LaNe RAGE: a new tool for genomic DNA flanking sequence determination

Daniel J. Park

Department of Zoology The University of Melbourne Gate 12, Royal Parade Victoria 3010, Australia Tel: 61 3 83444351 Fax: 61 3 83447909 E-mail: d.park@unimelb.edu.au

Financial support: This work was supported by the National Health and Medical Research Council of Australia grant no. 148630 to Professors J.A.M. Graves and M.B. Renfree, Associate Professor V.R. Harley and Dr. A.J. Pask.

Keywords: cloning, DNA fingerprinting, genomic DNA walking, PCR, sequencing, transposon mutagenesis.

Abbreviations: GAPDH: glyceraldehyde-3-phosphate dehydrogenase;

gDNA: genomic DNA;

LaNe RAGE: lariat-dependent nested PCR for rapid amplification of genomic DNA ends;

PCR: polymerase chain reaction;

PGK1: phosphoglycerate kinase 1;

RAPD: randomly amplified polymorphic DNA.

The determination of genomic DNA sequence flanking a known region is often problematic. Existing technologies depend on multiple, efficient enzymecatalysed preparative processing steps and/or rely on relatively inefficient 'one-sided' PCR mechanisms. I demonstrate the application of a simple 'two-sided' PCR-based approach, lariat-dependent nested PCR for rapid amplification of genomic DNA ends (LaNe RAGE), applied to the mouse GAPDH and PGK1 gene flanking sequences. This demonstration offers great promise in applications such as genome walking, transposon mutagenesis mapping and DNA fingerprinting.

Current methods for PCR-based genomic DNA (gDNA) flanking sequence determination variously require relatively large amounts of starting material, efficient enzyme reaction steps reliant on restriction endonuclease, ligase and/or exonuclease or dependence on relatively inefficient 'one-sided' PCR reactions (Mueller and Wold, 1989; Arnold and Hodgson, 1991; Sarkar et al. 1993; Devon et al. 1995; Myrick and Gelbart, 2002). This study set out to demonstrate the feasibility of a new technology, LaNe RAGE, requiring only limited starting material without the need for involved pre-processing steps and capable of unleashing the exquisite sensitivity of fully gene-specific 'two-sided' nested PCR series (Figure 1).

MATERIALS AND METHODS

gDNA extraction

gDNA was extracted from the mouse HC11 mammary

gland cell line using the DNeasy Tissue Kit (Qiagen, Clifton Hill, Vic., Australia).

LaNe RAGE

Primers were designed based on the GenBank accession sequences NT 039711.3 (BX469914.4) and NT 039206.3 (AL732526.8 reverse complement view) for PGK1 and GAPDH, respectively. Locations within these regions used for primer design are indicated after each primer sequence. 300 ng gDNA was used as template in 50 µl PfuTurbo® DNA Polymerase (Stratagene, Integrated Sciences, East Kew, Vic., Australia) -catalysed strand synthesis reactions primed by hybrid 5' gene-specific and 3' partially defined PGKN5CTCAC primers (5'-GAGAATGCCAAGACTGGCCAAGCNNNNNCTCAC-(5'-26744-26766) GAPN5CTCAC 3': or TGGTCTACATGTTCCAGTATGACTCNNNNNCTCAC-3': 3074-3098) for mouse PGK1 or GAPDH, respectively. The semi-random 3' termini were chosen with the expectation that under the appropriate annealing conditions, they would bind at relatively frequent intervals adjacent to a known region. As such, a ladder of gene-specific products would be anticipated. The reaction mix tubes were heated to 94°C for 3 min and then placed on ice prior to addition of 1.5 µl 10 mM dNTPs and PfuTurbo® DNA Polymerase. The reaction tubes were then heated from 18°C to 72°C at a rate of 0.1°C per 5 sec and held at 72°C for 10 min prior to placing on ice. Primers PGK1 (5'-TTGGCACATCTCTGTATGATGAAG-3': 26622-26645) (5'-TACTGTGGCCTCTGGTATACCT-3': and PGK4 26767-26788) GAP1 (5'or ATGGTGAAGGTCGGTGTGAACG-3': 2953-2974) and GAP4 (5'-CCGAGAATGGGAAGCTTGTCATC-3': 3131-3153), respectively, were added prior to thermocycling (36

cycles of 94°C for 1 min; 58°C for 1 min; 72°C for 2 min). The lariat structure formation and template generation depicted in the large boxed section of Figure 1 occurs inherently during the thermal cycling of LaNe RAGE.

1 µl product from each first round PCR was included as template in 50 µl hot-start second round PCRs employing PGK2 primers (5'-GGAGCCAAGATTGTCAAAGATCTC-3': 26648-26671) PGK5 (5'-CTGTTCGCTTATGGAAGGGATC-3': and 26814-26835) GAP2 (5'or CAGTGGCAAAGTGGAGATTGTTGC-3': 3018-3041) and GAP5 (5'-GGAGCGAGACCCCACTAAC-3': 3180-3198), respectively, and PfuTurbo® DNA Polymerase (40 cycles of 94°C for 1 min; 60°C for 1 min; 72°C for 2 min). Products were visualized by UV trans-illumination after agarose gel electrophoresis.

DNA sequencing

Major PCR products were gel extracted using the QIAEX®II bulk silica DNA purification system (Qiagen) and subjected to sequencing employing ABI PRISM®BigDye[™]Terminators v.3.1 (Applied Biosystems, Scoresby, Vic., Australia) primed by PGK2, PGK5, GAP2 or GAP5. Electrophoresis was performed by the Wellcome Trust Sequencing Facility, a facility of Monash Institute of Reproduction and Development and Prince Henry's Institute of Medical Research, Monash Medical Centre, Clayton, Vic., Australia.

RESULTS AND DISCUSSION

Figure 2 illustrates the agarose gel profiles resulting from application of LaNe RAGE to the mouse GAPDH and PGK1 genes. The 280 bp and 580 bp products for GAPDH and PGK1, respectively, as the major products, were chosen for DNA sequence analysis (Figure 3 and Figure 4). Nucleotide-nucleotide BLAST searching and Clustal W alignment analysis, indicated that they represented GAPDH and PGK1 flanking regions, respectively (Figure 5 and Figure 6). Furthermore, the presence of gene specific hybrid primer-derived sequence confirmed that the products had been vielded via a LaNe RAGE mechanism. Major bands were consistently observed in independent experiments and were confirmed to be gene-specific by DNA sequence analysis. The GAPDH profile exhibits an expected ladder of products while that for PGK1 shows a major product and relatively few minor distinct bands. It is possible that the gene-specific region of the PGKN5CTCAC primer contributed to enhance binding at a particular site, such that the resulting major product outcompeted others in the ensuing PCRs. The background smear observed in the GAPDH product could be due to primer binding at more regular intervals than anticipated (for example, TCAC or CAC sites). Altered primer design and annealing conditions change product profiles to suit the

requirements of the application. For example, the design of 'less random' primers for use in the initial annealing step should result in a ladder with longer products and bigger 'steps'. Nonetheless, the described experiments effectively serve to demonstrate the potential of LaNe RAGE.

This demonstration of a new simple and effective tool for derivation of gDNA flanking sequence requiring only limited starting material has implications not only for genome walking, but also for identification of insertion sites resulting from transposon mutagenesis screens and for DNA fingerprinting applied in contexts such as pathogenic bacteria strain typing. These applications should simply require optimization of primer design and thermocycling conditions. LaNe RAGE should offer greater specificity and sensitivity than existing approaches via a quick, simple and robust mechanism. At the heart of this is the capacity for 'two-sided' gene-specific nested PCR series without the requirement for involved enzyme pre-processing steps.

REFERENCES

ARNOLD, C. and HODGSON, I.J. Vectorette PCR: a novel approach to genomic walking. *PCR Methods and Applications*, 1991, vol. 1, no. 1, p. 39-42.

DEVON, R.S.; PORTEOUS, D.J. and BROOKES, A.J. Splinkerettes – improved vectorettes for greater efficiency in PCR walking. *Nucleic Acids Research*, May 1995, vol. 23, no. 9, p. 1644-1645.

MUELLER, Paul R. and WOLD, Barbara. *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. *Science*, November 1989, vol. 246, no. 4931, p. 780-786.

MYRICK, Kyl V. and GELBART, William M. Universal fast walking for direct and versatile determination of flanking sequence. *Gene*, February 2002, vol. 284, p. 125-131.

SARKAR, G.; TURNER, R.T. and BOLANDER, M.E. Restriction site PCR: a direct method of unknown sequence retrieval adjacent to a known locus by using universal primers. *PCR Methods and Applications*, 1993, vol. 2, no. 4, p. 318-322.

APPENDIX

Figures



Figure 1. Schematic representation of LaNe RAGE. Dashed lines represent gDNA. Dotted lines represent DNA polymerization. Solid lines represent DNA formed during the previous step. 3' indicates the 3' end of DNA strands. The small box represents a known region of gDNA. The large box indicates intrinsic steps which take place during the first PCR of the LaNe RAGE procedure. Numbers 1 to 5 indicate gene-specific primer binding sites within a known region of gDNA. Underlined numbers indicate reverse complementary sequence relative to the initial gDNA strand. Number 6 indicates an optional gene-specific primer binding site for direct sequencing of product. Arrows indicate sites of primer binding. NNN3 labels the special hybrid primer used to prime initial DNA synthesis, consisting of a 5' gene-specific sequence (corresponding to position number 3) adjacent to 3' terminal degenerate, part-defined or defined sequence.

Initially, the NNN3 hybrid primer binds in a semi-random fashion at regular intervals along genomic DNA to prime DNA synthesis. During thermal cycling, gene-specific primer 1 binds at a specific site within the known region of the product of NNN3-primed extension to prime another extension reaction. This results in a product with gene-specific sequence at its 5' and 3' (corresponding to gene-specific sequence introduced by the hybrid primer) ends, with unknown sequence between. During thermal cycling, this is able to form a lariat structure, which primes intra-strand extension. The resultant product possesses 5' gene-specific sequence and extended 3' gene-specific sequence with unknown sequence between, which is amenable to two-sided gene-specific PCR series.



Figure 2. Separation on 1.4% agarose 0.5xTBE agarose gel of products yielded from LaNe RAGE applied to mouse GAPDH and PGK1 genes. 100 indicates Promega 100 bp ladder (Promega Corporation, Annandale, NSW, Australia). Arrows indicate discrete product bands. The major products (280 bp and 580 bp for GAPDH and PGK1, respectively) were selected for sequence analysis.



Figure 3. Chromas generated chromatogram depicting sequence for the LaNe RAGE-derived 280 bp major *GAPDH* product. The sequencing reaction was primed using GAP2. Dashed lines represent 'unknown' sequence. Boxed regions represent sequence corresponding to primers. The product is consistent with the mechanism shown in Figure 1.



Figure 4. Chromas generated chromatogram depicting sequence for the LaNe RAGE-derived 580 bp major *PGK1* product. The sequencing reaction was primed using PGK2. Dashed lines represent 'unknown' sequence. Boxed regions represent sequence corresponding to primers. The product is consistent with the mechanism shown in Figure 1.

G280-2rc AL732526.8	ACAGTCAAGGCCGAGAATGGGAAGCTTGTCATCAACGGGAAGCCCATCACCATCTTCCAG	3180
G280-2rc AL732526.8	ATCAAATGGGGTGANGCCGGTGCTGAGTATGTCGTGGAGTCT GAGCGAGACCCCACTAACATCAAATGGGGTGAGGCCGGTGCTGAGTATGTCGTGGAGTCT ***********************************	42 3240
G280-2rc AL732526.8	ACTGGTGTCTTCACCACCATGGAGAAGGCCGGGGCCCACTTNAAGGGTGGAGCCAAAAGG ACTGGTGTCTTCACCACCATGGAGAAGGCCGGGGCCCACTTGAAGGGTGGAGCCAAACGG ***************************	102 3300
G280-2rc AL732526.8	GTCATCATCTCNNCCCCNTCTGCCGATGCCCCCNTGTTTNTGANGGGTGTCATCTCCGCCCCTTCTGCCGATGCCCCCATGTTTGTGATGGGTGTGAACCACGAG	150 3360
G280-2rc AL732526.8	AAATATGACAACTCACTCAAGATTGTCAGCAATGCATCCTGCACCACCAACTGCTTAGCC	3420

Figure 5. Clustal W generated alignment of the 'unknown region' of the GAPDH LaNe RAGE product (reverse complement of that depicted in Figure 3) with the reverse complement view of the corresponding GAPDH genomic segment AL732526.8. Numbers at the ends of the AL732526.8 lines indicate the base positions within the complete genomic segment.

P580-2rc BX469914.4	CCAAGCTACTGTGGCCTCTGGTATACCTGCTGGCTGGATGGTGAGTCAGCTGGCTG	26820
P580-2rc BX469914.4	ANAAATTTTNGGCATCATTCATTCATCCNTTCATTCATTTAGCAG CTTATGGAAGGGATCAGAAATTTTTGGCATCATTCATTCA	45 26880
P580-2rc BX469914.4	ATTTTTTTTGNGACATTGGNGATCCATNTGNGTGTTTCANAGATGTTAAGAAATGAGCTA ATTTTTTTTGTGACATTGGAGATCCATCTGTGTGTCTCATAGATGTTAAGAAATGAGCTA ********* ******* ******* ** **** ***	105 26940
P580-2rc BX469914.4	CCCCTGAGTTTTTCCCTCCAACCGTTTCCATGGGAACTGTCAACNCTGCAAATGATGCAG CCCCTGAGTTCTTCCCTCCAACCGTCTCCATGGGAACTGTCAACACTGCAAATGATGCAG ***********************************	165 27000
P580-2rc BX469914.4	NGNGGNCCGACCTGAGGGCAAATCCCTGGGGTAAAAGCAGNGNTGGCNCATTGTTGCTGT TGAGGACCGACCTGAGGGCAAATCCCTGGGGTAAAAGCAGTGCTGGCACATTGTTGCTGT * ** *******************************	225 27060
P580-2rc BX469914.4	TTTCATTAANAAGNGGGCAAACCCNCTTGATTTTTTTCAGNGNGNGTTTTAGCNCTATG CTTCATTAATAAGAGGGCAAACCCACTTGATTCTTTTTCAGTGTGTGT	285 27120
P580-2rc BX469914.4	TGCTACAGGGAGTACTANGNACTTTNTTTTTTTTTTTTTT	345 27180
P580-2rc BX469914.4	ATGTACTTTCATGGTTTGAACTGGTCTGGATTCCAGTGCTAATAGCTACAAGCCATTGTA ATGTACTTTCATGGTTTGAACTGGTCTGGATTCCAGTGCTAATAGCTACAAGCCATTGTA	405 27240
P580-2rc BX469914.4	AGCACACATTCAGGCCGAGCACAGTGGGTCAAAGTGCTTGCGACCCAAGCCTG	422 27300
P580-2rc BX469914.4	GCGGTCCTGCACTGGAAGGAAACAGTTCCAAAAATTGTCCTGACTTCTACATGACATGCG	27360

Figure 6. Clustal W generated alignment of the 'unknown region' of the *PGK1* LaNe RAGE product (reverse complement of that depicted in Figure 4) with the sense view of the corresponding *PGK1* genomic segment BX469914.4. Numbers at the ends of the BX469914.4 lines indicate the base positions within the complete genomic segment.