COMPUTER PROGRAMS

MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design

BRANT C. FAIRCLOTH

D. B. Warnell School of Forestry and Natural Resources, The University of Georgia, Athens, GA 30602, USA

Abstract

MSATCOMMANDER **is a platform-independent program designed to search for microsatellite arrays, design primers, and tag primers using an automated routine.** MSATCOMMANDER **accepts as input DNA sequence data in single-sequence or concatenated,** FASTA-**formatted files. Search data and locus-specific primers are written to comma-separated value files for subsequent use in spreadsheet or database programs. Binary versions of the graphical interface for** MSATCOMMANDER **are available for Apple OS X and Windows XP. Users of other operating systems may run the graphical interface version using the available source code, provided their environment supports at least Python 2.4, Biopython 1.43, and wxPython 2.8.** MSATCOMMANDER **is available from http://code.google.com/p/msatcommander/.**

Keywords: 5′-tails, microsatellites, primer design, primer development, primer tagging, Python

Received 18 March 2007; revision accepted 27 May 2007

Microsatellite array detection and subsequent design of primers to amplify these regions are often bottlenecks that arise during the isolation/development process. Although the use of 5′-tailed primers (Boutin-Ganache *et al.* 2001; Glenn & Schable 2005) has helped greatly reduce costs associated with this process, the careful addition of 5′-tails before polymorphism testing introduces additional time demands. The entire process is usually separated into three distinct phases: array detection, primer design, and primer tagging. Array detection requires accurate identification of repeat arrays and their location within DNA sequence data or sequence contigs. It is often important to identify the types of repeat arrays and the general classes into which they fall (i.e. dinucleotide, trinucleotide, etc.) before designing primers. Several programs have been designed to accomplish this task, including EPHEMERIS (Pentcheff 1999), MSATFINDER (Thurston & Field 2005), and REPEAT MASKER (Smit *et al.* 2004). Primer design requires the selection of specific regions flanking microsatellite arrays that will precisely amplify the target sequence. Commonly used programs accomplishing this task include PRIMER 3 (Rozen & Skaletsky 2000), OLIGO (Molecular Biology Insights), and PRIMER QUEST (Integrated DNA Technologies). Careful application of a 5′-tail to designed primers often requires

Correspondence: Brant C. Faircloth, Fax: 706 542 8356; E-mail: brant@uga.edu

primer design software and laborious user intervention in order to minimize the self-complementarity, pair-complementarity, and secondary structure of modified primer pairs. Because there are no programs that automate the process of identifying good or bad primers for 5′-tailing, many researchers simply accept lower success rates in identifying good primers when using this strategy.

I wrote MSATCOMMANDER to enable rapid and automated microsatellite array detection, locus-specific primer design, and 5′-tailing of designed primers. MSATCOMMANDER is written in Python (Python Software Foundation 2007) and is released under the GNU General Public License (GPL). Python is an open-source, interpreted, object-orientated programming language distributed for multiple platforms (Windows, Linux, and UNIX-like) and is comparable to Perl, Java, or Ruby. MSATCOMMANDER makes use of the BioPython and wxPython modules (available from http://biopython.org and http://wxpython.org, respectively) in addition to several modules distributed with the Python language. MSATCOMMANDER also utilizes the PRIMER 3 software (Rozen & Skaletsky 2000) as its primer design and primer-tagging engine.

I provide binary, executable, graphical user interface (GUI) versions of MSATCOMMANDER for the Apple OS X and Windows XP platforms (Fig. 1). The Apple OS X binaries are universal (native on both Intel and Power PC platforms). I distribute all binaries as archives appropriate for

each respective operating system, and they do not require that Python be installed on the machine running the software. MSATCOMMANDER is also distributed as a source-code tarball for other operating systems (UNIX-like). The GUI may be run on these systems, provided that at least Python 2.4, wxPython 2.8, and Biopython 1.43 are installed. Graphical user interface versions of MSATCOMMANDER have been tested successfully on Apple OS X (10.4) and Windows XP (Service Pack 2).

MSATCOMMANDER uses regular expression pattern matching within each DNA sequence to locate microsatellite arrays within user-selected repeat classes. Length of located repeat arrays within each class is a user-definable option. Repeat sequences are located using their lowest alphabetical, noncomplementary designation, and repeat sequences fitting these designations are written to the summary output file. To locate all microsatellite repeats fitting these designations, DNA sequences are first scanned in the 5′–3′ orientation. The program then takes a second pass through the complement of the sequence in the 3′–5′ orientation. Unknown bases (*N*) are allowed within microsatellite arrays of any class. Users are given a choice of searching for any combination of repeat classes (mononucleotide to hexanucleotide) or all repeat classes at once. Searches progress from simple to complex classes on a sequence-by-sequence basis. Users also have the option to combine multiple repeat arrays located within a single sequence provided they occur within a user-defined distance from one another.

Following the identification of microsatellite arrays, MSATCOMMANDER can design primers for these regions using PRIMER 3 as its primer design engine. PRIMER 3 is a program released under open-source license(s) (available from: http://primer3.sourceforge.net). Primer pairs are only considered when at least 10-bp distant from the start and stop positions of the detected array. Primers are designed to: amplify regions within a size range of 100–500 bp, possess an optimal T_M of 60.0 °C (range 57.0–62.0 °C), possess no greater than a 5.0 °C difference in T_M, possess an optimal GC content of 50%, posses at least a 1-bp GC clamp, possess low levels of self- or pair-complementarity, and have a maximum end-stability (ΔG) of 8.0. Users may additionally edit a single file to change primer features from their default values.

Assuming the option to tag primers is selected, MSATCOMMANDER will attempt to apply two default 5′-tails (CAG: 5′-CAGTCGGGCGTCATCA-3′ or M13R: 5′-GGAAACAG-CTATGACCAT-3′) to designed primers. Each sequence is applied to the 5′ end of the forward primer and the 5′ end of the reverse primer. Shared bases between each tag and primer combination are removed, and PRIMER 3 is used to assess the quality of each modified primer pair. The 5′-tail + primer (forward or reverse) combination yielding the

lowest penalty score is stored, along with primer properties and any warnings from PRIMER 3, within a summary output file. Users are given the option to use both/either of the default 5′-tails or a user-defined, custom 5′-tail.

Currently, MSATCOMMANDER accepts FASTA-formatted input files containing any number of sequences as input. Data are output to either a summary file (array detection only) or a directory at a user-selectable location. MSATCOMMANDER output files are written in comma-separated value (CSV) format. Additional, primer-specific output files created by the PRIMER 3 program are written in text (TXT) format. These files are then easily opened with or exported to common spreadsheet, database, or text-editor programs.

Running time is dependent on user-selected options, processor type and speed, operating system, and available RAM. As an example of the speed of the program with respect to array location, scanning an 80-kb fragment of the Virginia opossum (*Didelphis virginiana*) working draft sequence (GenBank Accession CU075922) for all microsatellite repeat classes took approximately 25 s on a 3.0 GHz Pentium 4 with 2 GB of RAM.

In preliminary testing, MSATCOMMANDER designed 23 tagged primers for a tree species, 19 of which (82.6%) amplified scorable products on an ABI PRISM 3730. I also used MSATCOMMANDER to design 13 tagged primers for an avian species, 11 of which amplified scorable products (84.6% success). These success rates are slightly lower than those achieved when designing primers by hand using commercially available software (88.9%, *n* = 9).

The documentation distributed with MSATCOMMANDER provides GPL license and warranty terms, PRIMER 3 license and warranty terms, and a sample data file. MSATCOMMANDER will be maintained and upgraded for the foreseeable future. All recent versions of the software along with the source code and build instructions are archived at http://code.google.com/p/msatcommander/.

References

Boutin-Ganache I, Raposo M, Raymond M, Deschepper CF (2001) M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods. *BioTechniques*, **31**, 24–28.

Glenn TC, Schable NA (2005) Isolating microsatellite DNA loci. In: *Methods in Enzymology 395, Molecular Evolution: Producing the Biochemical Data, Part B* (eds Zimmer EA, Roalson EH). Academic Press, San Diego, USA.

Pentcheff D (1999) EPHEMERIS version 1.0. Available from URL: http://www.uga.edu/srel/DNA_Lab/programs.htm (Accessed 4/26/2007).

Python Software Foundation (2007) Python version 2.5. Available from URL: http://www.python.org (Accessed 4/26/2007).

Rozen S, Skaletsky H (2000) PRIMER 3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz S, Misener S), pp. 365–386. Humana Press, Totowa, New Jersey.

Smit AFA, Hubley R, Green P (2004) REPEATMASKER Open-3.0. Available from URL: http://www.repeatmasker.org (Accessed 4/26/2007).

Thurston MI, Field D (2005) MSATFINDER: detection and characterisation of microsatellites. Distributed by the authors at URL: http://www.genomics.ceh.ac.uk/msatfinder/. CEH Oxford, Mansfield Road, Oxford OX1 3SR (Accessed 4/26/2007).