

WebSat - A web software for microsatellite marker development

Wellington Santos Martins¹, * , Divino César. Soares Lucas², Kelligton Fabricio de Souza Neves² and David John Bertoli³

¹Instituto de Informatica, Universidade Federal de Goias, Brazil; ²Departamento de Computacao, Universidade Católica de Goias, Brazil;

³Pós-graduação em Biotecnologia, Universidade Católica de Brasília, and Cenargen/Embrapa, Brazil; Wellington S. Martins* - Email: wellington@inf.ufg.br; * Corresponding author

received December 20, 2008; accepted December 28, 2008; published January 12, 2009

Abstract:

Simple sequence repeats (SSR), also known as microsatellites, have been extensively used as molecular markers due to their abundance and high degree of polymorphism. We have developed a simple to use web software, called WebSat, for microsatellite molecular marker prediction and development. WebSat is accessible through the Internet, requiring no program installation. Although a web solution, it makes use of Ajax techniques, providing a rich, responsive user interface. WebSat allows the submission of sequences, visualization of microsatellites and the design of primers suitable for their amplification. The program allows full control of parameters and the easy export of the resulting data, thus facilitating the development of microsatellite markers.

Availability: The web tool may be accessed at <http://purl.oclc.org/NET/websat/>.

Keywords: molecular-marker; microsatellite; SSR; primer; web-software

Background:

The traditional method for developing microsatellite (SSR) makers, using the construction of DNA libraries enriched for SSR sequences, can be expensive and time-consuming. With the increasing availability of sequence data, cheaper and faster *in silico* methods have been used. These are based on computational tools that easily screen sequence data and produce a complete list of SSRs [1]. Some of these tools concentrate on finding SSRs, while others include the additional function of designing PCR primers flanking the SSRs, thus facilitating the marker development process. Included in the latter category are Misa [2], SSRPrimer [3] and IMex [4]. These tools are very useful, providing a standalone version and, in some cases, a web online version as well. However, the web versions of the programs do not usually have a graphical representation for all the steps involved in the process. WebSat makes use of Ajax techniques, providing a rich, responsive user interface, allowing the submission of sequences, visualization of microsatellites, design of primers suitable for their amplification, and exportation of the resulting data.

Software input:

WebSat was written in PHP and JavaScript, making use of Ajax techniques. Its input can be either individual sequences, in raw or FASTA format, or a group of sequences in a multi-FASTA format. The user can also choose to upload a file, with a maximum of 150,000 characters. The input parameters are the motif length and the minimum number of motif repetitions. After entering the sequence(s), and choosing the parameters, the user submits the data, to be processed by an SSR finder program.

WebSat uses TROLL [5] as its SSR finder program due to its simplicity and efficiency. TROLL is targeted exactly at the types of repeats that are most useful for microsatellite marker development: repeats with di-, tri-, tetra-, penta- and hexa-nucleotide motif sizes. Although TROLL does not find imperfect repeats, we did not find this to be a problem. Long imperfect repeats largely consist of shorter perfect ones, and isolated short imperfect repeats are of very limited use for marker development.

Software output:

The output generated by WebSat lists the sequence(s), along with the SSRs found, colored yellow and underlined, in a table format. This is achieved by calling TROLL and, subsequently, parsing the resulting file so as to match the input parameters chosen by the user. To help the user localize the SSR coordinates, the lines are numbered, and groups of ten bases are separated by a space. By moving the mouse over an SSR, the user can find out its motif and length. The user can then click on any SSR to invoke a primer design program to design a pair of primers flanking the SSR.

Primer3 [6] was chosen as the primer design program since it is widely used and provides a rich set of, easily modifiable, parameters. WebSat's output includes, on the top of the window, some of the main primer3 parameter. By changing these values the user can design corresponding pair of primers flanking the SSR selected. Each time the user clicks on an SSR, WebSat creates an input file for primer3, calls it, and parses its output file. Thanks to Ajax techniques, the results are seamlessly integrated into the

WebSat output page. If primer design is successful and a pair of primers is designed, they are colored green along with the SSR in blue. If not, a message reporting the failure of primer design appears. When a primer sequence overlaps with an SSR, the user can still identify the SSR, which is underlined. By providing an easy way to repeatedly change primer3 parameters, the user can make an interactive use of WebSat.

WebSat uses a session control, allowing different users to work on the site at the same time. A user can click on many SSRs, one at a time, and each one associated with possibly different primer3 parameters. Once a SSR is selected, it is colored blue and, if the primer design is successful, the corresponding primer information is stored in a list. At any time the user can choose to save the current list to a CVS file. The exported file contains the following fields for each SSR: the sequence identification, SSR, product size, forward and reverse primer sequence, melting temperature, and coordinates of the primers within the sequence. The file can be easily visualized in a spreadsheet program, by using the option to import external data in CVS format.

Caveat and future development:

Due to server restrictions, large data sets cannot be uploaded and processed by WebSat. Large scale processing can be done, in an automatically way, by other tools such as the one described in [7]. As a future development we plan to incorporate large scale capability into the server.

References:

- [01] A. Merkel and N. Gemmell, *Briefings in bioinformatics*, 9: 355 (2008) [PMID: 18621747]
- [02] T. Thiel *et al.*, *Theoretical and Applied Genetics*, 106: 411 (2003) [PMID: 12589540]
- [03] E. Jewell *et al.*, *Nucleic Acids Research*, 34: W656 (2006) [PMID: 16845092]
- [04] S. B. Mudunuri and H. A. Nagarajaram, *Bioinformatics*, 23: 1181 (2007) [PMID: 17379689]
- [05] T. Castelo *et al.*, *Bioinformatics*, 8: 634 (2002) [PMID: 12016062]
- [06] S. Rozen, and H. J. Skaletsky, *Methods Mol Biol.*, 132: 365 (2000) [PMID: 10547847]
- [07] S. Martins *et al.*, *Nucleic Acids Research*, 34: e31 (2006) [PMID: 16493138]

Edited by P. Kanguane

Citation: Martins *et al.*, *Bioinformatics* 3(6): 282-283 (2009)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.