Journal of Genetics, Genomics & Plant Breeding **1(1) 1-4**

RESEARCH ARTICLE

Survey, analysis and functionality of microsatellite identification tools for the development of microsatellite or simple sequence repeat (SSR) markers for crop improvement

M.K. Rangari [1], S.K. Rangari[2] and A.A.Ramteke[3]

1. Brijlal Biyani Science College, Amravati, 444605, Maharashtra ,India

2. Modern College of Arts, Science and Commerce. Shivajinagar, Pune - 411005. Maharashtra, India
3. Department of Biotechnology, Dr. PDKV., Akola -444001

Corresponding authors email Id: manishakrangari@gmail.com

Manuscript received: March, 20, 2017; Decision on manuscript: June, 28, 2017; Manuscript accepted: July, 15, 2017

Abstract

Total sixteen tools were surveyed for their functionality in particular for the development of microsatellite loci. The surveyed tools are viz., ATR Hunter, IMEx, Misa, Mreps, Msatfinder, Poly, Repeat, Masker, SciRoKoCo, Sputnik,SSRF, SSRLocator, SSRIT, STAR, STRING, TRF, TROLL. The comparative analysis of the features of these tools is presented in paper. Hence, by comparing various features of the microsatellite identification tools it is concluded that a significant difference in the availability of various features and options was observed in these tools. MISA, SSRIT, Sputnik are among the mostly widely used software's for microsatellite identification in various crops e.g. Rice, Wheat, Barley and Maize. Hence, it is important for the researcher to choose the parameters carefully before using the bioinformatic tool to identify the microsatellite markers. Customized search criteria may prove better than the default settings while using these tools.

Key words: Bioinformatic tool, survey, microsatellite, marker, loci.

Introduction

Due to generation and availability of huge genomic information online about the crop plants, the *in silico* studies i.e. performed on computer or via computer simulation are important areas of interest for genomics researchers for comparative genomics study (Dudhe *et al.*, 2012). Microsatellites or simple sequence repeats (SSR) are short tandem repeats (STR), because a repeat unit consists of only 1 to 6 bp and the whole repetitive region spans less than 150 bp (Temnykh et al. 2001). Microsatellites have been variously classified depending upon their size, type of repeat unit and its location in the genome. The majority of DNA sequence and expressed gene sequence data generated these days comes from the next- or second-generation sequencing (NGS/2GS) technologies. NGS technologies produce vast quantities of short read data at a relatively low cost and within a short time which can be mined for genomic SSR development. In recent years, over seven million Expressed Sequence Tags (ESTs) from about 200 plant species have been deposited in public databases. ESTs serve as the main resource for Simple Sequence Repeats (SSRs). The computational approach for detecting SSRs and developing SSR markers from EST-SSRs is preferred over the conventional methods as it reduces time and cost to a great extent. Availability of ESTs with repeat motifs is the key for developing 'functionally characterized sequence motifs' or 'functional markers'. The available EST sequence databases, various web interfaces and standalone tools provide the platform for an easy analysis of the EST sequences leading to the development of potential EST-SSR Markers. Mining microsatellite markers with the help of most efficient microsatellite identification tools that are available freely for academic purpose is a major issue due to availability of number of tools available with different algorithms and features. Hence, understating the power and features of the software in effective mining of microsatellite loci is major challenge in effective mining of microsatellite loci (Temnykh et al. 2001). Traditionally, microsatellites loci discovery involves long and costly molecular biology experiments using magnetic beads for enrichment in specific repeats, followed by cloning, screening and sequencing. The level of polymorphism in the discovered loci is then assessed by PCR amplification from several individual of a population using primers specific for conserved regions flanking the microsatellite loci, followed by length estimation in a capillary-based or acrylamide gel electrophoresis instrument. Depending on the mutation level in the flanking regions, the primers may amplify successfully in genomes of related species of the same genus, or even in more distantly related genera. With the advent of genome sequencing projects, many bioinformatics tools have been developed for mining available sequence data for the presence of microsatellite loci. These tools have methodological bias, and mining one set of sequences with different algorithms commonly results in widely different results (detected loci). In all case, they represent a useful, fast and low-cost alternative for discovering novel microsatellites loci in a genome. Hence, the present survey was carried out to survey, analysis the functionality of microsatellite identification tools for the development of microsatellite or simple sequence repeat (SSR).

Materials and methods

We have surveyed the online microsatellite identification tools available in public domain for their functionality, in particular the program's algorithm and the search parameters used. Total of 16 tools were surveyed for their functionality in particular for the development of microsatellite loci. The tools are *viz*., ATR Hunter, IMEx, Misa, Mreps, Msatfinder, Poly, Repeat, Masker, SciRoKoCo, Sputnik,SSRF, SSRLocator, SSRIT, STAR, STRING, TRF, TROLL.

Results and discussion

There are numerous publications analyzing the microsatellite content of genomes but, in most cases, the results presented cannot be reproduced, mainly due to the lack of details on the microsatellite search process (particularly the program's algorithm and the search parameters used) and because the results are expressed in terms that are relative to the search process (i.e. measures based on the absolute number of microsatellites). A critical understanding of at least few important widely used available software tools designed to scan DNA sequences for microsatellites search is necessary. Till date more than 100 tools are available for SSR discovery. The available software tools vary in-terms of their features, output options and other utilities. Most of the researchers involved in the SSR research work may like the software to provide additional options, apart from extracting microsatellites such as designing primers, a filtering option to extract repeats of a particular size or motif, to get the standardized results, to know their locus information (coding /non-coding), to check the alignments etc. Detailed information on the features provided by each of the tool is provided in Table 1. The link to access these tools along with additional tools is given in Table 2. The software tools also vary in the type / size of repeats that they detect. SciRoKoCo, IMEx and Msatfinder are the only tools that pick repeats of all types (Perfect, Imperfect and Compound). Generally, the bioinformatics tools generate text outputs and therefore may need some post-processing in order to get desired results in legible formats. For example: an option to extract only the repeats of interest (eg. CAG, Poly A repeats) would be of much preference. Few researchers would like to extract only triplet repeats which are responsible for several neurodegenerative diseases in human also may be in need of amino acid produced by those triplet motif. So, an option to detect all repeats of a particular motif as well as an option to extract repeats of a particular size would be a great feature. This facility to extract only the preferred repeat type (such as only mono, or only hexa etc) or to search for a particular repeat (such as ATG, CAG) is provided only in IMEx. The feature to design primer sequences for the repeats detected is much preferred by the researchers. It is available in most of the tools. SSRLocator, IMEx, SciRoKoCo and few other tools can also standardize the repeats and report their corresponding statistics. The comparative analysis of the features of the some of the tools is presented here. Hence, by comparing various features of the microsatellite identification tools it can be concluded that a significant difference in the availability of various features and options was observed in these tools. MISA, SSRIT, Sputnik are among the mostly widely used software's for microsatellite identification in various crops e.g. Rice, Wheat, Barley and Maize. Hence, it is important for the researcher to choose the parameters carefully before using the bioinformatic tool to identify the microsatellite markers. Customized search criteria may prove better than the default settings while using these tools.

References

1. Dudhe**,** M. Y., Meena, H.P., Ranganatha, A. R. G., Mukta, N. and Lavanya, C. 2012. *In silico* identification of conserved domains from EST database in Safflower, J. Oilseeds Res., 29 (Sp. issue): 178-181.
2. Leclercq, S., Rivals, E., and Jarne, P.2007. Detecting microsatellites within genomes: significant variation among algorithms. BMC Bioinformatics, 8: 125.
3. Merkel, A., and Gemmell, N. J.2008. Detecting microsatellites in genome data: variance in definitions and bioinformatic approaches cause systematic bias. Evol. Bioinform., 4: 1-6.
4. Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour S. and McCouch, S. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon association, and genetic marker potential. Genome Res. 11: 1441–1452.

Table 1: Table listing all the microsatellite extraction software and all-important features of the tools.

Tool	Perfect	Imperfect	Compound	Web based	Repeat size (bp)	Primer option	Flanking sequences	Statistic information	Coding non coding region	Alignment	Graphical output
IMEx	Yes	Yes	Yes	**Yes**	1-6	Yes	Yes	Yes	Yes	Yes	No
ATR Hunter	No	Yes	No	Yes	1– 500	No	No	No	Yes	No	No
Misa	Yes	No	Yes	No	1-6	Yes	No	Yes	No	No	No
Mreps	Yes	Yes	No	Yes	1 - > 100	No	No	No	No	No	No
Msatfinder	Yes	Yes	Yes	Yes	1-6	Yes	Yes	Yes	No	No	No
Poly	Yes	No	No	No	1-4	No	No	Yes	No	No	No
Repeat Masker	Yes	Yes	No	Yes	Short (1-5),	No	No	No	Yes	Yes	No
SciRoKoCo	Yes	Yes	Yes	No	1-6	Yes	No	Yes	No	No	No
Sputnik	Yes	Yes	No	Yes	1-5	No	Yes	No	No	No	No
SSRF	Yes	No	No	Yes	1-6	Yes	Yes	No	Yes	No	No
SSRLocator	Yes	No	Yes	No	1-20	Yes	Yes	Yes	No	Yes	No
SSRIT	Yes	No	No	Yes	2-10	No	No	No	No	No	No
STAR	No	Yes	No	Yes	1-9	No	No	No	No	Yes	No
STRING	No	Yes	No	Yes	1-150	No	Yes	No	No	No	Yes
TRF	No	Yes	No	Yes	1–2000	No	Yes	No	No	Yes	No
TROLL	Yes	No	No	Yes	1-5 bp	Yes	Yes	No	NO	No	Yes

*** 'Perfect' indicates whether an option to search 'only perfect repeats' is present or not; An 'Yes' indicates the feature is a part of the software and a 'No' indicates that the feature is missing in that particular tool**

Table 2. List of the few sleeted microsatellite identification tools along with link

S. No	**Microsatellite identification tool**	**Link cited**
1	Perfect microsatellite repeat finder	http://sgdp.iop.kcl.ac.uk/nikammar/repeatfinder.html
2	Microsatellite repeats finder	http://www.biophp.org/minitools/microsatellite_repeats_finder/.php
3	Tandem repeats finder	http://tandem.bu.edu/trf/trf.html
4	Tandem repeat occurrence locator	http://finder.sourceforge.net
5	Mobyle@Pasteur	http://mobyle.pasteur.fr/cgi-bin
6	WebSat	http://wsmartins.net/websat
7	msatcommander	http://code.google.com/p/msatcommander
8	Microsatellite identification tool	http://pgrc.ipk-gatersleben.de
9	SSR PrimerII	http://flora.acpfg.com.au/ssrprimer2/cgi-bin/help
10	SSRIT	http://www.gramene.org/db/markers/ssrtool
11	SSR Finder	http://www.fresnostate.edu/ssrfinder/
12	Sputnik	http://abajian.net/sputnik/index.html
13	Modified Sputnik II	http://wheat.pw.usda.gov/ITMI/EST-SSR/LaRota/
14	Imperfect SSR Finder	http://ssr.nwisrl.ars.usda.gov
15	SciRoKo	www.kofler.or.at/Bioinformatics
16	IMEx	http://imex.cdfd.org.in/IMEX
17	ESMP	http://125.20.82.227/ESMP
18	MIcroSAtellite (MISA)	http://pgrc.ipk-gatersleben.de/misa
19	CUGIssr	http://www.genome.clemson.edu/projects/ssr/
20	SSRSEARCH	ftp://ftp.gramene.org/pub/gramene/software/scripts/ssr.pl
21	TandemSWAN	http://favorov.bioinfolab.net/swan/tool.html
22	MICAS	sunserver.cdfd.org.in:8080
23	Msatfinder	http://www.genomics.ceh.ac.uk/cgi-bin/msatfinder/msatfinder.cgi
24	TandemSWAN (Tandem Structure Word ANalyzer)	http://bioinform.genetika.ru/index.htm
25	mreps	http://www.loria.fr/mreps/.
26	TRStalker	bioalgo.iit.cnr.it.