PCR designer for restriction analysis of various types of sequence mutation

Xiayi Ke1,∗, Andrew Collins2 and Shu Ye2

1Wellcome Trust Centre For Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK and 2Division of Human Genetics, University of Southampton, Duthie Building, Southampton General Hospital, Southampton SO16 6YD, UK

Received on March 20, 2002; revised on May 14, 2002; accepted on May 24, 2002

ABSTRACT

Summary: Restriction analysis is widely used to detect gene mutations such as insertions, deletions and single nucleotide polymorphisms (SNPs). Although such mutation sites sometimes present some natural restriction sites to differentiate the wild-type and mutant sequences, mismatches are often needed in order to create artificial restriction fragment length polymorphisms (RFLPs). In this report, a computer program is described that screens for suitable restriction enzymes, introducing mismatches where appropriate and when necessary, designs primers using the information of the selected restriction enzymes, their recognition sequence and locations as well as the information about the mismatches if any. The program, supported by a WWW web interface, is intended to be used online.

Availability: The computer software, with related documentation, is available at http://cedar.genetics.soton.ac.uk/public_html/primer.html.

Contact: xiayi@well.ox.ac.uk

Single base-pair substitutions, deletions and insertions are the major types of human gene mutations (Cooper et al., 1998; Krawczak et al., 2000). According to the human gene mutation database (HGMD) (http://www.uwcm.ac.uk/uwcm/mg/hgmd0.html), as of 14/02/2002, among the 25 256 mutations recorded for 1132 genes in HGMD, 17 657 are single base-pair substitutions, 4149 are small (⩽20 bp) deletions, 1281 are gross (>20 bp) deletions, 1536 are small (⩽20 bp) insertions, 172 are small indels and 461 are other gross lesions such as complex arrangements. A variety of methods including restriction analysis have been developed for typing single base-pair substitutions (Syvänen, 2001), and computer programs also exist that facilitate such analysis (Ye et al., 2001; Ke et al., 2001). Restriction analysis has also been applied in typing small deletions and insertions (Dunleavey et al., 2000). The method usually introduces an artificial restriction site into a PCR product by the use of a primer with a single-base mismatch close to its 3′ end. However, to manually select an appropriate restriction enzyme and design corresponding primers is time-consuming and inefficient. In this communication, we describe a WWW-based computer program to facilitate such a process. The program is an advance of previous programs developed for SNP analysis and now enables the design of primers to detect the majority of the sequence mutations, i.e. deletions, insertions, indels as well as single base-pair substitutions. However, this program may not provide solutions for mutations involving simple repeats such as microsatellites.

The program uses all Type II restriction enzymes in the REBASE Restriction Enzyme Database (http://rebase.neb.com), and screens for only those enzymes that produce a restriction site either for the wild-type or the mutant sequence but not for both. If no natural discriminative restriction sites are found or the number of such sites is less than that requested by the user, mismatches are introduced on both sides of the mutation and on both strands to provide (extra) appropriate artificial restriction sites. Once the number of restriction sites requested by the user is satisfied or an exhaustive search on all possible mismatches has been carried out, the program will move to the primer design stage. The nearest-neighbor parameters (Breslauer et al., 1986) and the formula given by Rychlik et al. (1990) are used to calculate the primer melting temperatures. Primer pair searches always start with the optimum primer and product sizes according to the user inputs. If a mismatch is introduced, the primer that covers the mismatch will always be produced first and the other primer situated on the other strand is then produced to match the mismatch primer.

A simple Javascript-driven web interface is provided for users to input source sequence (up to 1kb) and mutation data and to set the requirements. All ambiguity letters (‘R’, ‘Y’, ‘M’, ‘K’, ‘S’, ‘W’, ‘H’, ‘B’, ‘V’, ‘D’ and ‘N’) in the source sequence and the recognition sequences of the enzymes will be expanded during the search.

*To whom correspondence should be addressed.
**PCR designer**

*********************************************************
No mismatch introduced
---Restriction Analysis Entry 1 (wild-type sequence)---
PCR primers recommended

<table>
<thead>
<tr>
<th>Primer pair sequence</th>
<th>3’ Tm</th>
</tr>
</thead>
<tbody>
<tr>
<td>103 GGCCGAGAAGTCTGTCTTTTC</td>
<td>128</td>
</tr>
<tr>
<td>235 GATGGATGCTCATACGGCAG</td>
<td>214</td>
</tr>
</tbody>
</table>

Recognition sequence: GTNAC
Locations in wild-type sequence: 139, 580
Enzymes: MAEIII

*********************************************************
Mismatch G introduced at position 126
---Restriction Analysis Entry 8 (mutant DNA)---
PCR primers recommended

<table>
<thead>
<tr>
<th>Primer pair sequence</th>
<th>3’ Tm</th>
</tr>
</thead>
<tbody>
<tr>
<td>103 GGCCGAGAAGTCTGTCTTGTC</td>
<td>128</td>
</tr>
<tr>
<td>228 GCTCATACGGCAGCCACATA</td>
<td>207</td>
</tr>
</tbody>
</table>

Recognition sequence: GTCTC
Locations in mutant DNA: 126, 429
Enzymes: ALW26I, BSCQII, BSMAI, BSOMAI

*********************************************************

**Fig. 1. A representative output of the program.** In the first entry, no mismatch was introduced. The second entry was the result of introducing a mismatch G at position 126. In both entries, line 1 indicates whether and where the mismatch is introduced; line 2 indicates the output entry number and whether the output design is intended for the wild-type sequence or the mutant DNA; line 3 to 6 describes the recommended PCR primers; line 7 is the restriction enzyme recognition sequence and in the case of entry 2 where a mismatch is introduced, a '*' symbol above the recognition sequence is used to indicate the position of the mismatch; line 8 gives all the locations of the recognition sequence in either the wild-type or mutant sequence; line 9 to end lists all relevant Type II restriction enzymes.

Figure 1 shows part of the output of using the program to design PCR primers and restriction enzymes for an indel mutation found in the human biotinidase (BTD) gene (Pomponio et al., 1995). The U03274 GenBank sequence (its first 720 base pairs) was used as the source sequence. This indel mutation corresponded to a 7 bp (gcggctg) deletion from position 133 coupled with a 3 bp (tcc) insertion at position 129. The output was obtained by using all the default requirements on the web interface, except the ‘starting position of mutation’ (set to 129), ‘Allele1’ (set to ‘ctctgcggctgt’) and ‘Allele2’ (set to ‘tccctct’). For long deletions or insertions, we suggest that the users set to zero the number of requested mismatches since sufficient number of discriminative restriction sites may have already been present without introducing any mismatches.

**REFERENCES**


