



Use of vectorette and subvectorette PCR to isolate transgene flanking DNA.

M J Allen, A Collick and A J Jeffreys

Genome Res. 1994 4: 71-75

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

A horizontal banner advertisement with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." in white. In the center, there is a white-bordered box containing the words "LEARN MORE" in teal. On the right, there is a photograph of a woman wearing a red superhero mask and a red cape, and the Cellecta logo, which consists of a cluster of green dots.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Copyright © Cold Spring Harbor Laboratory Press

Use of Vectorette and Subvectorette PCR to Isolate Transgene Flanking DNA

Maxine J. Allen, Andrew Collick, and Alec J. Jeffreys

Department of Genetics, University of Leicester, Leicester LE1 7RH, UK

Vectorette PCR permits the specific amplification of DNA segments flanking a known DNA sequence. It enables the application of the PCR where sequence information is only available for one primer site. We now show that vectorette PCR can be used for the systematic mapping and retrieval of transgene flanking DNA. We also show that the sequence of large vectorette PCR fragments can be obtained without cloning, by the production of subvectorette fragments.

PCR allows specific amplification of DNA fragments between two regions of known sequence to which primers can be directed. Often, however, molecular biological techniques require the isolation and subsequent sequence analysis of unknown regions of DNA. PCR can be used to retrieve this unknown sequence if it is flanking a region of known sequence. In this case, a primer can be designed in either orientation from the known sequence. Such techniques as inverse PCR, panhandle PCR, capture PCR, and vectorette PCR have all been described as PCR methods for genome walking.⁽¹⁻⁴⁾

The basic vectorette PCR strategy has been described previously.⁽⁵⁾ Briefly, a vectorette-linker library is constructed from digested DNA for use in the PCR. The duplex vectorette linker consists of two annealed oligonucleotides of complementary sequence flanking a nonhomologous region resulting in a "bubble" in the DNA. The vectorette primer has an identical sequence to the nonhomologous region of the lower oligonucleotide and therefore cannot anneal directly to the linker. The specificity of the PCR is achieved by performing primer extension reactions from a sequence-specific primer to a vectorette linker in the adjacent unsequenced DNA. This results in the synthesis of a strand complementary to the vectorette linker, thus generating a target sequence for the vectorette-specific primer in subsequent PCR cycles. In this way, a specific DNA fragment can be amplified from a complex fragment mixture and recovered for use in sequencing reactions.

Previously, vectorette technology has been used to isolate terminal sequences from yeast artificial chromosome (YAC) inserts, in the determination of exon

structures and in the recovery of unknown bacterial sequences.⁽⁴⁻⁶⁾ In this paper we detail its use in the systematic isolation of 5'- and 3'-flanking junction sequences around mouse transgenes of the minisatellite MS32. MS32 is a human minisatellite composed of a tandem array of 29-bp repeat-unit sequence variants. The interspersions of these sequence variants can be assayed in the minisatellite variant repeat (MVR)-PCR⁽⁷⁾, which utilizes repeat-unit-specific primers. MS32 has a high mutation rate in humans, and knowledge of the flanking mouse sequences around transgenes of this locus is important in ascertaining any positional effects on transgene mutation rate and process. It was therefore necessary to isolate at least one flanking sequence junction around each of eight transgenes of the MS32 locus. Vectorette PCR was chosen as a swift method for isolating transgene flanking sequence products, obviating the need for cloning of the junction regions. The double-stranded products obtained were sequenced using PCR-based double-stranded sequencing methods. Most double-stranded PCR sequencing methods will only give 200-300 bp from each end of a fragment. Therefore, to access full sequence information for the original vectorette fragment, subvectoretting procedures were developed to provide a series of overlapping contigs for sequencing. We also show how the specificity of the vectorette library can be increased by the use of size-selected fragments and how partially digested DNA can be used to give a series of flanking products, aiding structural analysis. In this paper vectorette analysis of three MS32 transgenic loci, called 110D, 102B, and 110A, is described.

MATERIALS AND METHODS

Vectorette Production

The sequences of the vectorette top and bottom strands are detailed in Figure 1. Oligonucleotides were synthesized in the Department of Biochemistry, University of Leicester, on an Applied Biosystems 380B DNA synthesizer, using reagents supplied by Cruachem. Ten micrograms of vectorette top strand and 10 μ g of vectorette bottom strand were annealed at 60°C for 30 min in 20 μ l of 0.1 M NaCl, 10 mM MgCl₂, and 10 mM Tris-HCl (pH 7.6) to produce a complete vectorette linker. Neither vectorette oligonucleotide was 5' phosphorylated.

DNA Digestion, Preparation, and Ligation

Approximately 2 μ g of genomic DNA digested with a six-cutter restriction enzyme was ligated to 0.2 μ g of annealed vectorette linker in 20 μ l of 50 mM NaCl, 20 mM Tris-HCl (pH 7.6), 5 mM DTT, and 0.1 mM ATP, with 2 units of T4 ligase (GIBCO BRL) overnight at 15°C for sticky end ligations and 4°C for blunt-ended ligations. The amount of vectorette linker in the reaction was increased to 2 μ g for total genomic DNA ligations cleaved with a restriction endonuclease recognizing a 4-bp cleavage site. Control ligations without DNA were also prepared, and ligation was monitored by removal of a 3- μ l aliquot and addition of 0.5 μ g of λ DNA cleaved with *Hind*III and then agarose electrophoresis after ligation. Ligations ("vectorette libraries") were diluted with 5 mM Tris-HCl (pH 7.6) to 5 μ g/ml of genomic DNA and stored at -20°C.

Vectorette PCR Amplification

Hot start PCR was performed to prevent the annealing of the top- and bottom-strand vectorette oligonucleotides and any subsequent template-directed extension from the free 5' end of the top vectorette strand, which would not have ligated to the 3' end of the template top strand. This obviated the need for 5' phosphorylation of the top strand of the vectorette oligonucleotide. Ten nanograms of vectorette genomic DNA was diluted to 10 μ l with water, denatured at 96°C for 3 min, and then held at 80°C. Ten microliters of 2 \times concentrated PCR buffer, primers, and *Taq* polymerase (Perkin-Elmer Cetus AmpliTaq) were added, to give a final concentration of 1 μ M transgene primer, 1 μ M vectorette primer, and 0.06 U/ μ l of *Taq* polymerase. PCR buffer was as described previously.⁽⁷⁾ Amplification was at 96°C for 1.2 min, 68°C for 1 min, 70°C for 3 min for 30 cycles on a Perkin-Elmer Cetus thermal cycler. Vectorette PCR products were detected by agarose gel electrophoresis followed by Southern blot hybridization or ethidium bromide staining and visualisation by long-wavelength UV. Appropriate DNA fragments were recovered by electroelution onto dialysis membrane and reamplified with vectorette and transgene primers for 25–30 cycles depending on yield. PCR products were repurified by agarose gel electrophoresis and electroelution followed by ethanol precipitation.

Sequencing

Fifty to 150 ng of double-stranded PCR product was sequenced by the method of

Winship⁽⁸⁾ using either the vectorette or transgene primer.

Subvectoring of Vectorette PCR Products

The full sequence of vectorette PCR products >500 bp in length was obtained by the production of subvectoring fragments. Full or partial digestions with enzymes cleaving within the initial vectorette product were performed on 200–300 ng of initial product. The digested DNA was phenol-extracted and recovered by ethanol precipitation. Twenty nanograms of blunt-ended vectorette linker was annealed to 10- to 15-ng DNA fragments in a 20- μ l reaction using the previous ligation conditions. Reamplifications were performed on 1 μ l of a 1:20 dilution of this subvectoring library for 25 cycles using the second vectorette primer plus either the transgene primer or the first vectorette primer. PCR products were repurified by agarose gel electrophoresis and electroelution followed by ethanol precipitation.

RESULTS

The structure of the MS32 allele used in transgenesis is detailed in Fig. 2. It has 71 repeat units of **a**- or **t**-type, each repeat being 29 bp long. Standard restriction digestion and Southern blot analysis were performed on mice positive for transgene insertion. This identified appropriate restriction sites around the transgene for use in vectorette PCR to recover mouse flanking DNA sequences.

Transgene 110D

Restriction mapping showed this to be a single-copy insertion of the transgene construct with the loss of 10 repeat units but with full 5' and 3' MS32 flanking sequences. There was a lack of common four- and six-cutter restriction sites in the flanking mouse DNA, but the enzyme *Ava*II cleaved frequently and was utilized in the preparation of a vectorette library. A size fraction of *Ava*II-digested 110D positive mouse genomic DNA, known to contain the transgene fragment, was recovered by electroelution from agarose and used in the preparation of the vectorette library. Because *Ava*II cleaves a degenerate G^A/TCC site, a 50% mixture of GAC and GTC 5' overhanging *Vec*1 top sequences was used in

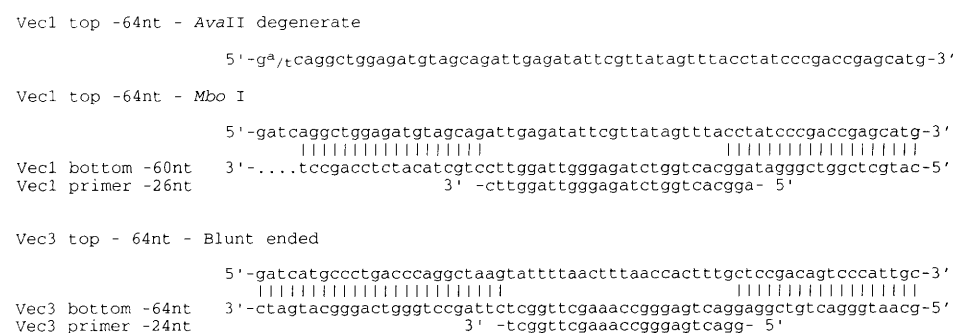


FIGURE 1 The oligonucleotides used in production of the vectorette units. One vectorette top strand and one vectorette bottom strand are annealed to produce a complete vectorette linker. Vectorette linkers for different restriction fragments are produced by varying the vectorette top strand sequence.

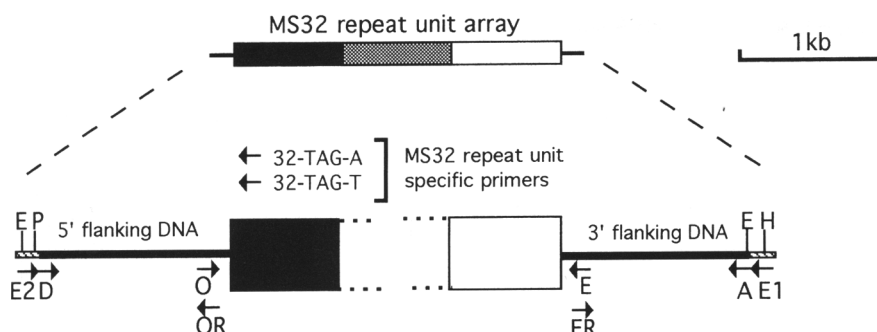


FIGURE 2 The structure of the MS32 transgene construct. The MS32 allele consists of 71 repeat units with 212 bp of 5' MS32-flanking DNA and 214 bp of 3'-flanking DNA. The total length of the construct is 2525 bp. The allele was PCR-amplified from human DNA using primers 32A and 32D with attached 20-bp synthetic linkers E1 and E2 (hatched). Diagnostic restriction sites in the linkers and flanking DNA are *Pst*I (P), *Hind*III (H), and *Eco*RI (E). The positions of the MS32 primers are indicated by arrows.

the preparation of the vectorette linker for this library. The first amplifications from this library were across the MS32 array, and products were verified by diagnostic restriction analysis and Southern hybridization (Fig. 3). Before reamplification of the product for se-

quencing, the size fraction containing the correct product was recovered by electroelution and amplified with a nested MS32 primer. This heminesting strategy was utilized to specifically recover the correct 5' and 3' products from the mixture of correct and Vec1/Vec1

products in the initial PCR. Sequencing the 5' and 3' junction fragments confirmed the unusual lack of restriction sites in the flanking DNA that was attributable to the γ satellite nature of the sequence.

Transgene 102B

Restriction mapping showed this single-copy transgene to have intact 3' MS32 human flanking DNA but no 5' flanking sequences. The mouse flanking DNA sequences again showed an unusual lack of most restriction sites other than for *Ava*II. *Ava*II was again chosen to construct a vectorette library using partially digested 102B genomic DNA. Vectorette PCR amplifications were used as a mapping tool to generate products extending out to a number of *Ava*II sites in the 5'- and 3'-flanking DNA (Fig. 4A,B). Amplifications with the 32E and Vec1 primers gave four bands, representing products extending to four flanking *Ava*II sites. E1 and Vec1 primer amplification only gave products extending to the first three of these *Ava*II sites, presumably owing to the favored amplification of shorter products. Mouse γ satellite DNA has a 234-bp repeating sequence.⁽⁹⁾ *Ava*II cleaves at a variant position in some γ satellite repeat units producing a 234- or 468-bp periodicity of digestion. This periodicity is seen in the flanking *Ava*II restriction map determined from vectorette PCR products, suggesting strongly that the flanking mouse DNA is gamma satellite.

As there was no 5' human MS32 flanking DNA, a heminested strategy for size reduction of the initial 5' vectorette PCR products (Vec1/E1 and Vec1/32E; Fig. 4B) was developed using MVR-PCR. For this reaction the Vec1 primer was used as the flanking primer, and the repeat unit-specific primers 32-TAG-A or 32-TAG-T were used as the transgene primers to produce a series of products extending from the closest *Ava*II site in the 5'-flanking mouse DNA to each **a**- or **t**-type repeat unit in the transgene MS32 array (Fig. 4C). The first repeat unit in the array is an **a**-type repeat, and the appropriate band from the MVR-PCR (indicated by an arrow in Fig. 4C) was electroeluted and amplified for sequencing, which confirmed the γ satellite sequence of the mouse DNA.

The 3'-flanking mouse DNA was recovered by vectorette PCR using primer

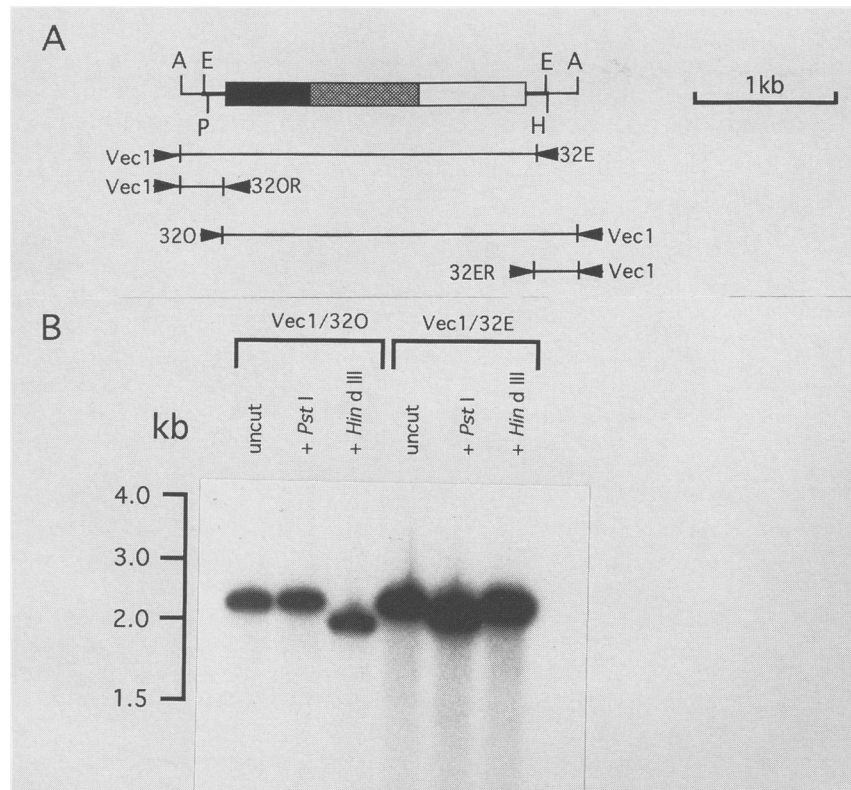


FIGURE 3 The structure of the 110D transgene and associated vectorette products. (A) (A) *Ava*II digestion site to which *Ava*II degenerate linkers were attached; (E) *Eco*RI; (P) *Pst*I; (H) *Hind*III. (B) Southern blot of the initial vectorette PCR products, Vec1/32E and Vec1/32O, probed with MS32 repeat-unit probe.

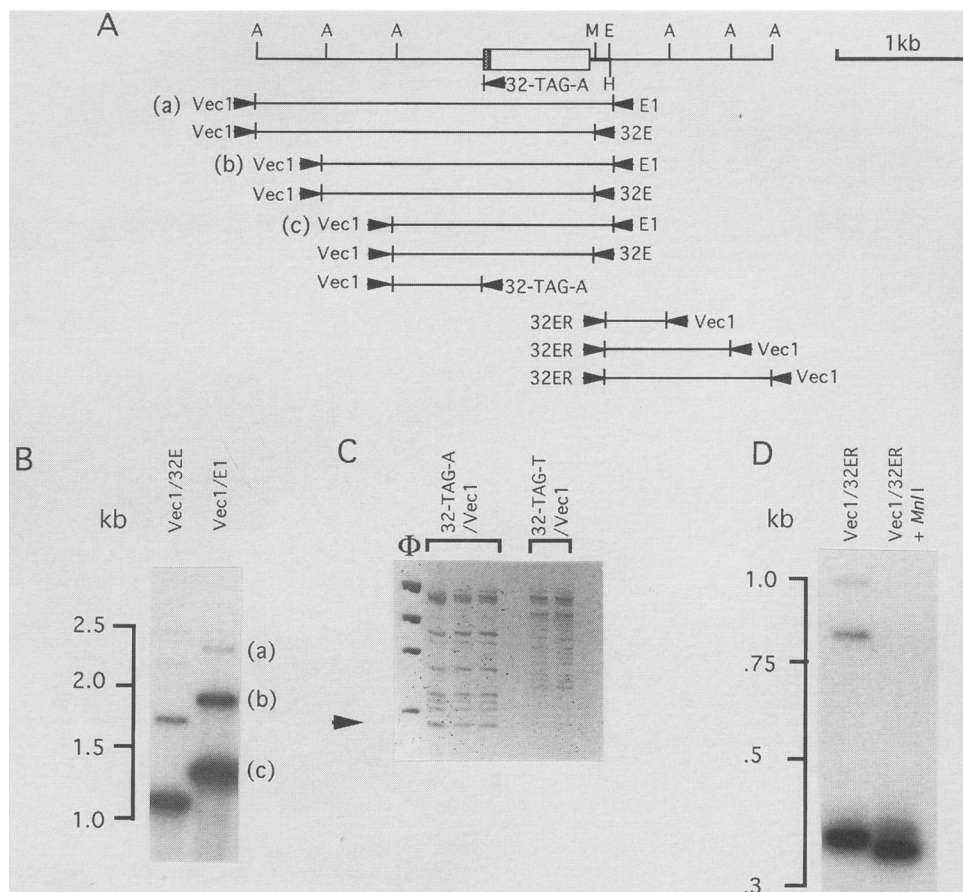


FIGURE 4 The structure of the 102B transgene and associated vectorette PCR products. (A) (E) *EcoRI*; (P) *PstI*; (M) *MnlI*; (A) *AvaII*. *AvaII* vectorette linkers were attached to DNA partially digested with *AvaII*, to produce vectorette PCR products extending to *AvaII* sites in both the 5'- and 3'-mouse flanking DNA (a, b, and c represent the Vec1/E1 products detailed in B). Primer 32-TAG-A is MS32 repeat-unit-specific. (B) Vectorette products extending from the 5' junction of the transgene. Using initial primer combinations Vec1/32E and Vec1/E1, there are three products extending out to three *AvaII* sites in the flanking DNA. There is also a product extending to a fourth *AvaII* site detectable in the Vec1/32E amplification (not detailed in A). (C) The MVR-PCR pattern visible by ethidium bromide staining when the electroeluted band from the Vec1/32E reaction was amplified with Vec1/32-TAG-A (triplicate reactions). Duplicate amplifications on the same electroeluted product using the 32-TAG-T repeat-unit-specific primer were also performed to generate the corresponding set of MVR-PCR products extending to each *t*-type repeat unit. (F) ϕ X174 cleaved with *HaeIII*. An arrow indicates the smallest product from the Vec1/32-TAG-A reaction that was electroeluted and used in sequencing reactions. (D) Vectorette products extending 3' of the transgene insertion. A Southern blot of amplification products with Vec1/32ER was probed with an MS32 3'-flanking probe (141 bp of the 3'-flanking DNA). Three vectorette products were amplified extending to *AvaII* sites in the 3'-flanking mouse DNA. The smallest product is cleaved by *MnlI*, which cuts in the MS32 flanking human DNA.

32ER in the human 3'-flanking DNA (Fig. 4D). Junction products were verified by Southern blot hybridization with a human 3'-flanking sequence probe and digestion with *MnlI*, which has a diagnostic site in the 3'-flanking DNA. A heminested strategy using the reverse primer of 32A (Fig. 1) with Vec1 was used to recover this fragment for sequencing, which showed that the mouse sequence was γ satellite.

Transgene 110A

Restriction mapping identified this transgene as a complex multicopy array of the original construct. The enzyme *MboI* does not cleave in the original construct but cleaves in the mouse DNA 1.3 kb upstream of the 5' end of the multicopy array. A complete *MboI* digest was performed on the genomic DNA, and *MboI* vectorettes were ligated. The initial

amplification using primer 32OR directed toward the 5'-flanking mouse DNA produced a 1.34-kb product detectable by ethidium bromide staining following agarose gel electrophoresis. This was too long for full double-stranded PCR sequencing. Restriction mapping of this product revealed a number of blunt-end restriction sites to which a universal blunt-end novel sequence vectorette (Vec3) could be annealed (Fig. 5). The 1.34-kb DNA fragment was either fully digested with *AluI* or *HincII* or partially digested with *RsaI* (because there were two sites to access with Vec3) and phenol-extracted and ethanol-precipitated, before the novel Vec3 was ligated. These subvectorette libraries were reamplified using 32OR and Vec3, and the products were recovered for sequencing. Sequence analysis of the original vectorette products and the four subvectorette products produced a complete sequence contig of the 5'-mouse flanking DNA, which was shown to be a unique sequence.

DISCUSSION

In this paper we discuss the use of vectorette PCR as a method of systematically isolating mouse DNA adjacent to transgene insertions as a prelude to DNA sequence analysis and show that vectorettes provide a highly flexible approach to the mapping and recovery of sequences from complex genomes. In this study we were able to isolate and sequence at least one junction from each of eight different transgenic inserts, without the need for cloning via genomic DNA libraries. Junction fragments up to 3.7 kb in length were successfully recovered by vectorette PCR.

The ability to recover specific vectorette PCR products depends on the complexity of genomic DNA, the frequency of restriction sites in the DNA used to prepare the vectorette libraries, and the specificity of vectorette and transgene primers. The latter was optimized by designing primers with a relatively high GC content (54–58%), which permitted the use of a high annealing temperature (67°C) during PCR. In some cases, specificity was sufficiently high to enable the correct vectorette PCR products to be identified after the first round of PCR by agarose gel electrophoresis and staining with ethidium bromide. In other cases, particularly if the vectorette library was

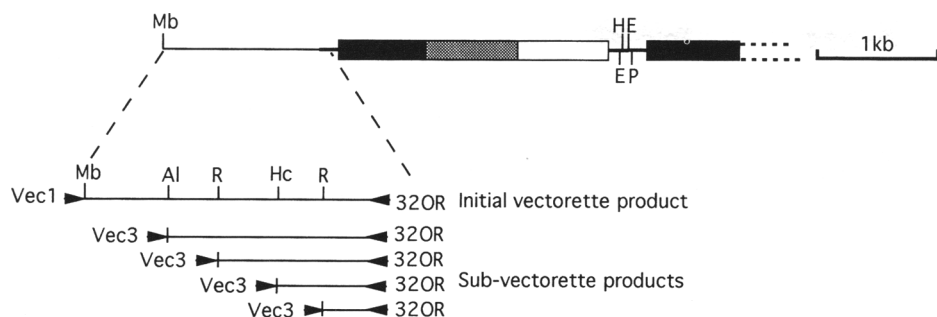


FIGURE 5 The structure of the 5' region of the multicopy transminisatellite 110A and associated vectorette and subvectorette PCR products. The initial Vec1 product used an *Mbo*I vectorette linker and the subvectorette was produced using a blunt-ended Vec3 linker ligated to sites for *Alu*I (Al), *Rsa*I (R), and *Hinc*II (Hc).

constructed from genomic DNA cleaved with frequently cutting restriction enzymes (e.g., *Mbo*I), relatively complex sets of PCR products were frequently obtained. Two approaches were used to circumvent this problem. First was knowledge of the size of the restriction fragment containing the transgene enabled restricted genomic DNA to be size-fractionated by agarose gel electrophoresis, followed by the recovery of a size fraction containing the transgene fragment. This type of fractionation was performed in the recovery of transgene 110D flanking DNA to reduce the complexity of the vectorette library 10- to 100-fold and to substantially improve the specificity of the vectorette PCR. The second approach was to perform a first round of vectorette PCR amplification, then to detect products by Southern blot hybridization, recover by gel fractionation, and reamplify using a second nested transgene primer; this heminesting strategy can also simplify the profile of vectorette PCR products.

The use of partially digested genomic DNA in vectorette libraries also provides a general method of mapping restriction sites adjacent to a transgene, as shown for transgene 102B. Such mapping data does not only provide clues about the nature of the mouse flanking DNA, but it can also generate primary vectorette products extending for any required distance into the flanking DNA, without the need for detailed restriction mapping prior to vectorette PCR. Similarly, vectorette PCR can also be used to explore the structure of the transgene itself. For example, it is possible to perform MVR-PCR directly on primary vectorette libraries (data not shown) to create PCR products extending from a

flanking restriction site to each **a-** or **t-**type repeat unit within the minisatellite array. This makes it possible to determine transgene structures in the absence of suitable flanking human DNA and gives direct information on the distance between the flanking mouse restriction site and the start of the MS32 repeat array.

Vectorette PCR can also be used as a tool to aid sequence analysis of large PCR products by subvectoretting. Our general strategy is to partially digest large PCR products with frequently cutting restriction enzymes that yield blunt-end DNA fragments, followed by ligation of a second vectorette and specific amplification of a set of DNA fragments extending from one of the original primers (the first vectorette primer or a transgene primer) to each of the blunt-end sites capped with the second vectorette. In this way, sets of nested deletions can be rapidly generated and used for sequence contig construction. We would note, however, that in our experience, blunt-end vectorettes work very inefficiently in PCR amplifications from primary genomic vectorette libraries and do not provide a useful means for isolating transgene/flanking DNA junctions from genomic DNA.

ACKNOWLEDGMENTS

We thank Debra Langton for oligonucleotide synthesis and Annette Macleod for helpful discussions on the vectorette PCR technique. Production of the transgenic mice was performed by Michael Norris and Sheila Barton at the Agricultural Food and Research Council Institute (Babraham, Cambridge, UK). The work of A.J.J. was supported in part by an

International Research Scholars award from the Howard Hughes Medical Institute and in part by grants from the Medical Research Council and the Royal Society.

REFERENCES

1. Triglia, T., M.G. Peterson, and D.J. Kemp. 1988. A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequence. *Nucleic Acids Res.* **16**: 8186.
2. Jones, D.H. and S.C. Winisforfer. 1992. Sequence specific generation of a DNA panhandle permits PCR amplification of unknown flanking DNA. *Nucleic Acids Res.* **20**: 595-600.
3. Lagerstrom, M., J. Parik, H. Malmgren, J. Stewart, U. Peterson, and U. Landegren. 1991. Capture PCR: Efficient amplification of DNA fragments adjacent to a known sequence in human and YAC DNA. *PCR Methods Applic.* **1**: 111-119.
4. Riley, J., R. Butler, D. Ogilvie, R. Finniar, D. Jenner, S. Powell, R. Anand, J.C. Smith, and A.F. Markham. 1990. A novel method for the isolation of terminal sequence from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.* **18**: 2887-2890.
5. Arnold, C. and I.J. Hodgson. 1991. Vectorette PCR: A novel approach to genome walking. *PCR Methods Applic.* **1**: 39-42.
6. Roberts, G.R., A.J. Coffrey, M. Bobrow, and D. Bentley. 1992. Determination of the exon structure of the distal portion of the Dystrophin gene by vectorette PCR. *Genomics* **13**: 942-950.
7. Jeffreys, A.J., A. Macleod, K. Tamaki, D.L. Neil, and D.G. Monkton. 1991. Minisatellite repeat coding as a digital approach to DNA typing. *Nature* **354**: 204-209.
8. Winship, P.R. 1989. An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. *Nucleic Acids Res.* **17**: 1266.
9. Hastie, N.D. 1989. Highly repeated DNA families in the genome of *Mus musculus*. In *Genetic variants and strains of the laboratory mouse*, 2nd ed. (ed. M.F. Lyon and A.G. Searle), pp. 559-574. Oxford University Press, Oxford, UK.

Received May 17, 1994; accepted in revised form July 12, 1994.