerful tools for the analysis of almost all aspects of vertebrate biology including development, physiology, and pathobiology. Mouse models have already been developed for many human diseases (1–6). The mouse especially offers major advantages for the molecular analysis of genetically defined phenomena by positional cloning techniques, leading, for example, to the molecular identification of the Brachyury gene (7), based to a large extent on genetic markers generated by microdissection/microcloning approaches.

To extend this approach to a global analysis of the genes in the mouse genome, highly efficient genetic mapping techniques have to be available, especially if genes involved in multifactorial traits are to be analyzed. Genetic crosses involving several thousand individuals can bring great precision to this kind of analysis, provided genotyping of closely spaced markers can be carried out at high efficiency. This will, in an ideal situation, result in a mapping system of sufficiently high resolution to be able to uniquely identify a single gene within the genome. Since we expect on average, genes to be spaced at 40-kb intervals, corresponding to a genetic resolution of 20–25 millimorgans, >4000 progeny would have to be analyzed to allow such a high resolution.

The second stage of this process, the molecular identification and analysis of the corresponding genes, can be carried out most efficiently if a physical map of the mouse genome is available that can be easily integrated with the genetic map. It is therefore important that the markers described here can be used to efficiently screen yeast artificial chromosome (YAC) clone libraries, allowing the construction of integrated physical and genetic maps of the genome, based on YAC clones identified by markers located with millimorgan resolution on the genetic map of the mouse.

Though in theory very high resolution can be provided by the use of simple repeat polymorphisms, this marker system is not well suited to very high throughput applications, since usually each genotype requires one gel lane, making the scoring of large crosses very costly. Similarly, the screening of large YAC libraries by PCR with large numbers of markers, though clearly feasible, is unnecessarily work-intensive and costly.

We have therefore developed a mapping approach, designed to exploit a readily available source of polymorphisms across the genome, to require little backcross DNA for mapping, to be suitable for automation in large-scale genetic and physical mapping projects, and to be relatively inexpensive to use. The method we describe here meets these criteria.

Our method depends on the use of interspersed repetitive sequence (IRS) PCR to generate DNA representations of the mouse genome that differ from strain to strain at many chromosomal sites. By using hybridization of a cloned IRS PCR product to a large number of IRS PCR products from individual animals spotted on a rehybridizable filter, many simultaneous genotyping assays can be performed on a large number of animals. The method is thus economical on time and materials, requiring just one hybridization to map each probe to high resolution, and has the advantage of generating probes that are both unique and easily hybridized to YAC library filters and Southern blots. This makes them ideal for quickly identifying YACs that map to particular loci, without going through the tedious and expensive multiple rounds of PCR required for sequence-tagged-site screening (8–11). By using this method, IRS PCR products can be genetically mapped and screened on genomic YAC libraries in a single hybridization, integrating the physical and linkage mapping efforts.

MATERIALS AND METHODS

C57BL/6 (B6) × *Mus spretus* (LS) Backcrosses. The progeny used were from the backcross to LS. DNAs were spotted

from 423 animals from the EUCIB (European Backcross Collaborative Group, Harrow, U.K.) and 94 Jackson Laboratory backcross animals (12). For the D_Leh probes, total genomic DNA from the parental strains and backcross animals was amplified by using a Cetus 9600 thermal cycler. The B1.2 primer is as described in Cox *et al.* (13) but lacked the 5' *Not* I sequence and included 5' cloning arms for annealing into the pAMP1 plasmid. DNA (100 ng) was added to 1 μ g of B1.2 primer/all four dNTPs (each at 0.125 mM)/10 mM Tris-HCl, pH 8.3 (at 25°C)/50 mM KCl/1.5 mM MgCl₂/2.5 units *Taq* polymerase (ABT), in a final volume of 100 μ l. A 5-min 94°C denaturation step was followed by 30 cycles of 94°C for 45 sec, 50°C for 1 min, and 72°C for 3 min, with a final 5-min extension at 72°C. Amplification of C57BL/6J for the D_Hun loci clones was performed essentially as described in Hunter *et al.* (14).

To size fractionate the PCR products, the product smear was electrophoresed on a gel. By using a device containing 30 razor blades spaced 2 mm apart, 6 cm of DNA smear was sliced into 29 size fractions. Each of these size fractions was then used in a secondary PCR with the same B1.2 primers, annealed in the pAMP1 plasmid (GIBCO/BRL), and electroporated into XL-1 Blue *Escherichia coli* cells. Non-size-fractionated IRS PCR products were also cloned in pAmp1.

Subtractive hybridization (15) was used to remove nonpolymorphic B6 IRS PCR products to generate a sublibrary of clones with an increased rate of $\pm$ polymorphism. LS DNA was PCR-amplified by using a biotinylated version of the B1.2 primer. These LS products were denatured and bound to streptavidin-coated magnetic beads (Dynal Dynabeads M-280 Streptavidin, Great Neck, NY). B6 IRS PCR products amplified with the unbiotinylated version of the B1.2 primer were hybridized to a 400-fold excess of the bound LS sequences. Two rounds of subtractive hybridization were carried out in 50 mM Hepes, pH 7.6/0.4 mM EDTA/0.5 M NaCl at 65°C for 3 h. After each hybridization, the unhybridized B6 products were removed. The final products were cloned as described above. The D_Hun clone library was constructed by digesting C57BL/6J inter-B1 products with *Eco*RII and performing one round of subtraction with a modified JBgl adaptor set, essentially as described by Lysitsin *et al.* (16).

Screening $\pm$ Polymorphisms on Spotted PCR Products of Animals Backcrossed to *M. spretus*. For D_Leh probes, $\approx$ 1.5 μ l of PCR product per backcross animal was robotically applied to nylon filters in a grid pattern. Probes showing presence/absence ($\pm$) polymorphism were labeled by random hexamer priming (17) with both dATP and dCTP. These probes were denatured at 100°C for 6 min with 105 μ g of LS B1 PCR products in 6 $\times$ SSC/0.1% SDS (200 μ l, final volume) and incubated at 65°C for 75 min. Hybridizations were performed in 0.5 M sodium phosphate, pH 7.2/7% (wt/vol) SDS/1% bovine serum albumin/1 mM EDTA (18) with the probe at 10⁶ cpm/ml at 65°C overnight. Each autoradiogram was scanned into an image file. Image analysis of these files with normalization of the spots for DNA concentration allows the assignment of a positive/negative value to each backcross animal. Autoradiograms were also manually scored as a control measure. The segregation data are entered in the EUCIB database to obtain a map position for each probe (19). The D_Leh probes are accessible in the Reference Library DataBase 2 (20) through World Wide Web (<http://gea.lif.icnet.uk/>). To access the data in the Mouse Backcross Database (MBX) at the Human Genome Mapping Project Resource Centre by email use helpdesk@hgmp.mrc.ac.uk. For D_Hun probes, DNAs from the Jackson backcross were amplified, spotted on filters, and hybridized essentially as described (14). Films were read manually, and the data were entered into The Jackson Laboratory database and analyzed by using the program MAPMANAGER version 2.6 (21). D_Hun data can be accessed on The Jackson Laboratory mapping panel through World Wide Web via The Jackson Laboratory Home Page (<http://www.jax.org>) or by email to lbr@aretha.jax.org, or in the Mouse Genome Database with accession numbers.

www.jax.org) or by email to lbr@aretha.jax.org, or in the Mouse Genome Database with accession numbers.

To facilitate the hybridization of large numbers of probes to the backcross filters with minimal effort, a large-scale hybridization box was designed. This perspex box was divided into 48 narrow chambers, each of which could contain two hybridization filters. The filters are supported on a rigid polyester backing to allow them to be easily manipulated in the chambers. Probes are labeled in microtiter plates and can be added to the hybridization box chambers by using a multichannel pipet. At 48 hybridizations per hybridization box, 192 hybridizations can be comfortably performed overnight, per person.

Hybridizations to YAC Library Filters. Each YAC clone in the Imperial Cancer Research Fund (ICRF) library was PCR-amplified by using the B1.2 primer, as described above for PCR of the inter-B1 clone library. The PCR products were robotically spotted onto nylon membranes in a high-density array, with 10,000 clones in duplicate per spotted membrane. Hybridization conditions were as above.

RESULTS

To assess the frequency of polymorphism between C57BL/6 (B6) and *M. spretus* (LS) inter-B1 products, clones from two libraries (randomly generated and a subtracted library) were screened for size polymorphism and $\pm$ polymorphism (presence or absence of the inter-B1 product) by hybridization to Southern blots of B6 and LS inter-B1 PCR products (Fig. 1). Inter-B1 probes showing $\pm$ polymorphism were hybridized to filters containing spotted IRS PCR products of the backcross animals. These probes gave clear positive/negative segregation patterns across the full set (517) of animals (Fig. 2), allowing the scoring of backcross mice as B6 or LS for each probe. Allele scores were then entered into the EUCIB MBX database or Jackson Laboratory database (12) to obtain genetic map positions. Inter-B1 probes mapped to each chromosome in the genome, although there are regions with high inter-B1 probe density that differ for each probe library. Closely linked or unseparable loci were screened by probe-to-probe hybridization to eliminate duplicate scorings of identical clones. A consensus map of chromosome 1 IRS PCR probes is shown in Fig. 3. The genetic map positions for 470 IRS PCR loci can be accessed through the following databases: D_Leh probes are accessible in the Reference Library DataBase 2 (20) through World Wide Web (<http://gea.lif.icnet.uk/>). To access

FIG. 1. Cloned IRS PCR products are screened for polymorphism on Southern blots of both *Rsa* I-digested and undigested LS (EUCIB *M. spretus*) and B6 IRS PCR products. Each of these probes shows clear presence/absence of sequence ($\pm$) polymorphism. The probes used are as follows. (A) D7Leh1. (B) D1Leh3. (C) 2B4. (D) D9Leh2. (E) D6Leh3. The two bands positive with D9Leh2 represent a single locus. The bands cosegregate on Southern blots of 75 backcross animals.

FIG. 2. (A) Hybridization of a $\pm$ polymorphic IRS PCR product (D9Leh2) to the IRS PCR products of 423 mice from the EUCIB. The B6-derived probe hybridized to the mice backcrossed to LS (EUCIB *M. spretus*) appears as a positive signal in those mice with the B6 allele and negative in the mice containing the LS allele. The allele scorings for each probe are then entered into the EUCIB database to calculate the map location. (B) Segregation pattern of probe D9Leh2 on Southern blots of *Rsa* I-digested IRS PCR products of EUCIB mice. The parental strain controls are on the right of the upper panel. B6 is positive and both the London (LS) and Paris (PS) strains of *M. spretus* are negative. The segregation pattern of the probe across these animals matches exactly its segregation pattern across the same animals on a spotted backcross filter (A). Both methods gave the same genetic map location on chromosome 9.

this data in MBX at the Human Genome Mapping Project Resource Centre by email use helpdesk@hgmp.mrc.ac.uk and ask for a registration form, indicating that you wish to be an MBX user. D_Hun data can be accessed on The Jackson Laboratory mapping panel through World Wide Web via The Jackson Laboratory Home Page (<http://www.jax.org>), by email to lbr@aretha.jax.org, or in the Mouse Genome Database (MGD) with MGD accession numbers.[†]

Eleven IRS PCR product probes were mapped on Southern blots of IRS PCR products from the backcross mice as well as on gridded filters. Identical segregation patterns were obtained with both methods. The hybridization pattern for probe D9Leh2 on both Southern blots and gridded filters is shown in Fig. 2. The general utility of the IRS PCR mapping technique on interspecific and intraspecific crosses was determined by hybridizing a subset of the polymorphic probes to Southern blots of IRS PCR products from 34 mouse strains, hamsters, and rats. Many of the probes showed size polymorphism between related laboratory mouse strains. However, $\pm$ —

polymorphisms were most frequently observed between more divergent strains (Fig. 4).

A number of the IRS PCR probes have been sequenced to confirm the separate identity of each probe and, if necessary, to allow regeneration of the probes by PCR. To date, an unexpectedly large fraction of the probes contain potentially polymorphic simple repeat sequences, allowing where appropriate, the use of internal simple sequence length polymorphism markers as genetic mapping tools. The probes are available at the National Center for Biotechnology Information data repository, at the following ftp address: [ncbi.nlm.nih.gov](ftp://ncbi.nlm.nih.gov) (130.14.20.1) in `dir/repos/RLDB/IRS_probes`.

To demonstrate the ability to integrate the IRS PCR mapping technique with the construction of a physical genomic map, most of the inter-B1 probes were hybridized to high-density gridded filters of the ICRF and Whitehead YAC libraries, containing spotted IRS PCR products of each YAC clone (Fig. 5). Most probes identified 2–10 YAC clones uniquely. Hybridization of inter-B1 PCR products to the inter-B1 PCR products of YAC clones is extremely efficient and can produce good signals in an overnight exposure to autoradiography film.

DISCUSSION

IRS PCR products are a source of polymorphism that has only begun to be exploited in the past 5 years (13, 22–25). *M. spretus* diverged from *Mus musculus* approximately three million years ago (26) and is the most distantly related species with which the

[†]MGD accession numbers for the D_Hun Loci are as follows, by chromosome: 1, MGD-CREX-203; 2, MGD-CREX-204; 3, MGD-CREX-205; 4, MGD-CREX-206; 5, MGD-CREX-207; 6, MGD-CREX-208; 7, MGD-CREX-209; 8, MGD-CREX-210; 9, MGD-CREX-211; 10, MGD-CREX-212; 11, MGD-CREX-213; 12, MGD-CREX-214; 13, MGD-CREX-215; 14, MGD-CREX-216; 15, MGD-CREX-217; 16, MGD-CREX-218; 17, MGD-CREX-219; 18, MGD-CREX-220; 19, MGD-CREX-221; X, MGD-CREX-222.

FIG. 4. Segregation of $+/-$ probe, D0Leh50, across a panel of 34 mouse strains, a hamster, and a rat.

laboratory mouse is fertile. This has resulted in a high degree of polymorphism between the species because the position and sequence of repetitive elements in the genomes of both species vary, reflecting time of insertion and evolution of the elements

after insertion (27). As a result, IRS PCR size polymorphisms and $+/-$ polymorphisms are easily detected on Southern blots, making them more convenient to use as genetic markers than sequence polymorphisms. We have described (13) a genetic mapping method that typed probes on the IRS PCR products of interspecific backcross animals. We have further developed this method to type cloned IRS PCR products in a single hybridization on spotted backcross IRS PCR products.

The use of spotted backcross filters instead of gels should dramatically reduce the time and expense involved in genetic mapping, while increasing the number of animals screened per hybridization. In theory, up to 10,000 backcross animals could be screened on a single backcross filter. The spotted filters have the additional advantage that few manipulations are required in their preparation. The PCR products are robotically spotted directly onto nylon membranes. Such high-resolution genetic mapping from a single hybridization makes this a powerful method for rapidly increasing the probe density of the mouse genetic map, with minimal effort.

Potential limitations of this technique are the preferential cloning of short IRS PCR products and the preferential isolation of a subset of the possible products due to competitive amplification. To avoid the former limitation, the B6 IRS PCR products were size-fractionated and cloned. To address the latter limitation, individual probes were hybridized to spotted clone filters to determine the frequency of multiple isolations. In initial tests, D9Leh2 was found to represent 15% of the D_Leh probes screened, whereas a number of other probes

FIG. 3. Map of chromosome 1 showing the distribution of 40 IRS PCR probes. These genetic locations were obtained using the EUCIB MBX database or Jackson Laboratory (JAX) database. The loci in boldface type and underlined are the anchor markers used in the MBX or JAX database. This consensus map represents a best estimate of relative positions of D_Leh and D_Hun probes.

FIG. 5. Hybridization of probe D1Leh2 to ICRF mouse YAC library gridded filter. These filters contain the IRS PCR products (B1.2 primer) of each YAC in the library. All the genetically mapped inter-B1 sequences were hybridized to the YAC library.

were duplicated at a much lower frequency. Based on these results, we estimate that $\approx 75\%$ of the 40,000-clone library consists of independent isolates. Preliminary analysis of the LS-nonsubtracted clones indicates a $26\% \pm$ polymorphism rate, indicating that with an estimated 30,000 clones, we could theoretically identify ≈ 7500 genetic markers. This would map an IRS PCR probe to every ≈ 500 kb, if randomly spread across the genome.

More than 400 of the IRS PCR probes have identified YACs when screened on the ICRF mouse YAC library. The combined genetic and physical mapping of IRS PCR probes, therefore, represents a powerful tool for the mouse genome physical mapping effort. The IRS PCR product mapping technique complements the simple sequence length polymorphism-based mapping approach (8–11). Both methods exploit repetitive sequences in the mouse genome, but by concentrating on SINE and LINE elements rather than microsatellites, we have generated markers at sites not mapped by the simple sequence length polymorphism or other approaches. A combination of these markers with those already placed on the genome provides the marker saturation necessary for identifying genes involved in polygenic traits and results in a high-confidence physical map of the mouse genome.

L.M. thanks the EUCIB organization for making available the resources necessary for this work and Maria Kelly and Dilip Taylor in particular for assistance in the genetic analysis of the D₁Leh probes. K.H. thanks Mary Barter and Lucy Rowe for assistance in the genetic analysis and Lucy Rowe for useful discussions and critical review of this manuscript. W.N. acknowledges financial support under the European Community BioMed1 program. This work was supported in part by Public Health Service Grant 1-F32 GM-14788-01 to K.H. and National Institutes of Health Grant HG00299 and Program of Excellence Grant HL41484 to D.H.

1. Ton, C. C., Nirvonen, H., Miwa, H., Weil, M. M., Monaghan, P., Jordan, T., Van Heyningen, V., Hastie, N. D., Meijers-Heiboer, H., Drechsler, M., Royer-Pokora, B., Collins, F., Swaroop, A., Strong, L. C. & Saunders, G. F. (1991) *Cell* **67**, 1059–1074.
2. Fleischman, R. A., Saltman, D. L., Stastny, V. & Zneimer, S. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10885–10889.
3. Epstein, D. J., Vekemans, M. & Gros, P. (1991) *Cell* **67**, 767–774.
4. Tassebehji, M., Read, A. P., Newton, V. E., Harris, R., Balling, R., Gruss, P. & Strachman, T. (1992) *Nature (London)* **355**, 635–636.
5. Montagutelli, X., Lalouette, A., Coude, M., Kamoun, P., Forest, P. & Guenet, J. L. (1994) *Genomics* **19**, 9–11.
6. Janocha, S., Wolz, W., Sren, S., Srsnova, K., Montagutelli, X., Guenet, J. L., Grimm, T., Kress, W. & Muller, C. R. (1994) *Genomics* **19**, 5–8.
7. Herrman, B. G., Labeit, S., Poustka, A., King, T. R. & Lehrach, H. (1990) *Nature (London)* **343**, 617–622.
8. Love, J. M., Knight, A. M., McAleer, M. & Todd, J. (1990) *Nucleic Acids Res.* **18**, 4123–4130.
9. Cornall, R. J., Aitman, T. J., Hearne, C. M. & Todd, J. A. (1991) *Genomics* **10**, 874–881.
10. Hearne, C. M., McAleer, M. A., Love, J. M., Aitman, T. J., Cornall, R. J., Ghosh, S., Knight, A. M., Prins, J. B. & Todd, J. A. (1991) *Mamm. Genome* **4**, 273–282.
11. Dietrich, W., Katz, H., Lincoln, S. E., Shin, H.-S., Friedman, J., Dracopoli, N. C. & Lander, E. S. (1992) *Genetics* **131**, 423–447.
12. Rowe, L. B., Nadeau, J. H., Turner, R., Frankel, W. N., Letts, V. A., Eppig, J. T., Ko, M. S. H., Thurston, S. J. & Birkenmeier, E. H. (1994) *Mamm. Genome* **5**, 253–274.
13. Cox, R. D., Copeland, N., Jenkins, N. A. & Lehrach, H. (1991) *Genomics* **10**, 375–384.
14. Hunter, K. W., Ontiveros, S. D., Watson, M. L., Stanton, V. P., Gutierrez, P., Bhat, D., Rochelle, J., Graw, S., Ton, C., Schalling, M. & Hausman, D. (1994) *Mamm. Genome* **5**, 597–607.
15. Lebeau, M.-C., Alvarez-Bolado, G., Wahli, W. & Catsicas, S. (1991) *Nucleic Acids Res.* **19**, 4778.
16. Lysitsin, N., Lysitsin, N. & Wigler, M. (1993) *Science* **259**, 946–951.
17. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
18. Church, G. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **87**, 7757–7761.
19. European Backcross Collaborative Group (1994) *Human Mol. Genet.* **3**, 621–627.
20. Zehetner, G. & Lehrach, H. (1994) *Nature (London)* **367**, 489–491.
21. Manly, K. F. (1993) *Mamm. Genome* **4**, 303–313.
22. Cox, R. D. & Lehrach, H. (1991) *BioEssays* **13**, 193–198.
23. Simmler, M.-C., Cox, R. D. & Avner, P. (1991) *Genomics* **10**, 770–778.
24. Sinnet, D., Deragon, J.-M., Simard, L. R. & Labuda, D. (1990) *Genomics* **7**, 331–334.
25. Ledbetter, S. A., Nelson, D. L., Warren, S. T. & Ledbetter, D. H. (1990) *Genomics* **6**, 475–481.
26. Bonhomme, F. & Guenet, J.-L. (1989) in *Genetic Variants and Strains of the Laboratory Mouse*, eds. Lyon, M. F. & Searle, A. G. (Oxford Univ. Press, London), pp. 649–662.
27. Hastie, N. D. (1989) in *Genetic Variants and Strains of the Laboratory Mouse*, eds. Lyon, M. F. & Searle, A. G. (Oxford Univ. Press, London), pp. 559–573.