An effective method for allele-specific sequencing using restriction enzyme and biotinylation (ASSURE B)

DE-XING ZHANG and G. M. HEWITT
Population Biology Sector, School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK

DNA sequencing has become an important tool in many areas of molecular biology. In particular, direct sequencing of a polymerase chain reaction (PCR) product serves as an essential technique for detecting genetic polymorphisms in population studies and for revealing mutations in molecular medicine. Because PCR will amplify any target present in the reaction mixture, genomic amplification will yield heterogeneous products if more than one target exists. This is often the case when nuclear DNA is amplified from non-haploid organisms where there is heterozygosity and when the amplified sequence has a multiple copy number. In such situations, a rapid and correct determination of the DNA sequences of each target (allele) is required, but direct sequencing runs into difficulty, because more than one chain termination will occur at a given polymorphic site, thereby producing ambiguous sequences.

Existing methods are not entirely satisfactory in resolving this problem. One method involves cloning of PCR product and then sequencing a number of different clones to determine the sequences of individual targets. However, cloning of PCR products is not straightforward and is very time-consuming, and a poorly represented target requires the sequencing of a large number of clones. Multiple sequencing will increase the cost and labour, especially in large-scale analyses such as population studies. Another method is the use of allele-specific PCR/sequencing primers to amplify or sequence a specific target. There are several disadvantages with this. For example, in a molecular population analysis many different alleles could exist in the populations, and the cost of synthesizing many allele-specific oligos for analysis is therefore substantial. Also, it is not always possible to design a suitable allele-specific primer, as a primer requires a number of features to work effectively. Furthermore, if allele-specific PCR primers are used, different amplifications are needed in order to sequence all the alleles.

So a cost-effective and time-saving method suitable for large scale sample analysis in population and clinical studies is most desirable. We describe below such a method for Allele-Specific Sequencing directly from PCR products Using Restriction Enzyme and Biotinylation (named ASSURE B).

The strategy and procedure of the ASSURE B method is schematized in Fig. 1(a). Double-stranded PCR is carried out using one biotinylated primer and one normal primer. If direct solid-phase sequencing (Stahl et al. 1988; Hultman et al. 1989) reveals heterogeneity, the ambiguous sequence obtained will be analysed for allele-specific restriction sites. Let us assume for example that two different alleles (assigned as ‘A’ and ‘a’, respectively) exist, and allele-A can be specifically cut by the enzyme MspI (Fig. 1a). The PCR product is then digested using the allele-specific enzyme (MspI for allele-A in Fig. 1a) to chop up the target allele (A in Fig. 1a). Using magnetic beads streptavidin, part of the digested DNA will be removed (as they do not contain biotinylated nucleotides) and single-stranded DNA prepared. The unbiotinylated complementary PCR primer, or any internal primer upstream to the last cutting site, is used for solid-phase sequencing. As the remaining part of allele-A does not contain the priming site any more (Fig. 1a), only allele-a will be sequenced. Once the sequence of allele-a is determined, that of allele-A can be deduced from the ambiguous sequence obtained in the first sequencing analysis. Otherwise, if allele-a specific restriction enzymes exist, ‘A’ can be sequenced using the same procedure.

How widely could the ASSURE B method be applied in practice? The key factor in its application is to find suitable restriction enzymes which cut only at polymorphic sites. Note that this method is based on the same principle as the widely employed RFLP techniques in finding polymorphic restriction sites. What is more, this method can predict which enzymes will be useful by analysing raw DNA sequence data. In addition, because restriction enzyme is required only for fragmenting unwanted targets, all three types of restriction endonucleases can be employed (the commonly used enzymes are the simplest and belong to type II, which recognize symmetric sequences and cleave within the sequences at specific sites). This greatly increases the chance of finding a suitable enzyme for a given allele.
Fig. 1 ASSURE B method for allele-specific sequencing. (above) Strategy and procedure of the ASSURE B method for allele-specific sequencing. 'A' and 'a' represent two different alleles, respectively. A biotinylated strand is symbolized with a circled letter 'B' at its 5' end. M in a square represents a magnetic particle coated with streptavidin which binds biotinylated DNA. See text for detailed description. (right) Autoradiograph showing an example using the ASSURE B method to determine unambiguously allele-specific sequences. Double stranded PCR was carried out with one biotinylated and one normal primer, amplifying a heterozygous non-coding nuclear marker in the desert locust *Schistocerca gregaria* (diploid). ASSURE B method was used as described in the text to determine the sequences of each allele. PCR product was directly digested with restriction enzymes without additional purification. Digested DNA was extracted once with chloroform to remove oil before strand separation. Dynabeads M280 Streptavidin (Dynal) was used to remove unwanted DNA and prepare single-stranded DNA following the manufacturer's recommended protocols. Three independent sets of sequencing reactions from the same PCR run were performed: uncut, using intact uncut DNA; Msp, using *MspI* digested DNA for sequencing allele-a (*MspI* cuts specifically allele-A); Pst, using *PstI* digested DNA for sequencing allele-A (*PstI* cuts specifically allele-a). The 8 polymorphic sites differentiating the two alleles are indicated with black dots and arrows. Clear resolution of DNA sequences of alleles is obtained by comparing corresponding bands between the 3 runs. Sequencing order: ACGT. The Sequenase V2.0 manual sequencing kit (USB-Amersham) was used.
Table 1 How many polymorphic sites are cuttable by restriction enzymes? Notes. (a) There are very few sequence data on heterozygous nuclear markers at the present. Here six sets of published DNA sequence data on genetic polymorphisms (nuclear or mtDNA) were analysed to show how many polymorphic sites could fall in a restriction site. What we have done is: (i) randomly choose such data in the literature (most coming from recent publications in Molecular Ecology); (ii) make a consensus sequence by putting ambiguous nucleotides for polymorphic sites just as if a heterozygous nuclear marker is involved (i.e. if both a 'T' and a 'A' were observed at a given site, they will be shown as a 'W' in the consensus sequence); (iii) then find how many polymorphic sites would be cut by restriction enzymes by computer analysis; (iv) list enzymes which only cut at polymorphic sites; and finally (v) identify unique-cutting enzymes suitable for ASSURE B method. It is worth noting that ASSURE B-suitable enzymes do not need to be unique-cutting. If a second site occurs near to the nonbiotin-tagged end, allele-specific sequencing can still be carried out using internal primers. Only part of the data of Cooper et al. (1995) was used. (b) Data were analysed using the DNASTAR computer software package (EditSeq V3.75 & Mapdraw V2.88, 1994).

<table>
<thead>
<tr>
<th>DNA markers and organisms involved</th>
<th>Types of markers</th>
<th>Size (bp)</th>
<th>No. of observed polymorphic sites</th>
<th>REs cuttable polymorphic sites</th>
<th>Examples of REs suitable for the ASSURE B method</th>
<th>Sources of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sgmnl, locust</td>
<td>nuclear</td>
<td>250</td>
<td>19</td>
<td>13 (68%)</td>
<td>Hhal, Mfel, MapI, PstI</td>
<td>This paper (unpublished)</td>
</tr>
<tr>
<td>Cprnl, grasshopper</td>
<td>nuclear</td>
<td>421</td>
<td>19</td>
<td>14 (74%)</td>
<td>Accl, DraI, Hpyd, SphI</td>
<td>Cooper et al. 1995</td>
</tr>
<tr>
<td>MHC intron I, cichlid</td>
<td>nuclear</td>
<td>227</td>
<td>10</td>
<td>7 (70%)</td>
<td>Apel, MaeIII, Rsal, SphI</td>
<td>Klein et al. 1993</td>
</tr>
<tr>
<td>ATPase VI, trout</td>
<td>mtDNA</td>
<td>315</td>
<td>11</td>
<td>9 (82%)</td>
<td>HpyII, MbolI, Msel, NbaI</td>
<td>Giuffra et al. 1994</td>
</tr>
<tr>
<td>Cyt b, trout</td>
<td>mtDNA</td>
<td>295</td>
<td>6</td>
<td>5 (83%)</td>
<td>AcclIII, AluI26I, HaeIII, SphI</td>
<td>Giuffra et al. 1994</td>
</tr>
<tr>
<td>Cyt b, Atlantic cod</td>
<td>mtDNA</td>
<td>307</td>
<td>14</td>
<td>11 (78%)</td>
<td>Ball, BglI, Hhal, Smal</td>
<td>Carr et al. 1995</td>
</tr>
<tr>
<td>D-loop, trout</td>
<td>mtDNA</td>
<td>310</td>
<td>9</td>
<td>6 (67%)</td>
<td>HpaII, Mund, Smal, Xmal</td>
<td>Bernatchez &amp; Osinov 1995</td>
</tr>
</tbody>
</table>

RE(s), restriction endonuclease(s).

especially for highly polymorphic sequences. Table 1 shows the frequency of polymorphic sites within restriction sites, and the chance of finding unique-cutting sites suitable for the ASSURE B method in six sets of DNA sequence data randomly chosen from recent publications (plus our own unpublished data). (In fact, ASSURE B-suited enzymes do not need to be unique-cutting. If a second site occurs near to the nonbiotin-tagged end, allele-specific sequencing can still be carried out using internal primers.) Looking at Table 1, the potential usefulness of this method is obvious. For example, among the 19 polymorphic sites detected in a 250-bp nuclear DNA marker in S. gregaria, 13 sites (i.e. 68%) can be cut by at least one commercially available restriction enzymes (Table 1, authors’ unpublished data).

In practice, we have noticed that due to the presence of single-stranded product in some PCR amplifications, sometimes the DNA of the unwanted targets cannot be cut completely by allele-specific digestion. However, this does not hinder the use of this method because the signals for the digested target are very much reduced on the sequencing gel, allowing correct determination of allelic sequences. Alternatively, adjustment of PCR primer concentration can help to overcome this problem. Figure 1(b) shows an example using the ASSURE B method to determine allele-specific sequences directly from PCR product. It is clear that sequences of both alleles can be obtained unambiguously. Therefore, we believe that this simple and effective method will find wide application in various research fields. Also, it is possible to automate the whole procedure because of its simplicity and the use of solid-phase techniques. For example, part of the method can be incorporated into an automatic DNA sequencer, by developing analytical software to find useful allele-specific enzymes for ‘ASSURE B’, and to output allelic sequences.

Acknowledgements

This work was supported by a grant from EEC on biodiversity research. We are grateful to Dr T.A. Burke for helpful suggestions on an earlier version of this manuscript.

References


