

Review Paper:

Application of Bioinformatics resources for mining of simple sequence repeats (SSRs) marker in plant genomes: An Overview

Satpathy Raghunath

School of Biotechnology, Gangadhar Meher University, Amruta Vihar, Sambalpur, Odisha, INDIA
rnsatpathy@gmail.com

Abstracts

In the case of plants, the prediction of simple sequence repeats (SSRs) is important for the purpose of gene mapping, biodiversity study and detection of genes with desired characters. In addition to this, due to their polymorphic nature and distribution throughout the genome, it is considered as an ideal marker in plants. However, the traditional methods for the detection of SSR-based polymorphism cause difficulty. As an alternative method, the bioinformatics approaches have been used extensively in the study of these molecular markers, in a economic way.

The next generation sequencing (NGS) techniques generate a huge amount of genomic data which is stored in the public databases in the form of whole genome and EST. This provides the opportunity to implement the bioinformatics tools to predict and annotate the SSR associated with desired gene of the plants. Several computational programs along with the pipelines have been developed to detect the SSR sequences automatically by using the genomic information from the database. In this review, the application and availability of specific bioinformatics resources and the methods of application involved for SSR discovery have been presented by taking suitable examples from the literature.

Keywords: Simple sequence repeats, Bioinformatics tools, Next generation sequencing, Plant breeding, Biodiversity study, Gene mapping, Expressed Sequence Tags.

Introduction

For the last 30 years, the applications of molecular markers are being regularly utilized in plant genomics and breeding areas. Several types of these molecular markers are used for generating the fingerprinting of a plant, genetic map construction including mapping alleles for desirable traits, marker-assisted selection (MAS) and so on.^{44,45} Among different types of molecular markers, the simple sequence repeats (SSRs) found in both the coding and non-coding region are associated with the specific trait and affect gene regulation activity respectively. A microsatellite is a specific representative form of repetitive DNA in which the repetition of DNA occurs commonly in the range of 2-5 base pairs. The microsatellites occur throughout the genome of an

organism and exhibit greater genetic diversity, hence can be used as a suitable molecular marker.

The origin of the repeat patterns in the case of SSRs of a genome due to the mechanism of *slippage* during replication has been reported.^{11,47} The microsatellites such as SSRs are having some of the important unique features like locus specificity, high reproducibility, co-dominance inheritance and hypervariability studied by Squirrell et al.⁴²

The traditional experimental methods that are involved in finding SSR markers, constitute complex time-consuming techniques like genomic libraries construction and the isolation followed by the sequencing of clones.

Another problem that is encountered by the traditional method is the gel electrophoresis techniques used, may not display the perfect genomic sequence that contain the SSRs, so there is a chance to obtain the error-prone information during PCR amplification.^{36,55} But because of cost-effectiveness, the use of microsatellites in plant species has been reduced, therefore, here the use of *in silico* approach has been appreciated being less cost-effective and less time consuming to identify the SSRs, to investigate their functionality and to screen the applicable SSRs for primer design. Bioinformatics is a sophisticated computational approach (storing, processing and analyzing) used to derive meaningful information by using molecular data and play a major role in the integration of the data.

Several computer programs have been developed by the researchers for mining the SSR sequences from the genomic data. As SSR markers are available throughout the genomic sequences also in the ESTs, hence can be detected by use of these programs and available pipelines.^{12,15} Many of the important biological functions can be detected by identification of the SSR sequences, such as regulation of transcription factor and gene expression of a particular trait, other molecular information such as genomic methylation, chromatin organization and so on in the case of plants.^{20,49}

Due to the rapid advancement of genome sequencing techniques, thousands of plant genomes have been completed and the updated data is available in the database. This published genomic information opens the path for the implementation of the available bioinformatics tools for the scientific community to identify microsatellite containing sequences that lead to the development of SSR markers.³⁵

Apart from the genomic sequences, the expressed sequence tags (ESTs) consist of unitary DNA sequences and can be utilized to discover the SSRs known as EST-SSRs. Since the EST represents the coding sequence; the EST-based SSRs represent the physical association with the functional molecular markers of interest for marker assisted selection (MAS) purpose.

The expressed sequence tags (ESTs) are the easiest and cheapest source for SSR improvement and a popular method used in plant molecular biology due to their abundant in the genome, reproducibility, co-dominant inheritance pattern, multi-allelic nature and exhibit a higher level of polymorphism.^{16,17,28} The EST sequence from the database search followed by subsequent computational analysis builds the platform for the direct mode analysis and prediction of EST-based SSR markers. Advances in next-generation sequencing (NGS) technology generate the data consisting of millions of small sequences that can be read and analysed by using computer programs.

Currently, the NGS data are studied for the large-scale mining of molecular markers. Among many of the NGS platforms, the Illumina platform is very widely used for SSR marker mining purposes due to its sequencing accuracy. Also presently, the transcriptome NGS data have been successfully reported for SSR marker development in many plants by using the *de novo* sequencing methods without any reference genome.^{40,43} The objective of the review is to address the bioinformatics techniques, resources available

that are suitably used for the mining of EST-based SSR and their potential application in several plant species by narrating the literature.

Computational resources and techniques in the SSR development: To predict the SSR molecular markers from the NGS and EST information, follow a multi-step procedure starting from retrieval and quality control of NGS data, assembly, prediction and database matching to annotate the function (Figure 1). The EST-SSRs offer various opportunities for discovery of markers for comparative mapping and evolutionary analysis of plant species. However, the possible discovery of the SSR marker can also be done by using the genomic sequence information available in the database. Therefore, presently, the use of bioinformatics techniques is extensively implemented for the mining of the SSR markers from the huge genomics information available in the biological database.^{8,18,38}

The contigs and singleton sequences obtained from the assembly programs are further processed to find out the sequences which contained the types SSR repeats by the use of several programs. Some of the tools and their availability are given in table 1. In a subsequent approach, the functional annotation of the predicted SSRs is carried out by using computational tools that provide functional analysis of protein sequences by classifying them into families and predicting the functional domains.^{2,51} Some of the popular tools for the SSR annotation are given in table 2.

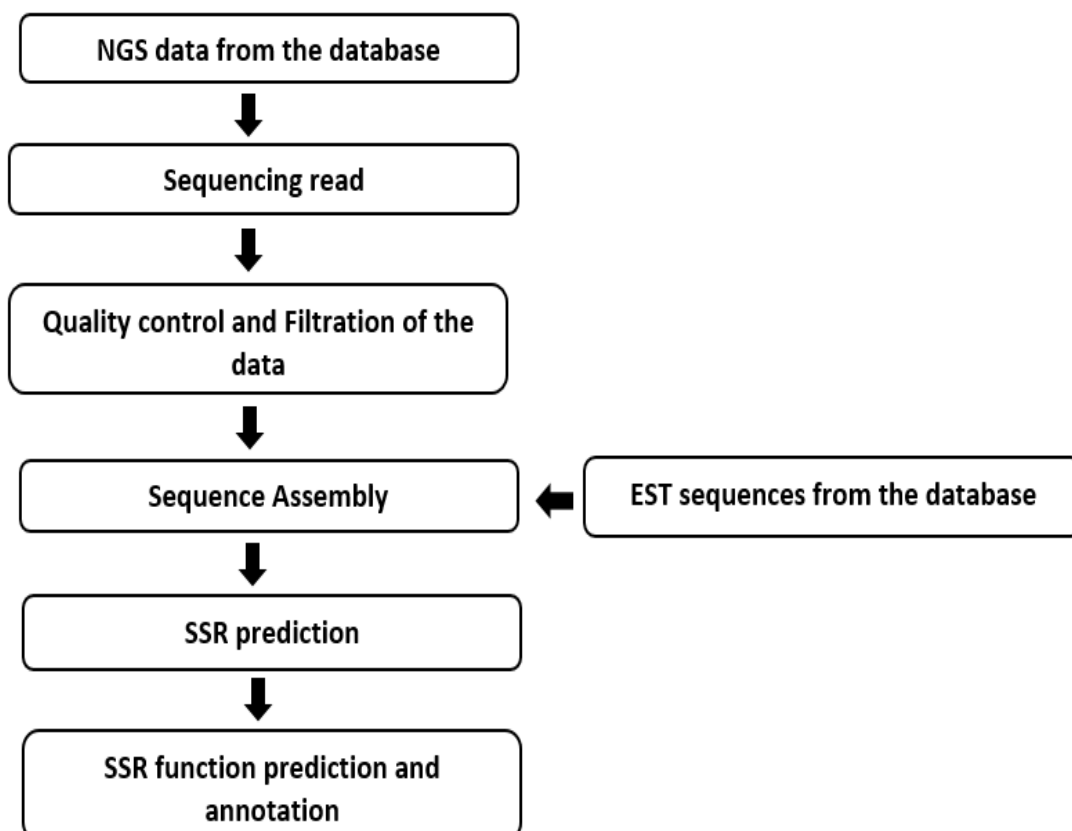


Figure 1: Roadmap to the Bioinformatics based EST-SSR prediction from the genomic sequences

Table 1
SSR prediction tools from the assembled contigs

S.N.	Tool name	Availability	Application
1	WebSat software	https://bioinfo.inf.ufg.br/websat/	Used for prediction of microsatellite based molecular marker
2	SSRIT (SSR marker search tool)	http://archive.gramene.org/db/markers/ssrtool	Valuable computational resource for comparative functional genomics and SSR mining purpose
3	SSRome	http://mggm-lab.easyomics.org/	Contains integrated database and pipelines for microsatellites mining and other allied application
6	MISA-web	http://misaweb.ipk-gatersleben.de/	A web-based computational tool assists the development of microsatellite markers
7	SSR Locator	http://microsatellite.org/ssr.php	SSR prediction
8	Genome-wide Microsatellite Analyzing Tool Package (GMATA)	http://sourceforge.net/projects/gmata/?source=navbar .	Used for SSR mining and calculation of statistics at genome level
10	EST Analysis Pipeline Plus (ESAP Plus)	http://gbp.kku.ac.th/esap_plus/	Used for multi-purpose activities such as pre-processing, clustering and assembly, SSR mining and SSR primer design.
11	Rapid Identification of SSRs and Analysis of primers (RISA)	http://sol.kribb.re.kr/RISA/)	Identification of SSR markers from high-throughput genomics data (200 Mbp).
14	MICAS	http://www.mcr.org.in/micas/	Server used to find non-redundant microsatellites
15	Msatfinder	http://www.genomics.ceh.ac.uk/msatfinder	Used for detection of microsatellite repeats (1-6 bp) in nucleic acid
16	polymorphic SSR retrieval (PSR)	http://sourceforge.net/projects/polyssr/ .	SSR length polymorphism detection tool from the NGS data
17	Poly	http://www.bioinformatics.org/poly/wiki/	Quantitative analysis of simple sequence repeats (SSRs) in the given DNA sequences

Table 2
Tools for the functional annotation of the predicted SSR

S.N.	Name	Availability	Applications
1	Inter Pro Scan	https://www.ebi.ac.uk/interpro/search/sequence/	Functional domains identification
2	Blast-X	https://blast.ncbi.nlm.nih.gov/	Protein functional annotation
3	Blast2GO	https://www.blast2go.com/	High quality functional annotation and analysis of genomic data with visualization effect
4	AmiGO	http://amigo.geneontology.org/amigo	Interactively tool for the search the Gene Ontology data for annotations purpose
5	REViGO	http://revigo.irb.hr/	Prediction of Gene Ontology terms and with visualization effect
6	QuickGO	https://www.ebi.ac.uk/QuickGO/	A web-based platform allows the user for easy browsing of the Gene Ontology
7	NaviGO	http://kiharalab.org/web/navigo .	Tool for visualization and functional similarity analysis with the gene ontology
8	Gene Ontology Functional Enrichment Annotation Tool (GO FEAT)	http://computationalbiology.ufpa.br/gofeat/	free web platform for functional annotation and enrichment of genomic and transcriptomics data

Table 3
Examples of different bioinformatics analysis of the SSR mining in case of some important plants

S.N.	Name of the plant/Scientific name	Family	Application
1	Maize <i>Zea mays</i>	Poaceae	For breeding purposes ⁴⁹
2	Rice <i>Oryzasativa</i>	Poaceae	Mapping of rice genome, that provide the links between the genetic, physical and sequence-based maps ⁴⁶ Study of genetic diversity, primer development for selective breeding ¹⁹
3	Wheat <i>Triticum</i>	Poaceae	Genetic analysis and exploitation of the genetic basis of functional diversity of wheat and related species ³⁰ Investigation of the genetic diversity and evolution ¹³ Mining of salt responsive gene based SSRs in wheat highly useful in elucidating the functional diversity among wheat genotypes ³⁹
4	Safflower Crop <i>Carthamustinctorius</i>	Asteraceae	Development of valuable SSR markers for the genetic analysis of safflower and related species ⁶ Utilization of the developed EST-SSR markers to assess the genetic purity of hybrids ²⁹ Mapping of the flower colour gene in <i>C. tinctorius</i> . ²⁵
5	Carrot <i>Daucuscarota</i>	Apiaceae	Application of the SSR marker as a molecular tools for germplasm characterization and gene tagging in plant breeding studies ⁴⁸
6	Chick Pea <i>Cicerarietinum</i>	Fabaceae	Application in the genetic enhancement studies in chickpea. ³¹
7	Cucumber <i>Cucumissativus</i>	Cucurbitaceae	Application in the enhancement of the efficiency of genetic research and marker-assisted breeding in cucumbers. ²²
9	Melon <i>Cucumismelo</i>	Cucurbitaceae	Study of genetic linkage map, construction, molecular mapping and marker-assisted selection (MAS) in melon. ⁵⁶
10	Sesame <i>Sesamumindicum</i>	Pedaliaceae	Genetic level of assessment, genomic studies and breeding advancement in sesame ⁷
11	Subtropical Bamboo, <i>DendrocalamusHamilt onii</i>	Poaceae	Application in the systematic breeding and implementation for sustainable utilization of bamboo tree. ⁵
12	<i>Arachis Hypogea</i>	Fabaceae	Application in the genetic analysis, comparative genome mapping, population genetic structure and phylogenetic inferences among wild and allied species of <i>Arachis</i> . ³³
13	Barnyard millet (<i>Echinochloa sp.</i>)	Poaceae	Effectively applications to study diversity and mapping aspects of Barnyard millet ³
14	genus <i>Vigna</i>	Fabaceae	Application in genome mapping and comparative mapping among <i>Vigna</i> species ³⁷
15	Date Palm <i>Phoenix dactylifera</i>	Arecaceae	Application in population genetics, selection of the signature, sequence and analysis of the interspecific gene flow studies in Phoenix and other genera. ¹ Application for identification of high diversity and abundance of Transposable genomes (TEs) in the genomes ⁹
16	<i>Sorghum bicolor</i>	Poaceae	Enhancement of resolution of genome analysis in sorghum. ⁴¹
17	Mango (<i>Mangiferaindica L.</i>)	Anacardiaceae	Genetic improvement of mango ²³
18	<i>Phaseolus vulgaris L.</i>	Fabaceae	Gene diversity study in common bean molecular breeding programs for enhancement of common bean yield. ²⁶
19	Winged bean (<i>Psophocarpustetragon olobus</i>)	Fabaceae	Plant breeding system for the genetic improvement of winged bean ⁵⁴

Applications of bioinformatics based SSR mining:

Particularly, the simple sequence repeats (SSR) areas of plant genomics such as genetic analysis and mapping, quantitative trait locus (QTL) mapping and marker-assisted selection (MAS) are included for plant breeding purposes.^{4,34} In addition to this, the plant biodiversity study by the SSRs enables the researchers to explore and select the desired trait in the plant species and improve the varieties by plant breeding methods. In these aspects, the SSR identification by the aid of bioinformatics approach plays a major role and provides efficient and cost-effective methods to plant breeders.^{21,24}

Also, the SSRs are used for tagging genes in many of the crop plants to enhance their yield. To explore and understand the specific traits like stress resistance mechanisms (such as drought tolerance, salt tolerance, disease resistance) of the plants, the resources can be used to develop specific SSRs.

This EST-based method is specifically applied whenever there is no whole-genome sequence data available for the selected plant species.⁵² Another advantage of use of Expressed Sequence Tags (ESTs) based SSR prediction is to identify the presence of novel genes in these aspects.^{14,26}

In recent years, significant computational efforts have been made to develop the SSR markers in several important crop species like maize, wheat, safflower, also crops like wheat, rice, maize, barley, sorghum etc. Many of these SSR mining of these plants are developed and studied by using the ESTs sequences available in public databases. These SSR sequences provide a theoretical estimation of the diversity in the gene expression and can be used in many molecular applications as stated above. The *in-silico* based SSR mining has been implemented in case of several crop species towards various applications and examples of some of them are represented in table 3.

Table 4
Some of the databases contain the plant-specific SSR markers

S.N.	Name of the database	Availability	SSR resources for Plant species
1	MSDB (Microsatellite Database)	http://tdb.ccmb.res.in/msdb	Contains SSR data for 864 plant species
2	PMDBase	http://www.sesame-bioinfo.org/PMDBase	Contains SSR data for 110 plant species
3	Kazusa Marker DataBase	http://marker.kazusa.or.jp	Contains SSR for 14 plant species: <i>Solanumlycopersicum</i> , <i>Arachishypogaea</i> , <i>Raphanussativus</i> , <i>Glycine max</i> , <i>Trifoliumrepens</i> , <i>Trifoliumpratense</i> , <i>Lotus japonicus</i> , <i>Eucalyptus camaldulensis</i> , <i>Capsicum annuum</i> , <i>Fragaria x ananassa</i> , <i>Capsicum annuum</i> , <i>Viciafaba</i> , <i>Vignaangularis</i> , <i>Hieracium</i> species, <i>Trifoliumsubterraneum</i>
4	Gramene Markers Database	https://archive.gramene.org/markers	Basic information about the different markers of plants
5	PiNigSSRdb	http://www.nbgr.ernet.in:9091	SSR markers for black pepper
6	Swiss Vitis Microsatellite Database (SVMD)	http://www1.unine.ch/svmd	SSR markers for grapevines that grow in Switzerland
7	MusatransSSRDB	http://nrdb.res.in/nrcbbio/about.html	SSR resources for <i>Musa</i>
8	citSATdb	http://bioinfo.usu.edu/citSATdb	Contains SSR marker data for 8 citrus species
9	Oryzabase	http://shigen.nig.ac.jp/rice/oryzabase/marker/detail/5588	SSR markers for <i>Oryzasativa</i>
10	Wheat Microsatellite Database	http://webtom.cabgrid.res.in/wheatssr	SSR markers for Wheat
11	TeaPGDB	http://eplant.njau.edu.cn/tea	SSR markers for Tea plant
12	LegumeSSRdb	http://bioinfo.usu.edu/legumeSSRdb	SSR markers for Legumes (13 species) : <i>Glycine max</i> , <i>Cicerarietinum</i> , <i>Medicago truncatula</i> , <i>Trifoliumpratense</i> , <i>Arachishypogaea</i> , <i>Vignaunguiculata</i> , <i>Arachispaensis</i> , <i>Cajanuscajan</i> , <i>Lupinusalbus</i> , <i>Lupinusangustifolius</i> , <i>Vignaangularis</i> , <i>Vignaradiata</i> , <i>Phaseolus vulgaris</i>
13	CicArMiSatDB	https://cegresources.icrisat.org/CicArMiSatDB	SSR markers for Chickpea
14	VigSatDB	http://webtom.cabgrid.res.in/vigna_ssr/ssr.php	SSR markers for Available two species of <i>Vignaradiata</i> and <i>Vignaangularis</i>

Challenges and future perspectives: At present, several bioinformatics tools have been developed for SSR marker prediction along with their specific pipelines of function. These tools are to be implemented in a sequential order to predict the SSR starting from the pre-processing of genomic data to SSR mining and primer designing (Figure 1).

However, no such complete bioinformatics programs are available to do such tasks alone. Currently, available SSR prediction tools have several major limitations. For example, many of these tools require a long run time while analyzing the whole genomic level data. Some tools having limited statistical analyses are provided by the available software. In addition, certain tools display platform dependence, specifically working particular operating systems.

Many programs do not contain the GUI, so may not be user-friendly to the common researchers having little knowledge in bioinformatics. The next limitation is the available software which does not have the capacity for perfect SSR marker development and also contains time-consuming modules for prediction purposes.^{10,35,52} Above all, the proper prediction/identification and classification of SSR markers in terms of their biological functions are still a challenges in genomic studies. Several plant-specific databases containing predicted SSR data have been developed by the researchers for the easy mining polymorphic SSRs, however, the numbers of the predicted plants are few (Table 4).

Hence, the following aspects of bioinformatics applications are desirable:

- Construction of predicted SSR based database of specific economically important crop species to enhance the marker-based breeding research.
- Development and implementation of novel algorithm to identify the correct SSR flanking sequences to avoid the false-positive results in the SSR mining process
- Design of knowledge-based gene interaction models in the case of specific plant species, so that the mined SSR function can be predicted and the gene regulation mechanism can be easily understood and correlated.

Conclusion

Simple sequence repeats (SSRs) constitute one of the most informative molecular markers greatly exploited in plant functional genomics area due to plant breeding. Due to labor-intensive, time-consuming and costly procedure of experimental discovery of SSRs, the available sophisticated bioinformatics approaches are used for the purpose. A wide range of computer programs is available to process the genomic/EST data to process and predict and annotate the SSR in a multi-step manner. The process consists of retrieval of NGS genomic data from the database and is followed by quality control, assembly and SSR prediction. The predicted SSRs are used as the representative to evaluate the genetic polymorphism at specific loci in a selected plant species and are also subsequently applicable for plant breeding purposes.

However, due to limitations in the numbers of plant species for which the SSRs have been predicted, more SSR predictions of the economically important plant species are required in the current time for the enhanced breeding program. Similarly, the SSRs also can be identified by the EST sequence in the absence of the whole genome sequence data. In this review, the roadmap of bioinformatics-based mining of SSRs from the plant genome by using bioinformatics-based approaches has been presented with adequate examples taken from the literature.

References

1. Aberlenc-Bertossi F. et al, In silico mining of microsatellites in coding sequences of the date palm (Arecaceae) genome, characterization and transferability, *Appl. Plant Sci.*, <https://doi.org/10.3732/apps.1300058>, **2** (2014)
2. Atia M.A., Osman G.H. and Elmenofy W.H., Genome-wide in silico analysis, characterization and identification of microsatellites in *Spodoptera littoralis* multiple nucleopolyhedrovirus (SpliMNPV), *Sci. Rep.*, <https://doi.org/10.1038/srep33741>, **6** (2016)
3. BABU B.K. and Chauhan R., In-Silico Identification of EST Based Microsatellite Markers and SNPs and Comparative Genomic Analysis of ESTs in Barnyard Millet for their Omics Applications, *Curr. Agri. Res. Jour.*, <http://dx.doi.org/10.12944/CARJ.5.3.03>, **5**, 279 (2017)
4. Bernardi J. et al, Use of microsatellites to study agricultural biodiversity and food traceability, In *Microsatellite markers*, *IntechOpen*, Doi: 10.5772/64863 (2016)
5. Bhandawat A., Discovery and utilization of EST-SSR marker resource for genetic diversity and population structure analyses of a subtropical bamboo, *Dendrocalamus hamiltonii*, *Biochem. Genet.*, <https://doi.org/10.1007/s10528-019-09914-4>, **57**, 652 (2019)
6. Chapman M.A. et al, Development, polymorphism and cross-taxon utility of EST-SSR markers from safflower (*Carthamus tinctorius* L.), *Theor. Appl. Genet. Theor. Angew. Genet.*, <https://doi.org/10.1007/s00122-009-1161-8>, **120**, 85 (2009)
7. Dossa K. et al, Development of highly informative genome-wide single sequence repeat markers for breeding applications in sesame and construction of a web resource: SisatBase, *Front. Plant Sci.*, <https://doi.org/10.3389/fpls.2017.01470>, **8**, 1470 (2017)
8. Du Q.Z., Zhang D.Q. and Li B.L., Development of 15 novel microsatellite markers from cellulose synthase genes in *Populus tomentosa* (Salicaceae), *Am. J. Bot.*, <https://doi.org/10.3732/ajb.1100308>, **99**, e46 (2012)
9. Filho J.A.F. et al, In silico approach for characterization and comparison of repeats in the genomes of oil and date palms, *Bioinformatics Biol. Insights*, <https://doi.org/10.1177/1177932217702388>, **11** (2017)
10. Gou X. et al, SSRMMD: A rapid and accurate algorithm for mining SSR feature loci and candidate polymorphic SSRs based

- on assembled sequences, *Front. Genet.*, <https://doi.org/10.3389/fgene.2020.00706>, **11**, 706 (2020)
11. Gupta P.K. et al, Microsatellites in plants: A new class of molecular markers, *Curr. Sci.*, **45**, 45-54 (1996)
12. Gupta S. et al, Bioinformatics tools for development of fast and cost effective simple sequence repeat (SSR) and single nucleotide polymorphisms (SNP) markers from expressed sequence tags (ESTs), *African Journal of Biotechnology*, 10.5897/AJB12.1410, **12** (2013)
13. Han B. et al, Genome-wide analysis of microsatellite markers based on sequenced database in Chinese spring wheat (*Triticum aestivum* L.), *PLoS One*, <https://doi.org/10.1371/journal.pone.0141540>, **10**, e0141540 (2015)
14. Haq S.U. et al, Identification and characterization of microsatellites in expressed sequence tags and their cross transferability in different plants, *Int. J. Genomics*, <https://doi.org/10.1155/2014/863948>, **2014**, 863948 (2014)
15. Ilyin S.E., Belkowski S.M. and Plata-Salamán C.R., Biomarker discovery and validation: Technologies and integrative approaches, *Trends Biotechnol.*, <https://doi.org/10.1016/j.tibtech.2004.06.005>, **22**, 411 (2004)
16. Kantety R.V. et al, Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat, *Plant Mol. Biol.*, <https://doi.org/10.1023/a:1014875206165>, **48(5-6)**, 501 (2002)
17. KoohiDehkordi M.K., Beigzadeh T. and Sorkheh K., Novel in silico EST-SSR markers and bioinformatic approaches to detect genetic variation among peach (*Prunus persica* L.) germplasm, *J. For. Res.*, <https://doi.org/10.1007/s11676-019-00922-z>, **31**, 1359 (2020)
18. Kumar B., Kumar U. and Yadav H.K., Identification of EST-SSRs and molecular diversity analysis in *Mentha piperita*, *Crop J.*, <https://doi.org/10.1016/j.cj.2015.02.002>, **3**, 335 (2015)
19. Kumar D. et al, Mining and characterizing the SSR markers for black rice using the Illumina sequencing platform, *Preprints*, <https://doi.org/10.20944/preprints202003.0119.v1> (2020)
20. Li Y.C. et al, Microsatellites within genes: Structure, function and evolution, *Mol. Biol. Evol.*, <https://doi.org/10.1093/molbev/msh073>, **21**, 991 (2004)
21. Liang W. et al, Genetic diversity, population structure and construction of a core collection of apple cultivars from Italian germplasm, *Plant Mol. Biol. Report*, <https://doi.org/10.1007/s11105-014-0754-9>, **33**, 458 (2015)
22. Liu J. et al, Development of genomewide simple sequence repeat fingerprints and highly polymorphic markers in cucumbers based on next-generation sequence data, *Plant Breed.*, <https://doi.org/10.1111/pbr.12304>, **134**, 605 (2015)
23. Mahato A.K. et al, Leaf transcriptome sequencing for identifying genic-SSR markers and SNP heterozygosity in crossbred mango variety "Amrapali" (*Mangifera indica* L.), *PLoS One*, <https://doi.org/10.1371/journal.pone.0164325>, **11**, e0164325 (2016)
24. Martins S. et al, Western European wild and landraces hazelnuts evaluated by SSR markers, *Plant Mol. Biol. Report*, <https://doi.org/10.1007/s11105-015-0867-9>, **33**, 1712 (2015)
25. Mayerhofer R. et al, Development of molecular markers and linkage maps for the *Carthamus* species *C. tinctorius* and *C. oxyacanthus*, *Genome*, <https://doi.org/10.1139/g10-002>, **53**, 266 (2010)
26. Mir R.R. et al, Allelic Diversity, Structural Analysis and Genome-Wide Association Study (GWAS) for Yield and Related Traits Using Unexplored Common Bean (*Phaseolus vulgaris* L.) Germplasm from Western Himalayas, *Front. Genet.*, **11**, 1797 (2020)
27. Mochida K. and Shinozaki K., Advances in Omics and Bioinformatics Tools for Systems Analyses of Plant Functions, DOI: 10.1093/pcp/pcr153, *Plant Cell Physiol.*, **52**, 2017 (2011)
28. Nagaraj S.H., Gasser R.B. and Ranganathan S., A hitchhiker's guide to expressed sequence tag (EST) analysis, *Brief Bioinform.*, <https://doi.org/10.1093/bib/bbl015>, **8**, 6 (2007)
29. Naresh, V., et al, EST-SSR marker-based assay for the genetic purity assessment of safflower hybrids. *Euphytica*, <https://doi.org/10.1007/s10681-009-9995-3>, **170**, 347 (2009)
30. Nicot N. et al, Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs), *Theor. Appl. Genet. Theor. Angew. Genet.*, <https://doi.org/10.1007/s00122-004-1685-x>, **109**, 800 (2004)
31. Parida S.K. et al, Development of genome-wide informative simple sequence repeat markers for large-scale genotyping applications in chickpea and development of web resource, *Front. Plant Sci.*, <https://doi.org/10.3389/fpls.2015.00645>, **6**, 645 (2015)
32. Prohens J. and Picó B., Application of genomic tools in plant breeding, *Curr. Genomics*, <https://doi.org/10.2174/138920212800543084>, **13**, 179 (2012)
33. Sardar S.S. et al, In silico mining of EST-SSRs in *Arachis hypogaea* L. and their utilization for genetic structure and diversity analysis in cultivars/breeding lines in Odisha, India, <https://doi.org/10.1007/s11032-016-0466-y>, *Mol. Breed.*, **36**, 49 (2016)
34. Scarano D. and Rao R., DNA markers for food products authentication, *Diversity*, <https://doi.org/10.3390/d6030579>, **6**, 579 (2014)
35. Sharma P.C., Grover A. and Kahl G., Mining microsatellites in eukaryotic genomes, *Trends Biotechnol.*, **25**, 490 (2007)
36. Sharp M.J. and Sharp P.J., Targeted development of informative microsatellite (SSR) markers, *Nucleic Acids Res.*, <https://doi.org/10.1093/nar/29.8.e44>, **29**, E44 (2001)
37. Shukla N. et al, Mining and analysis of simple sequence repeats in the chloroplast genomes of genus *Vigna*, *Biotechnol. Res. Innov.*, <https://doi.org/10.1016/j.biori.2018.08.001>, **2**, 9 (2018)
38. Simko I., Development of EST-SSR markers for the study of population structure in lettuce (*Lactuca sativa* L.), *J. Hered.*, <https://doi.org/10.1093/jhered/esn072>, **100**, 256 (2009)

39. Singh A.K. et al, Identification, analysis and development of salt responsive candidate gene based SSR markers in wheat, *B.M.C. Plant Biol.*, <https://doi.org/10.1186/s12870-018-1476-1>, **18**, 249 (2018)
40. Singh V.K. et al, Next-generation sequencing (NGS) tools and impact in plant breeding. In *Advances in plant breeding strategies: Breeding, biotechnology and molecular tools*, 563–612, Springer, DOI:10.1007/978-3-319-22521-0_20 (2015)
41. Song J., Genome-Wide In Silico Analysis of Microsatellites in Sorghum bicolor, In *International Conference on Future BioMedical Information Engineering (FBIE)*; IEEE Publications, 556–559 (2009)
42. Squirrell J. et al, How much effort is required to isolate nuclear microsatellites from plants?, *Mol. Ecol.*, <https://doi.org/10.1046/j.1365-294x.2003.01825.x>, **12**, 1339 (2003)
43. Taheri S. et al, Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants, *Molecules*, <https://doi.org/10.3390/molecules23020399>, **23**, 399 (2018)
44. Tautz D. and Renz M., Simple sequences are ubiquitous repetitive components of eukaryotic genomes, *Nucleic Acids Res.*, <https://doi.org/10.1093/nar/12.10.4127>, **12**, 4127 (1984)
45. Tautz D. and Schlötterer C., Simple sequences, *Curr. Opin. Genet. Dev.*, [https://doi.org/10.1016/0959-437x\(94\)90067-1](https://doi.org/10.1016/0959-437x(94)90067-1), **4**, 832 (1994)
46. Temnykh S. et al, Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations and genetic marker potential, *Genome Res.*, <https://doi.org/10.1101/gr.184001>, **11**, 1441 (2001)
47. Tóth G., Gáspári Z. and Jurka J., Microsatellites in different eukaryotic genomes: Survey and analysis, *Genome Res.*, <https://doi.org/10.1101/gr.10.7.967>, **10**, 967 (2000)
48. Uncu A.O. and Uncu A.T., High-throughput simple sequence repeat (SSR) mining saturates the carrot (*Daucus carota* L.) genome with chromosome-anchored markers, *Biotechnol. Biotechnol. Equip.*, <https://doi.org/10.1080/13102818.2019.1701551>, **34**, 1 (2020)
49. Varshney R.K. et al, Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice, *Plant Sci.*, <https://doi.org/10.1016/j.plantsci.2004.08.001>, **168**, 195 (2005)
50. Varshney R.K. et al, In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species, *Cell Mol. Biol. Lett.*, **7**, 537 (2002)
51. Vásquez A. and López C., In silico genome comparison and distribution analysis of simple sequences repeats in cassava, *Int. J. Genomics*, <https://doi.org/10.1155/2014/471461>, **2014**, 471461 (2014)
52. Wang X. and Wang L., GMATA: An integrated software package for genome-scale SSR mining, marker development and viewing, *Front. Plant Sci.*, <https://doi.org/10.3389/fpls.2016.01350>, **7**, 1350 (2016)
53. Wang Z. et al, De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*), *BMC Genomics*, [10.1186/1471-2164-11-726](https://doi.org/10.1186/1471-2164-11-726), **11**, 1 (2010)
54. Wong Q.N. et al, Development of gene-based SSR markers in winged bean (*Psophocarpus tetragonolobus* (L.) DC.) for diversity assessment, *Genes*, <https://doi.org/10.3390/genes8030100>, **8**, 100 (2017)
55. Yang J. et al, Target SSR-Seq: A novel SSR genotyping technology associate with perfect SSRs in genetic analysis of cucumber varieties, *Front. Plant Sci.*, <https://doi.org/10.3389/fpls.2019.00531>, **10**, 531 (2019)
56. Zhu H. et al, Development of genome-wide SSR markers in melon with their cross-species transferability analysis and utilization in genetic diversity study, *Mol. Breed.*, <https://doi.org/10.1007/s11032-016-0579-3>, **36**, 1 (2016).

(Received 03rd January 2022, accepted 07th March 2022)