SNPicker: a graphical tool for primer picking in designing mutagenic endonuclease restriction assays

Tianhua Niu¹,∗ and Zhenjun Hu²

¹Division of Preventive Medicine, Brigham and Women’s Hospital, Harvard Medical School, 900 Commonwealth Avenue, MA 02215, USA and ²Bioinformatics Program, Boston University, Boston, MA 02215, USA

Received on May 2, 2004; revised on May 29, 2004; accepted on May 31, 2004
Advance Access publication June 16, 2004

ABSTRACT

Summary: Simple, low-cost and accurate genotyping methods for single nucleotide polymorphisms (SNPs) are in high demand in the post-genome-sequencing era. We present a graphical tool called SNPicker, implemented in Java, which significantly facilitates the design of mutagenic endonuclease restriction assays. SNPicker uses the online NEB REBASE to automatically scan for all possible designs of mutagenic primers that can facilitate the picking of mismatched PCR primers to artificially introduce or abolish a restriction site at the target SNP site. We successfully applied SNPicker in designing endonuclease restriction assays for 14 SNPs for the MTHFR gene, the Coagulation Factor II gene and the Coagulation Factor V gene. The SNP assays designed using SNPicker were cross-validated using the MassARY technology.

Availability: SNPicker, as a software tool in the Web-based SeqVISTA Suite, is freely available at http://zlab.bu.edu/SeqVISTA/. A tutorial for SNPicker is available at http://zlab.bu.edu/SeqVISTA/manual/SNPicker.htm

Contact: tniu@hsph.harvard.edu

INTRODUCTION

Single nucleotide polymorphisms (SNPs), accounting for >90% of all human DNA variants (Twyman and Primrose, 2003), occur at a frequency of approximately 1 SNP/kb throughout the genome when the sequences of any two randomly selected individuals were compared (Bentley, 2000). SNPs are widely used in disease gene mapping (Bell, 2002), genetic testing (Simesk et al., 2002) and pharmacogenomic studies (Ambrose, 2002). Endonuclease restriction assay is a well-developed enzymatic method for SNP genotyping by analysis of DNA restriction patterns. In the case when a SNP of interest does not change a restriction site, one can engineer a mutagenic oligonucleotide primer, creating a ‘half-site’ that constructs a ‘full-site’ for a specific restriction endonuclease jointly with the target mutation (i.e. the complementary ‘half-site’). We previously reported a prototype for a primer design computer program for assisting mutagenic endonuclease restriction assays (MER A)—SNPkit (Hao et al., 2002), which has drawn a significant attention from the genetics community, resulting in a large volume of requests. However, SNPkit, as a ‘bare-bone’ prototypic software, only generates text-based output files containing the lists of various choices of mutagenic primers and their corresponding restriction enzymes. When there are many different mutagenic primer choices, it is not easy for the user to decide which mutagenic primer to pick. For example, for any given mutagenic primer designed by SNPkit, it is difficult for the SNPkit user to check whether its corresponding restriction endonuclease has undesirable cutting sites at other places of the polymerase chain reaction (PCR) amplicon. To pick the ‘best-fit’ primer, the user has to do a manual screen of all the possible mutagenic primer designs using tools such as NEB Cutter (Vincze et al., 2003). This manual process is labor-intensive, time-consuming and error-prone. To overcome this weakness of SNPkit, we developed SNPicker, a software tool written in Java that not only implements the function of SNPkit, but also graphically displays the locations of all the respective recognition sites corresponding to the restriction endonucleases of all automatically designed mutagenic primers. We found that SNPicker significantly lessened the burden of primer picking.

METHODS AND IMPLEMENTATION

Primer design algorithm

We employed the following algorithm using the Java language. The primer design algorithm keeps at least two perfect matches at the 3‘ end of the mutagenic primer to ensure that the primer can extend properly during the PCR. The site of the SNP, in the context of its flanking sequence, is searched exhaustively online against a
T. Niu and Z. Hu

Fig. 1. The schematic illustration of picking a mutagenic primer for a mutagenic endonuclease restriction assay of the C15096T polymorphism of the delta-aminolevulinic acid dehydratase gene by using SNPicker. The left-hand side shows a screenshot of SNPicker. By clicking on the ‘SNPicker’ icon in the Advanced menu of SeqVISTA, a dialog window is displayed (upper right). For this polymorphism, it is a C→T substitution at position 96 in the PCR amplicon (15001–15913 nt at GenBank accession no. X64467), and the ‘number of mismatch’ is selected as 1. SNPicker runs through the online NEB REBASE, and generated a large number of compatible mutagenic primer choices. We picked the mutagenic primer that includes a A→C mutation in the 5′ primer 4-bp upstream of the mutant site, and this mutagenic primer generates a restriction site for endonuclease PvuI, in the C allele, but not in the T allele (lower right).

comprehensive restriction endonuclease database–NEB (New England Biolabs)’s REBASE to see whether the SNP alters a restriction site. If the search is successful, the SNP can be directly genotyped using the standard restriction endonuclease assay. If not, one, two or three mismatches within the 5–10 bp flanking region spanned by the primer sequence can be selected to be introduced into the primer by randomly permuting each ‘mismatch’ position with all four possible bases (A, T, C, G), and those qualified mutagenic primers that do create a ‘half-site’ that constructs a ‘full-site’ for a specific restriction endonuclease jointly with the target mutation will be screened out (Fig. 1). The opposing primer can be designed using either the Primer3 software (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) or other custom software programs such as DNAMAN (Niu et al., 2001) by choosing a primer with a comparable melting temperature (the difference of melting temperature between the upstream and downstream primers is within 2°C), which together with the mutagenic primer, gives rise to a PCR amplicon of a desirable length, typically between 100–180 bp. In addition, SNPicker can be used to perform functions similar to NEBCutter (Vincze et al., 2003, http://tools.neb.com/NEBcutter) by specifying the value of SNP position as −1.

Graphical display of the restriction sites for various designs

The SNPicker program displays all mutagenic primers found by the primer design algorithm described above. The names of the restriction endonucleases, as well as their corresponding cutting sites within the entire PCR amplicon are shown. We put SNPicker as a widget of the SeqVISTA suite (Hu et al., 2003), such that the user can use other functions of SeqVISTA during their primer picking process. For example, the RepeatMasker function of SeqVISTA can be used to mask any repeated sequences. Moving the mouse over the SNP position will invoke a MouseOver event showing a text box that displays the information of the SNP. The graphical window can also be zoomed in or out by the user to view local or holistic features.

Example To illustrate the use of SNPicker, we describe here the design of an MERA for one SNP identified by Niu et al. (2001) using denaturing high-performance liquid chromatography (HPLC), C15096T in the delta-aminolevulinic acid dehydratase gene. This SNP is a C→T substitution at position 96 in a 913 bp PCR amplicon (15001–15913 nt at GenBank accession no. X64467). By specifying one mismatch in applying SNPicker, it can be seen that by substituting ‘A’ with ‘C’
residue in the upstream primer 4 bp upstream of the target site (Fig. 1), the extended PCR product generates a restriction site for endonuclease PvuI in the C allele, not in the T allele. Furthermore, the PvuI site is absent in the rest of the PCR amplicon, which indicates that this is a desirable design. We tested the assay and the results were satisfactory.

Application We applied SNPicker in designing MERAs for 14 SNPs for the 5,10-methylenetetrahydrofolate reductase (MTHFR) gene (rs1994798, rs2066470, MTHFR C-677T, MTHFR A-1298C), the Coagulation Factor II (FII) gene (FII-005389, FII-011111, FII-G20210A) and the Coagulation Factor V (FV) gene (rs4524, rs4525, rs6019, FV Leiden, rs6020, rs6022, and rs6030). All 14 SNPs have also been genotyped independently in a blinded manner at a different laboratory (Harvard Partners Center for Genetics and Genomics, Harvard Medical School, Boston, MA) using the primer-extension chemistry–MassARRAY assay (Jurinke et al., 2002). For each SNP, 96 DNA samples were tested and the results of the different chemistries for each DNA sample were compared. We found the concordance rate was 100% for all the assays designed using SNPicker. Although MERA is a robust and convenient procedure, potential limitations of this technique in SNP analysis include (1) MERA is less amenable to automation because it is a gel-based method; (2) the cost of SNP genotyping can be high for restriction endonucleases with high price tags; and (3) this method has a relatively low throughput due to its manual nature. Therefore, MERA is useful primarily for genotyping a limited number of samples. Potential applications of this method include (1) candidate gene association studies (Szombathy et al., 2000), (2) genome-wide gene mapping using a high-density SNP map (Wicks et al., 2000), (3) genome-wide gene mapping using a high-density SNP map (Wicks et al., 2000), and (3) molecular diagnostics (Yamaki et al., 1998; Jassim et al., 1999; Love-Gregory et al., 2001).

ACKNOWLEDGEMENTS
We thank Dr Richard Roberts (New England Bio Labs, Beverly, MA) for insightful discussions. This research was supported in part by the National Institutes of Health grants R01 HG002518-01, P20GM066401-01 and A08-PGM66401A.

REFERENCES