

TECHNICAL NOTE

A strategy for the rapid isolation of new PCR based DNA polymorphisms

Paul R Hoban, M F Santibanez-Koref, J Heighway

DNA sequence polymorphisms in the form of restriction fragment length polymorphisms (RFLPs) are powerful tools for studying linkage or loss of constitutional heterozygosity (LOCH). Most have been identified by and are used in Southern blotting protocols. However, this technique is limiting owing to the requirement of reasonable amounts of high quality DNA as a starting material. In addition, the procedure is time consuming and labour intensive when handling large numbers of samples. Here, we describe a strategy for the rapid detection of new PCR based DNA polymorphisms. PCR technology is faster, more sensitive, and reduces the requirement for high quality DNA, accessing poor quality material such as DNA obtained from paraffin embedded samples. Boehringer *Taq* polymerase and reaction buffer was used in all PCR amplifications as the reaction buffer does not contain detergent which might interfere in subsequent reactions. In all of the examples given below, at least one frequent polymorphism was detected in the target gene, at the step indicated by following the strategy through steps 1 to 4.

Step 1. Analysis of DNA sequence databases

Comparing DNA sequences (particularly non-coding regions of genes) in databases often shows previously described sequence variants between individual clones. These will sometimes create or destroy a restriction enzyme site and are therefore potentially useful DNA polymorphisms. Primers are selected to amplify a fragment encompassing the potential site and PCR carried out on total DNA from 10 subjects. The amplified products (usually a tenth of the final volume) are then digested with the appropriate restriction enzyme. On confirmation that the sequence variant does constitute an RFLP and is not a cloning artefact, further characterisation is performed as described in step 5. Polymorphisms within the insulin gene have been detected using this approach.^{1,2} If the variant does not alter a restriction enzyme recognition sequence then try step 4. If no sequence differences are identified go to step 2.

Step 2. Restriction analysis of amplified products

Primers are chosen to regions of sequence most likely to contain variations (non-coding, such as introns and 3' or 5' untranslated regions). A 1 to 2 kb fragment is amplified from a panel of 10 unrelated subjects. The amplified PCR products are then digested with a series of four or five base recognition sequence restriction enzymes (the chosen fragment will usually contain several sites for each). Digests are analysed on high percentage agarose gels. Any RFLPs identified are characterised as described in step 5. Polymorphisms within the 3' untranslated region of the candidate Wilms' tumour gene³ (original amplification 952 bp), c-Ki-ras^{4,5} (two separate regions amplified 1300 and 450 bp) and PRAD1⁶ (1200 bp amplification) have been isolated using this approach. If this step is unsuccessful or a greater number of independent polymorphisms at the locus is desired step 3 is used.

Step 3. Sequencing of PCR amplified products

After treatment of the amplification products from six unrelated subjects with Klenow fragment polymerase and polynucleotide kinase, the fragments from each subject are cloned into *Sma*I cut, alkaline phosphatase treated, M13 vector. After transformation, a number of white plaques from each sample are picked into sterile PBS (100 µl), vortexed, and two 3 µl aliquots removed and used to set up two PCR reactions (a and b). Amplifying with one of the original primers and (a) extended normal or (b) extended reverse M13 primers in each tube, the presence and orientation of the insert can be confirmed. The PBS suspension can then be used to inoculate a culture for preparation of single stranded DNA. Six clones in each orientation are then sequenced. The amount of sequencing carried out will depend on the importance of the locus to the operator. However, using rapid protocols, such as the Amersham Multiwell System and sequencing the clone in both directions, the time taken is minimised. Variant sequences identified can then be screened for restriction site alteration using computer mapping. If a potential RFLP is located, it can be confirmed as genuine and not a PCR artefact, by digestion of amplified products from a number of

CRC Department of
Cancer Genetics,
Paterson Institute for
Cancer Research,
Wilmslow Road,
Manchester M20 9BX.
P R Hoban
M F Santibanez-Koref
J Heighway

Correspondence to
J Heighway.

Received 5 September 1991.
Revised version accepted
4 October 1991.

subjects and further characterised as described in step 5. Using this approach, sequencing approximately 400 bp, an RFLP within the N-ras⁷ gene (original amplification 2000 bp) has been identified. If no polymorphism is detected then a separate region can be amplified and the procedure repeated from step 2 onwards. However, if sequence variants are identified but they do not alter a restriction enzyme recognition sequence then go to step 4.

Step 4. Design of modified primers

Sequence variants that do not alter restriction sites can often be used as polymorphisms by designing primers containing mismatches close to the 3' end, which, on amplification, incorporate the polymorphic base into a restriction site (usually the polymorphic base would be the first base 3' to the primer). The polymorphism can then be assayed for by restriction of the PCR product. As the result of this digest will be cleavage of the primer from the product, care should be taken that this will be easily visible on conventional agarose gels (for example, final amplification product 150 to 200 bp). Again, the region can be amplified from a panel of subjects to ascertain correct PCR conditions (which may be critical) and further characterised as described in step 5. Using this approach polymorphisms within the Ha-ras (Hoban *et al*, unpublished data) and Ki-ras⁸ genes have been identified.

Step 5. Characterisation of polymorphism

New polymorphisms are characterised by determining the allele frequency in a population of 50 unrelated subjects after PCR amplification. Mendelian inheritance is confirmed by characterisation of the polymorphism within families and screening for codominant segregation.

Implications

We feel that by using this or similar strategies it is possible to detect multiple, easily assayed, PCR based polymorphisms within virtually

any known gene. The combination of these should lead to high PIC value within the target gene.^{4,5,8} If this proves to be true, then it is probable that within a relatively short time PCR analysis will replace Southern blotting as the major technique in human genetics. The advantage of this strategy over, for example, the search for highly informative dinucleotide repeat polymorphisms is the accuracy with which the polymorphism can be targeted, not just close to a gene of interest but to a particular location within a gene. If the polymorphism is detected in a region of the sequence that is encoded within the mRNA, then using RT-PCR with a cDNA target enables the study of relative expression levels from each parental allele. This approach has been shown to work well for c-Ki-ras, c-Ha-ras, and PRAD1 (unpublished data).

Other methods exist for the detection of variation in DNA sequence (single strand conformational or temperature gradient gel electrophoresis). However, we feel that the use of multiple simple DNA polymorphisms provides a rapid, accessible assay for individual alleles. Given the level of variation in non-coding sequences it does not seem worthwhile to use these techniques initially to detect variants but subsequent polymorphisms identified through the strategy could be analysed in this way, especially if they can not be used in restriction assays.

This work was funded by the CRC and Royal Manchester Children's Hospital, Pendlebury Leukaemia Research Fund, which supports PRH.

- 1 Hoban PR, Kelsey AM. PstI polymorphism within the 3' untranslated region of the insulin gene detectable by the polymerase chain reaction. *Nucleic Acids Res* 1991;19:4576.
- 2 Hoban PR, Kelsey AM. A MaeIII polymorphism within intron A of the insulin gene detectable by PCR. *Nucleic Acids Res* 1991;19:5091.
- 3 Hoban PR, Kelsey AM. HinfI polymorphism within the 3' untranslated region of the candidate Wilms tumour gene. *Nucleic Acids Res* 1991;19:1169.
- 4 Heighway J. RsaI polymorphism in c-Ki-ras. *Nucleic Acids Res* 1991;19:968.
- 5 Heighway J. DdeI polymorphism within sequence encoding 3' untranslated region of c-Ki-ras2. *Nucleic Acids Res* 1991;19:5092.
- 6 Heighway J. HaeIII polymorphism within 3' untranslated region of PRAD1. *Nucleic Acids Res* 1991;19:5452.
- 7 Heighway J. BsmI polymorphism in N-ras intron I. *Nucleic Acids Res* 1991;19:682.
- 8 Heighway J. Primers to convert c-Ki-ras to a four allele RsaI system. *Nucleic Acids Res* 1991;19:6966.