Polymorphisms generated by arbitrarily primed PCR in the mouse: application to strain identification and genetic mapping

John Welsh, Charles Petersen and Michael McClelland*
California Institute of Biological Research, 11099 North Torrey Pines Road, La Jolla, CA 92037, USA

Received October 22, 1990; Revised and Accepted December 11, 1990

ABSTRACT

Polymorphisms in genomic fingerprints generated by arbitrarily primed PCR (AP-PCR) can distinguish between strains of almost any organism. We applied the technique to the mouse (*Mus musculus*). The characteristic differences in the AP-PCR genomic fingerprints between strains will be of value in strain identification and verification. Using one primer, we genetically mapped four polymorphisms in a set of C57BL/6J x DBA/2J recombinant inbreds. One of these polymorphisms is a length variant. The method will allow rapid genetic mapping of DNA polymorphisms without Southern blotting.

INTRODUCTION

Simple and reproducible fingerprints of complex genomes can be generated using single arbitrarily chosen primers and the polymerase chain reaction (PCR) (1,2). An arbitrarily primed (AP-PCR) fingerprint is generated by subjecting a small amount of template DNA to PCR at relaxed stringency with any randomly selected PCR primer (with a few minor restrictions). Divergence of even a fraction of a percent between two genomes often results in a different fingerprint pattern because a somewhat different set of sites in the genome have the best matches with the primer. PCR products that are shared between only some individuals act as polymorphic markers, equivalent to other polymorphic characters used in phylogenetic and genetic mapping methods. Each primer gives a different pattern of AP-PCR products, each with the potential of detecting polymorphisms between strains. Thus, the data produced allows the differentiation of even closely related strains of the same species. These polymorphisms can be mapped genetically. We have applied the method to strain identification and genetic mapping in the mouse.

METHODS

Strains

DNA from *M. musculus domesticus* strains AKR/J, C3H/HeJ, C57BL/6J, C57L/J, DBA/2J, MEV/1Ty, SWR/J. and *M. musculus castaneus* strain CAST/Ei was prepared by Dr Benjamin Taylor at the Jackson Lab.

Primers

The primer KpnR (5'-CCAAGTCGACATGGCACRTGTATA-CATAYGTAAC), the 12-mer, Alu278, (5' GTAAGACTCTG) and the universal M13 sequencing primer, Uni, (5'-TTATGT-AAAACGACGGCCAGT), were designed for other purposes and chosen arbitrarily for these experiments. These primers do not match any sequences in the GENBANK database of mouse sequences. The oligonucleotides were obtained from Genosys, Houston, TX.

AP-PCR amplification

10 μl reactions were prepared using 0.025 units of Taq polymerase and 1 x Taq polymerase buffer (Stratagene) adjusted to 4 mM $MgCl_2$, 0.2 mM of each dNTP, 10 μM primer and DNA at various concentrations as indicated in the figure legends. A high Mg^{2+} concentration was selected to enhance the stability of primer/template interactions. The reaction was overlaid with oil and and subjected to two cycles through the following temperature profile: 94°C for 5 min. to denature, 40°C (for KpnR) and 48°C (for Uni) for 5 min. for low stringency annealing of primer and 72°C for 5 min. for extension. This temperature profile was generally followed by ten high stringency cycles: 94°C for 1 min., 60°C for 1 min. and 72°C for 2 min. At the end of this reaction, 90 μl of a solution containing 2.25 units of Taq polymerase, 1 x Taq buffer, 0.2 mM dNTPs and 5 μCi alpha-[^{32}P] dCTP was added and the high stringency cycles were continued for an additional 30 rounds. This protocol was designed to allow for high primer concentration during the low stringency steps (1).

For the 12-base primer the annealing temperature was kept at 40°C throughout.

RESULTS AND DISCUSSION

We investigated the number of polymorphisms that can be detected by AP-PCR in various inbred strains of mice. Two examples are presented in Figures 1 and 2. In Figure 1 the universal sequencing primer was used to generate a fingerprint. There was one polymorphism, Uni-230, that distinguishes some *M. mus. domesticus* strains from each other; lanes 1, 2, 4, 7,

* To whom correspondence should be addressed

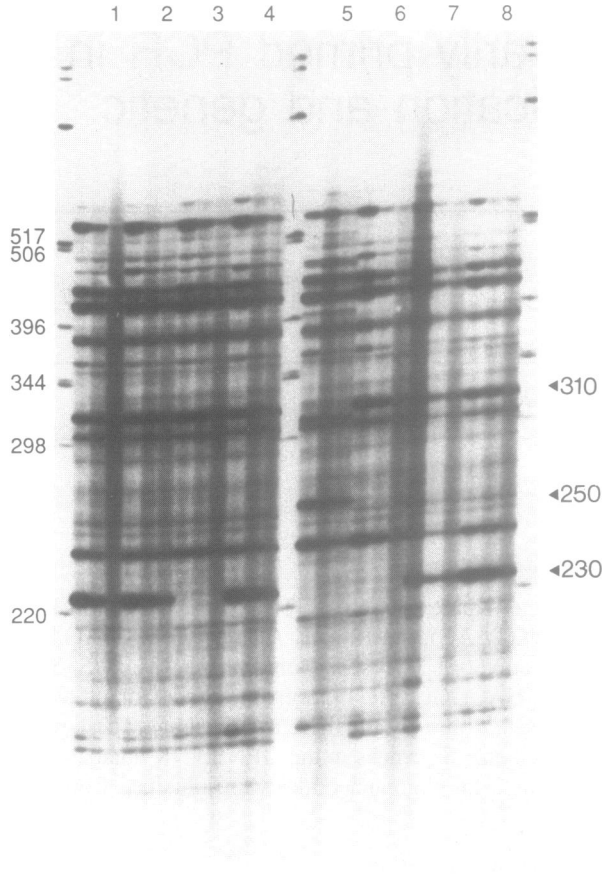

FIGURE 1. Mouse inbred lines show AP-PCR polymorphisms with the universal sequencing primer: AP-PCR was performed using three amounts of template, 250, 50 and 10 ng, under the standard AP-PCR conditions. The 20-mer M13 sequencing primer was used. The initial low temperature annealing step was 48°C. The resulting products were resolved by electrophoresis in 1X TBE through 5% Acrylamide-50% Urea and exposed by autoradiography at −70°C using a screen for 10 hours. Lanes 1, AKR/J. Lanes 2, C3H/HeJ. Lanes 3, C57BL/6J. Lanes 4, C57L/J. Lanes 5, CAST/Ei. Lanes 6, DBA/2J. Lanes 7, MEV/1Ty. Lanes 8, SWR/J. Size markers are from BRL/Gibco.

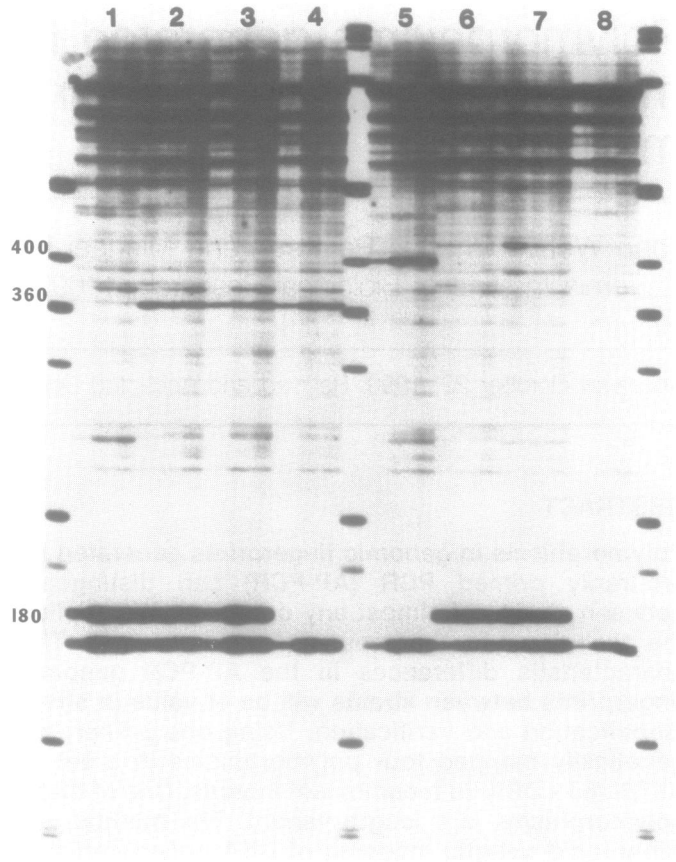

FIGURE 2. Mouse inbred lines show AP-PCR polymorphisms with a twelve base primer: Same as figure 1 except AP-PCR was performed using the 12-mer Alu278. PCR used the 40°C annealing temperature throughout.

and 8 are distinguished from lanes 3 and 6 by this polymorphism. Two polymorphisms, Uni-250 and Uni-310 distinguish *M. mus. castaneus* (lane 5) from *M. mus. domesticus*. We have named AP-PCR polymorphisms based on the primer name and the length of the polymorphism detected. In Figure 2 we used a 12 base primer. In this case there were two polymorphisms, Alu278-360 and Alu278-180 that distinguish some *M. mus. domesticus* strains from each other and a further polymorphism Alu278-400 in the *M. mus. castaneus*. Using these two primers we can distinguish all eight strains from each other, except the pair AKR/J and MEV/1Ty. This demonstrates the utility of the technique in strain identification.

There are only a few methods for the genetic mapping of DNA polymorphisms, including identifying restriction fragment length polymorphisms (RFLPs) by Southern blotting (3). One PCR method involves the use of specific primers to amplify across variable number tandem repeats (3a). This should prove very useful in genetic mapping. However, the VNTRs must be identified in clones by southern hybridization then sequenced and specific primers constructed.

As an alternative, one can take advantage of the heritable characteristics of the polymorphic bands produced by AP-PCR fingerprinting. Using fingerprinting of DNA from recombinant inbreds we have been able to determine map positions of such polymorphisms very quickly. For example, a primer that gave a pattern that distinguished the C57BL/6J from DBA/2J was applied to a series of recombinant inbreds derived from these strains. Recombinant inbred lines are derived by repeated sibling crosses starting with the F2 generation. This results in lines that are homozygous. The proportion of polymorphisms that cosegregated in the RI lines is an indication of the genetic linkage between the polymorphisms. For instance, two polymorphisms that are very closely linked genetically in a parental line will generally both be present or will both be absent in any particular RI line. As markers that are more distantly linked are examined, the chance of cosegregation diminishes because of recombination between chromosomes from the parentals during the breeding process. The example presented shows four mappable polymorphisms in one gel (Figure 3). Three of these were scored as presence or absence of a band and one was a length polymorphism. The pattern of inheritance can be compared to the pattern of inheritance for other markers that have already been mapped in this set of recombinant inbreds. We supplied our data to Dr. Benjamin Taylor at the Jackson Lab, Bar Harbor, Maine. He used a database containing all polymorphisms that had been mapped in this collection of RIs to determine the location of the KpnR polymorphisms on the genetic map. He found that the polymorphism KpnR-310 cosegregates with other polymorphisms that had previously been linked together on

chromosome 10, including a polymorphism at the locus *Mpmv-5* (3 crossovers in 21 recombinant inbreds). KpnR-185/175 maps near to an anonymous RFLP on chromosome 10 (1/21) and Xmv-31 (4/21). KpnR-115 maps to distal ch12 near the Igh-V immunoglobulin variable region gene cluster (2/21). Polymorphism KpnR-235 may map to ch 2 near *Psp*. However placement of this locus on either side of *Psp* requires double crossovers, so the assignment is not certain. KpnR-360 is not closely linked to previously mapped loci. About 10% of all polymorphisms fall in this latter class.

The polymorphisms were removed from the gel and PCR amplified to be made generally available. This was done by aligning radioactive ink spots on the gel with spots on the autoradiogram, cutting out the band, soaking in 50 μl of TE for a few hours, then amplifying by conventional PCR. Because this is a new technique, we blunt end ligated the amplified products into pBSKII$^+$ (Stratagene) and completely sequenced all the resulting clones. In principle, when two different arbitrarily selected primers are used to generate a polymorphic band, sequencing of some of these polymorphisms might be performed directly after amplification.

The KpnR primer, which contains long stretches of RY sequence (where R = G or A and Y = T or C), may have identified an $(RY)_n$ sequence polymorphism (7) in KpnR-235 with the sequence:

5′ ACATG.GCACRTGTATACATAYGTA.AC>ACACACATGCA.ACACACA 3′
primer > mouse

The length polymorphism KpnR-185/175 proved to be due to the formation or deletion of a perfect ten base pair repeat:

5′ primer ---15 bp --- $(TGTTACCTGA)_{1 \text{ or } 2}$ --- 150 bp --- primer 3′

Amplifications were checked by cleaving the DNA with restriction endonucleases with six base recognition specificity for which sites had been found in the clones. Greater than 95% of the amplified material was cleaved in all six cases, indicating substantial purity of the amplified material and that we had obtained the correct clones. We are currently using Southern blotting to determine whether polymorphisms are derived from single copy regions. To block contaminating PCR products in the amplified polymorphism probe we are experimenting with material amplified from the parent that does not carry the polymorphism. In the future the cloning or blotting steps would only be done as the probes become of interest in physical mapping or screening of libraries.

Most primers give at least one polymorphism when compared against a panel of *M. mus. domesticus* inbred lines. However, those primers that give the most polymorphisms will be particularly useful. It is likely that certain sequence motifs are more likely to be polymorphic, including the hypermutable CpG (4,5,6) and $(RY)_n$ minisatellite repeats (7). Of ten primers we have used so far, the KpnR primer in Figure 3 and two others (not shown) produce the most polymorphisms (three to six per gel) and have in common a long stretch of RY repeats and many CpG sequences or the CpG transition mutation products CpA and TpG. Thus, AP-PCR primers can be chosen for their greater potential to identify regions that are more polymorphic than the average. An alternative strategy we are also pursuing is to generate simple fingerprint patterns by PCR using consensus primers for *interspersed repetitive sequences (IRS)* (8,9,10,11, 13) may have associated variable number tandem repeat (VNTR) polymorphisms (14,15).

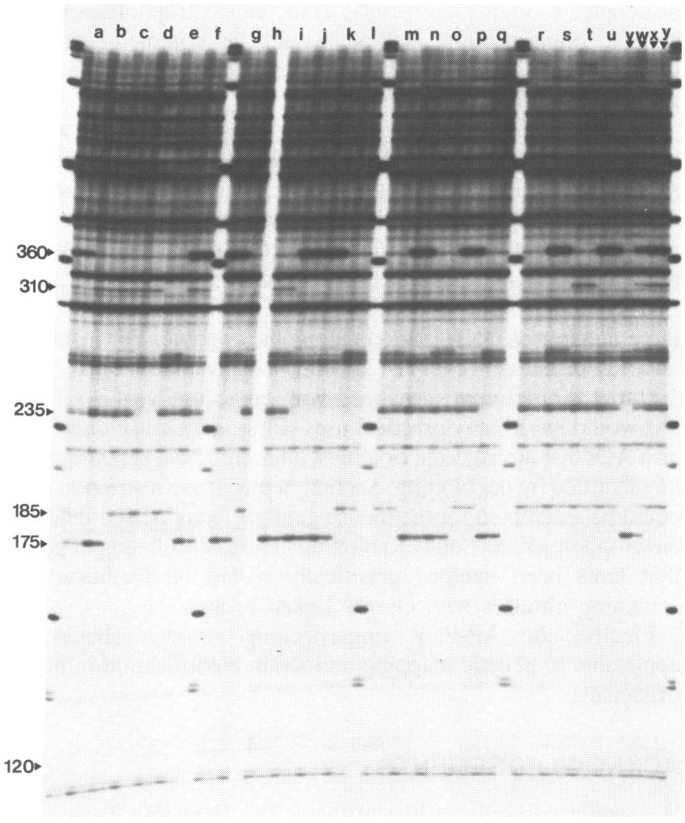

FIGURE 3. Segregating polymorphisms in mouse recombinant inbreds: Genomic DNA from males from the parentals, F1 hybrid and 26 recombinant inbred lines were investigated with the primer KpnR using the standard AP-PCR protocol and 40°C initial low stringency steps. The resulting products were resolved by electrophoresis in 1× TBE through 5% Acrylamide-50% Urea.

Most reactions were loaded in pairs, with two different amounts of template, 50 ng on the right and 25 ng on the left. Lanes **v** and **x** are patterns generated by AP-PCR of the genomes of the two parentals C57BL/6J and DBA/2J, lanes **a** and **w** are from the F1, lanes **b** to **u** and **y** are from 21 recombinant inbreds. Polymorphisms can be seen at 360, 310, 235, 185, 175 and 115 bases. The polymorphisms at 185 and 175 are mutually exclusive in the RIs and represents a length polymorphism at a single locus.

The fact that AP-PCR requires less than 1/100 the amount of genomic DNA per lane compared to that needed to prepare a Southern blot for conventional RFLP analysis is important when mapping in F2 crosses or backcrosses, for which a finite amount of DNA is available. In the mouse, interspecific backcrosses and F2 crosses are used in genetic mapping because they have a high level of polymorphism relative to most other inbred lines (eg 16,17). For example, one collection of (*M. mus. domesticus*×*M. spretus*)×*M. mus. domesticus* backcrosses has been scored for over 300 genetic markers (17).

AP-PCR fingerprinting is, in many respects, dramatically easier and faster than established methods for genetic mapping. We can begin a mapping program without having to first identify RFLP probes. No clones need be made and no plasmids purified. Polymorphisms can be generated by almost any primer we choose. Most of the steps in AP-PCR are automatable. The method can use ethidium detection, fluorescent detection or only tiny amounts of labelled bases relative to Southern hybridization. Furthermore, AP-PCR generated DNA polymorphisms can be isolated directly from gels and reamplified to use as probes in genome walking or restriction mapping strategies.

There are major efforts going into improving electrophoresis-based DNA separation technology. Many of these methods could

be adapted to PCR fingerprints. Thus, an electrophoresis-based genetic fingerprinting method would be complemented by improvements in electrophoresis technology, including data acquisition, storage and comparison.

A restriction map and an array of clones covering the whole mouse genome is highly desirable. However, even the most sophisticated clone mapping methods available will require a major effort to complete the map and may be hindered by potential limitations of the cloning vectors or vector hosts. For this reason pursuing a variety of approaches is advisable. AP-PCR could be used to produce a genetic map consisting of a very large number of DNA markers along the genome. For instance, an array of markers every 0.25 centiMorgans (about once every 100 to 1,000 kilobases) mapped with an accuracy of +/− 0.5 cM would have many practical uses. First, genomic clones such as YACs that are adjacent or almost adjacent in the genome could be identified by dot blotting. Second, a megabase restriction map could be assembled for the mouse genome using pulsed field gel electrophoresis and Southern blotting. Third, the cloning of genes that have been mapped genetically would be facilitated by screening libraries with closely linked probes.

Finally, the AP-PCR fingerprinting strategy should be applicable to genetic mapping and strain identification in many organisms.

ACKNOWLEDGEMENTS

We would particularly like to thank Dr. Benjamin Taylor for supplying DNAs, invaluable advice, interpreting the polymorphisms and critical reading of the manuscript. Without his assistance and exceptional generousity this work would not have been possible. We also thank Manuel Perucho and Waldemar Lernhardt for critical reading of the manuscript. This work was supported by grants to M.M. from the US Dept. of Energy and National Institutes of Health.

REFERENCES

1. Welsh J. and McClelland M. (1990) *Nucleic Acids Res.* **18** (in press).
2. Williams J.G.K., Kubelik A.R., Livak K.J., Rafalski J.A., Tingey S.V. (1990) *Nucleic Acids Res.* **18** (in press).
3. Botstein D., White R.L., Skolnick M. and Davis R.W. (1980) *Am. J. Hum. Genet.* 32:314−331.

3a. Jeffreys, A.J., Neumann, R. and Wilson, V. (1990) *Cell* **60**:473−485.

4. Coulondre C., Miller J.M., Farrabaugh P.J. and Gilbert W. (1978) *Nature* **274**:775.
5. M^{c}Clelland M. and Ivarie R. (1982). *Nucleic Acids Res.* 10:7865−7877.
6. Bird A.P. (1987) *Trends Genet.* **3**:342−47.
7. Litt M. and Luty J.A. (1989) *Am. J. Hum. Genet.* **44**:397−401.
8. Krayev A.S., Markusheva T.V., Ryskov A.P., Skryabin K.G., Bayev A.A. and Georgiev G.P. (1982) *Nucleic Acids Res.* **10**:7461−7474.
9. Kalb V.F., Glasser S., King D., and Lingrel J. (1983) *Nucleic Acids Res.* **11**:2177−2184.
10. Burton F.H., Loeb D.D., Voliva C.F., Martin S.L., Edgell M.H. and Hutchison III C.A. (1986) *J. Mol. Biol.* **187**:291−304.
11. Hwu H.R., Roberts J.W., Davidson E.H. and Britten R.J. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**:3875−3879.
12. Nelson D.L., Ledbetter S.A., Corbo L., Victoria M.F., Ramirez-Solis R., Webster T.D., Ledbetter D.H. and Caskey C.T. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**:6686−6690.
13. Ledbetter S.A., Nelson D.L., Warren S.T. and Ledbetter D.H. (1990) *Genomics* 6:475−481.
14. Economou E.P., Bergen A.W., Warren A.C. and Antonarakis S.E. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**:2952−2954.
15. Zuliani G. and Hobbs H.H. (1990) *Am. J. Hum. Genet.* **46**:963−969.
16. Buchberg A.M., Bedigian H.G., Taylor B.A., Brownell, Ihle J.N., Nagata S., Jenkins N.A., Copeland N.G., (1988) *Oncogene Res.* **2**:149−165.
17. Avner P., Amar L., Dandolo L, Guenet J.L. (1988) *Trends. Genet.* **4**:18−23.