Molecular Ecology Resources (2008) **8**, 575–577 doi: 10.1111/j.1471-8286.2007.02035.x

COMPUTER PROGRAMS

CISPRIMERTOOL: software to implement a comparative genomics strategy for the development of conserved intron scanning (CIS) markers

B. JAYASHREE,* V. T. JAGADEESH* and D. HOISINGTON†

Bioinformatics Unit, ICRISAT, Patancheru, Andhra Pradesh 502324, India, †GTL-Biotechnology, ICRISAT, Patancheru, Andhra Pradesh 502324, India

Abstract

The availability of complete, annotated genomic sequence information in model organisms is a rich resource that can be extended to understudied orphan crops through comparative genomic approaches. We report here a software tool (CISPRIMERTOOL) for the identification of conserved intron scanning regions using expressed sequence tag alignments to a completely sequenced model crop genome. The method used is based on earlier studies reporting the assessment of conserved intron scanning primers (called CISP) within relatively conserved exons located near exon-intron boundaries from onion, banana, sorghum and pearl millet alignments with rice. The tool is freely available to academic users at www.icrisat.org/gt-bt/CISPTool.htm.

Keywords: CISP markers, CISprimer, intron scanning primers

Received 2 July 2007; revision accepted 21 September 2007

As the number of expressed sequence tags (EST) deposited in databanks increase, there is a concomitant increase in the number of protocols reported to maximize the information content and usefulness of these sequences through comparative approaches. ESTs can serve as a resource for molecular markers and have been successfully mined for simple sequence repeats (SSR) and single nucleotide polymorphism (SNP) markers. Molecular markers are very useful in the characterization of available germplasm through DNA fingerprinting and estimation of genetic diversity, the preparation of molecular maps and in marker-assisted breeding. Some of these markers are reported to function across taxa (Gutierrez *et al*. 2005; Xiao-Ping *et al*. 2007), indicating their transferability across species. A recently reported method to leverage information from model crops for the betterment of understudied crops is the detection of conserved intron scanning primers (CISP) in orphan crops (Feltus *et al*. 2006). Introns are more variable than coding sequences and this variation can be exploited as molecular markers. The method described by Feltus *et al*. (2006) utilizes intron variation between taxa to identify molecular markers, since there is a high level of conservation in the location but not the sequence of the introns across taxa. Model genomic sequences are well annotated for intron and exonic regions, while ESTs provide a means to identify conserved exonic regions. This allows the identification of conserved exons located near exon–intron boundaries to which primers can be designed to assess diversity in the introns suitable for DNA marker identification.

This method has been used to identify markers in onion, banana, sorghum and pearl millet through their EST alignments with rice (Feltus *et al*. 2006; Lohithaswa *et al*. 2007). These markers are useful in exploring poorly characterized genomes for DNA polymorphisms resulting from the sequence variation of the introns. Most published reports on the mining of CISP markers have manually pipelined data through existing bioinformatics tools with in-house scripts to search for conserved introns. We report here the availability of a tool that can automate the steps in identifying conserved intron-spanning exons, design primers and verify them electronically. The software can be accessed through informative user interfaces.

The methodology involves identification of conserved intronic locations between two exonic sequences on the model crop genome (Fig. 1). This is made feasible through

Correspondence: B. Jayashree, Fax: 91-40-30713074; E-mail: b.jayashree@cgiar.org

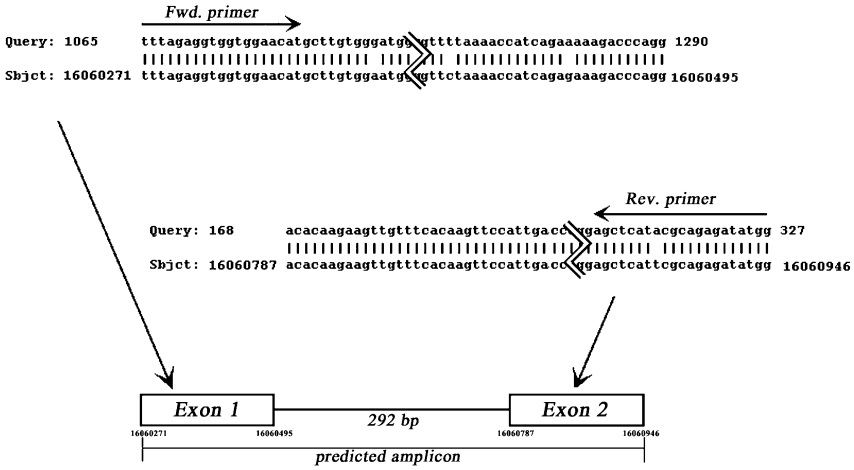

Fig. 1 The design of primers to conserved regions spanning introns.

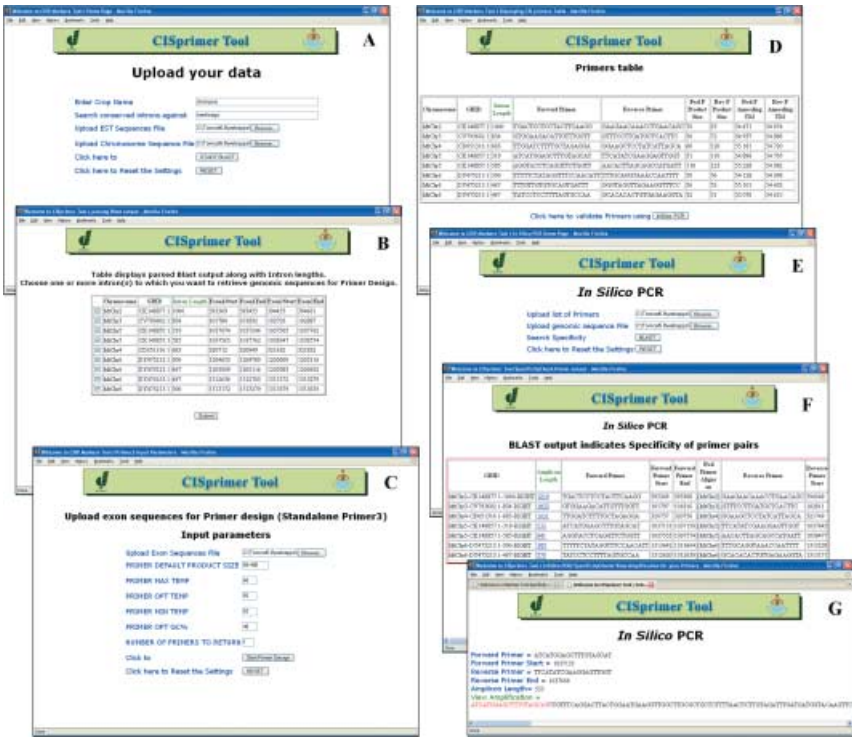

Fig. 2 CISPRIMERTOOL screen displays — (a) an interface to upload data sets and set up search against the model genome, (b) parsed result table, (c) an interface to submit primer design parameters, (d) parsed PRIMER 3 output, (e) the *in–silico* PCR interface, (f, g) interfaces to view *in-silico* PCR output.

alignments of high stringency between the ESTs from the crop of interest with the model crop genome. Primers can then be designed to the model crop exonic regions or the ESTs from the crop of interest that spans an intron of desired length. The CISPRIMERTOOL incorporates a series of steps using tools like BLAST, parsing of output files, substring extractions and primer design, along with an additional step of 'virtual' verification before moving to real world application. The tool has been developed as a stand-alone software with a graphical user interface (GUI) and implements the following steps: (i) perform a BLAST search of the unigene EST data set against the model crop genome; (ii) parse the output and identify ESTs that return two hits on the same chromosome with an intron length spanning 200–2000 bp (at an e-value specified by the user, and with > 95% identity over a minimum of 50 bases of query sequence); (iii) design primers to the conserved exons using the PRIMER 3 software program; and (iii) validate the designed primers first using BLAST to check for specificity of primers and then return the *in-silico* polymerase chain reaction (PCR) product. The *in-silico* PCR code has been written along the lines of the algorithm used in the Primer-Unigene Selectivity (PUNS) system for enhanced primer design (Boutros & Okey 2004).

The software has been implemented in Java and the GUI has been written in Java server pages (JSP). After installation, the user can access the program through the browser. A MySQL database stores data parsed from the output files at each step of the pipeline for later use. The software can be run on either Linux or Windows operating systems. The bioinformatics tools that need to be pre-installed are standalone BLAST (ftp://ftp.ncbi.nih.gov/blast/) and standalone PRIMER3 (https://sourceforge.net/projects/primer3). The other application dependencies are the Apache Tomcat server, a Java runtime environment and the MySQL5.0 database.

The user needs to upload the EST data set as well as the complete genomic sequence database in FASTA format to use in the BLAST searches. The EST data set should preferably be a nonredundant unigene data set. The data sets can be uploaded through the user interface (Fig. 2a). The user can choose e-value settings for the BLAST search. The parsed output file is displayed as a table with the relevant fields (Fig. 2b). The user determines which exonic sequences he would like to use based on the intron length, chromosome number, etc. Once the user determines the data set to use, he needs to specify the primer design conditions (Fig. 2c). These include annealing temperature, expected product size, optimum GC% and number of primer pairs to retrieve. In the event of no primer pairs being returned by PRIMER3, the user may return to the web page and change parameter settings. Based on the parsed PRIMER3 output (Fig. 2d), the user chooses the primers to check for amplification specificity using BLAST, selecting a database or databases that will determine transcriptomic and genomic specificity of the primer pairs (Fig. 2e) followed by amplification using *in-silico* PCR (Figs 2f, g). The output interface returns the chromosome number, product and product size. This report allows the user to accept or reject potential primer pairs for experimental use.

Plans for future improvements include implementing the pipeline with the parallel version of BLAST to reduce run times when searching large genomic databases, and allowing query of the MySQL database. The tool is available at www.icrisat.org/gt-bt/CISPTool.htm. Installation instructions and test files can also be downloaded from this page.

References

Boutros PC, Okey AB (2004) PUNS: Transcriptomic and genomic *in silico* PCR for enhanced primer design. *Bioinformatics*, **20**, 2399–2400.

Feltus FA, Singh HP, Lohithaswa HC *et al*. (2006) A comparative genomics strategy for targeted discovery of single-nucleotide polymorphisms and conserved noncoding sequences in orphan crops. *Plant Physiology*, **140**, 1183–1191.

Gutierrez MV, VazPatto MC, Huguet T *et al*. (2005) Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. *Theoretical and Applied Genetics*, **110**, 1210–1217.

Lohithaswa H, Feltus F, Singh H *et al*. (2007) Leveraging the rice genome sequence for monocot comparative and translational genomics. *Theoretical and Applied Genetics*, **115**, 237–243.

Xiao-Ping J, Yun SUS, Yan-Cun S *et al*. (2007) Development of EST-SSR in foxtail millet (*Setaria italica*). *Genetic Resources and Crop Evolution*, **54**, 233–236.