DOI: 10.15740/HAS/AU/11.2/163-172
*Agriculture Update*
Volume **11** | Issue 2 | May, 2016 | 163-172
Visit us : *www.researchjournal.co.in*
e ISSN-0976-6847

**A REVIEW :**

# Tools for simple sequence repeat (SSR) markers

**ALOK KUMAR SINGH, N.K. SINGH, V.K. SINGH, D.P. SINGH AND N.P. SINGH**

**ARTICLE CHRONICLE :**
**Received :**
*01.02.2016;*
**Accepted :**
*24.04.2016*

**SUMMARY :** SSRs or microsatellites are tandem repeats of 2-8nt units of DNA and are ubiquitous in all genomes studied so far. SSR markers have many advantages over the other marker systems. The first advantage is their high reproducibility, which would be the most important in genetic analysis. The second advantage of the SSR marker system is the polymorphic genetic information contents. The third advantage has to do with the co-dominant nature of SSR polymorphisms. The fourth advantage of the SSR marker system is their abundance and distribution in genomes. A fifth advantage of the SSR marker system is that SSRs are preferentially associated with non-repetitive DNA. This review focuses on some of the reasons for SSR mutations that occur due to replication or repair process which may depend on not only the motif size but also the nucleotide composition of each motif as well as orientation of repeats or position with reference to replication origin. In this review tools for SSRs available are given with their advantages and disadvantages.

**How to cite this article :** Singh, Alok Kumar, Singh, N.K., Singh, V.K., Singh, D.P. and Singh, N.P. (2016). Tools for simple sequence repeat (SSR) markers. *Agric. Update*, **11**(2): 163-172 **(DOI : 10.15740/HAS/AU/11.2/163-172).**

**KEY WORDS :**
SSRs marker, RFLP, RAPD, AFLP, PCR, CID, SAT, TROLL, MISA

**A**uthor for correspondence :
**ALOK KUMAR SINGH**
Krishi Vigyan Kendra (NDUAT), MAU (U.P.) INDIA
See end of the article for authors' affiliations

## BACKGROUND AND OBJECTIVES

SSRs or microsatellites are tandem repeats of 2-8nt units of DNA and are ubiquitous in all genomes studied so far. SSR have varying density and motif distribution that may be species specific even in genomes with low SSR density. For example, among eukaryotes, bivalves have one of the lowest SSR densities but molluscan specific repeat density variations are reported (Cruz *et al*., 2005). Similarly, variations are seen in SSRs in mitochondrial DNA of *Procypris rabaudi* (Tchang) (an endemic fish species in China), (Zhang *et al.,* 2009). These species specific variations can be explained on the basis of differences in repair systems (Pérez *et al*., 2005) or mutation rates.

SSRs have functional roles in gene regulation, chromatin modeling, recombination, evolution, development of new genes, evading host immune response in pathogens, adaptation, resistance to environmental stresses, binding of regulatory proteins to RNA structure by formation of hairpin loop, affect efficiency of exon splicing, protein functions etc. whether they are present in coding or non-coding or intragenic regions (Amador *et al.*, 2004; Kashi and King, 2006; Mrázek, 2006; Sreenu *et al*., 2007 and Coil *et al*., 2008). In primates, repeats like ATn could be involved in segmental duplications that can mediate evolutionary rearrangements (Kehrer-Sawatzki and Cooper, 2008). AT-rich

repeats affect replication dynamics by possibly blocking replication, decrease the efficiency of nucleosome assembly and DNA supercoiling (Yamakoshi *et al*., 2005; Mrázek, 2006 and Lukusa and Fryns, 2008). However, all SSRs may not function as contingency loci or recombination hot spots (Mrázek, 2006).

Despite several functions attributed to repeats, it is also known that mutations in SSRs can be source of disorders including cancer and neurodegeneration (Hancock *et al.,* 2001; Hancock and Santibanez-Koref, 1998 and Bacolla and Wells, 2009). SSRs have high mutation rates which are more in intergenic sequences compared to coding sequences. These mutation rates within a genome may differ depending not only on the motif types but also on length of repeats (Jacob and Eckert, 2007 and Eckert and Hile, 2009). Repeat expansion or deletion can happen due to replication slippage which is considered as one of the main reasons for SSR mutations. The other reasons that cause repeat instability have been attributed to faulty repair/ recombination processes. However, as gleaned from other studies, repeats also expand or contract independent of errors during replication, repair or recombination events. This review focuses on some of the reasons for SSR mutations that occur due to replication or repair process which may depend on not only the motif size but also the nucleotide composition of each motif as well as orientation of repeats or position with reference to replication origin. This review does not focus on details of mechanisms of SSR mutations.

**Biological functions of SSRs :**

SSRs were generally deemed to be evolutionarily neutral. However, numerous lines of evidence have demonstrated that SSRs are not distributed randomly in the genome (Morgante *et al.*, 2002 and Tóth *et al*., 2000). It is estimated that 14 per cent of the genes in eukaryotic species contain repeated sequences, approximately three times more than in prokaryotes (Marcotte *et al*., 1999). Incorporation of repeat sequences in eukaryotic genomes may confer an evolutionary advantage of adaptability to new environments (Marcotte *et al*., 1999; Wren *et al*., 2000). Debates on the functional role (s) of the SSRs on species adaptation and survival have been well documented (Li *et al*., 2002 and 2004). However, the findings of expansion and contraction of the SSR motifs within genes have encouraged the assignment of a biological role to SSRs. Thus, far, the best known cases of SSRs with phenotypic effects are the human loci of Huntington's disease and fragile-X (Cummings and Zoghbi, 2000). SSRs on the UTR regions may also be involved in the regulation of expression of nearby genes as shown by a GT repeat in the Tilapia prolactin 1 gene in fish, in response to a salt-challenged environment (Streelman and Kocher, 2002). Intronic SSRs can regulate gene expression by influencing mRNA splicing or by translocation of mRNA to cytoplasm, as shown by the CCTG repeat in the first intron of the human zinc finger protein 9 (ZNF9), in which an expansion of the repeat causes one intron splicing to fail, thus, leading to myotonic dystrophy (Liquori *et al*., 2001). Although biological roles for SSRs in plants have not been reported as yet, similar roles are expected for these molecular markers in plant genes.

If some SSRs are functional and confer an adaptive advantage, are these functional SSRs suitable in the assessment of biodiversity and ecological conservation of endangered species? Most of the molecular markers that have been utilized in population genetics have not undergone selection and, therefore, have been essentially neutral. In neutral theory, the frequency of alleles is determined by molecules, purely stochastic processes (Li and Graur, 1991). In conservation biology, neutral molecular markers may be useful in providing fundamental information about the types of mating in a population, gene flow and the population history of a species (Ennos, 1996). However, there was a large discrepancy between genetic divergence as measured by neutral RAPD markers and that measured by quantitative genetic traits of the monkey puzzle tree (*Araucaria araucaucana*), a vulnerable tree endemic to southern South America (Bekessy *et al*., 2003). Van Tienderen *et al*. (2002) contend that gene-targeted functional markers can contribute to *ex-situ* management of genetic resources, studies on ecological diversity and conservation of endangered species. Holderegger *et al*. (2006) proffer a theory on the adaptive versus neutral diversity for landscape genetics in which the diversity measured by neutral markers is well suited for the study of processes of gene flow within landscapes, whereas diversity assessed by quantitative genetic experiments using functional markers is best suited for measuring the evolutionary or adaptive potential of a population or species. They concluded that ecologists must recognize

these differences between neutral and adaptive genetic variation when interpreting the results of landscape genetic studies. However, it should be remembered that variation in functional genes might reflect the past influence of selection, which can be variable in each gene and can affect the profiles of variation from the history, migration and drift (Holderegger *et al*., 2006). While genomic SSR markers are mostly neutral, genic SSRs from EST's or cDNAs may retain some adaptive roles. This duality in selection and adaptation ascribes another advantage to the utilization of SSRs in characterizing the genetic diversity of the resources that are preserved in different germplasm institutes.

**Cross-species applications :**

If SSRs are isolated for which primers can be designed, there is no doubt that the SSR marker system has many advantages over other marker systems. If the sequence information is insufficient to develop SSR markers, it may be advantageous to utilize primer sequences identified for one species in the analysis of other closely related species, given the high cost of developing useful SSR markers. SSRs from non-coding regions were not successful in cross-species amplifications due to the sequence variation surrounding SSR motifs, whereas SSRs from coding regions were successful in a wide range of species. In the most extensive cases, 17 SSRs were able to amplify across fish, which diverged about 470 million years ago (mya) (Rico *et al.*, 1996); six SSRs isolated from marine turtles amplified freshwater turtles, which diverged about 300 mya (Fitz Simmons *et al*., 1995). Peakall *et al*. (1998) demonstrated that although 31 per cent of soybean SSR loci were transferable to other legume species, the useful transferability was restricted to congeners. In grass species, Chen *et al*. (2002) selected 11 SSR markers from *Oryza sativa* with the following criteria: (i) high allelic variation in *O. sativa* cultivars, (ii) a variety of perfect and compound SSR motifs, and (iii) minimal stutter bands. The 11 SSR loci were all amplifiable among the *Oryza* species having the same A genome, whereas 73 per cent (8/11) of primers amplified *Oryza* species having the other genomes B and C, and 27 per cent (3/11) amplified species in other genera. Transferability of EST-SSR markers is high. Gupta *et al*. (2003) demonstrated that 43 of 78 EST-SSR markers exhibited transferability from *Triticum* to *Hordeum*, indicating that the sequences flanking SSR motifs were conserved not only within a single genus but also between related genera in the Poaceae family. However, SSR markers derived from genomic SSR-enriched library showed poor cross-species amplification between species from a different genus. Only two of eleven SSR markers from an enrichment library of *Swietenia humilis* showed amplification across the Meliaceae family (White and Powell, 1997). Regardless, it can be concluded that SSR primers from a SSR-enrichment library are still useful in the analysis of species within a genus. Recently, we isolated 12 SSR loci from *Amaranthus hypochoriacus* and were able to demonstrate cross-amplification of these SSR markers to 18 other wild species in the genus *Amaranthus* (Lee *et al*., 2008). Similar results were obtained with the SSRs isolated from the common buckwheat (*Fagopyrum esculentum*) in cross-species amplifications with other species in the genus *Fagopyrum* (Ma *et al.*, 2009).

One should be cautious in using transferable SSR markers for assessing species relationships since the maintenance of allele sizes among species is complex. The complexity of the mutation process in SSRs as well as size homoplasy may complicate the interpretation of SSR variations. Size homoplasy was frequently detected among cross-species amplified SSR markers between species. In the study of cross-species amplifications in *Oryza* species Chen *et al*. (2002) demonstrated that mutations occurred not only in the repeat units but also in the flanking regions to show allele size homoplasy as well as cryptic alleles. Similar results have been reported in studies with *Pinus* species (Kostia *et al*., 1995) as well as legume species (Peakall *et al*., 1998). Therefore, it is recommended that inferences *vis-a-vis* species relationships using SSRPs be accompanied by the information underlying sequences.

**Advantages of SSR analysis :**

SSR markers have many advantages over the other marker systems. The first advantage is their high reproducibility, which would be the most important in genetic analysis. While reproducibility of the SSR profile is as robust as it is with RFLPs, experimental procedures for SSR analysis are much simpler and require only a small amount of template DNA. Since SSR analysis does not require restriction with enzymes, it can reproduce the same profiles regardless of the state of the template DNA. It also does not require template DNA to be ultrapure, which is a requirement in AFLP analysis since

contaminated or impure DNA is often recalcitrant in restriction enzyme digestions to produce nonspecific spurious bands. This is a real benefit when one is dealing with specimens that are dry, contaminated, mummified or even in fossilized form in the wild (Manen *et al*., 2003 and Boder *et al*., 2006).

The second advantage of the SSR marker system is the polymorphic genetic information contents. The hyper-variable nature of SSRs produces very high allelic variations even among very closely related varieties. A literature survey showed that the number of alleles varied from 1 to 37 with diversity indices of 0.29–0.95 in major crop species (Powell *et al*., 1996). The level of genetic variation detected by SSRPs analysis was almost two times higher than that detected by RFLPs, with 61 soybean lines (Morgante *et al*., 1994). In a comparative study of the utility of RFLP, RAPD, AFLP and SSR marker systems for germplasm analysis, SSRs showed the highest expected heterozygosity, while AFLPs had the highest effective multiplex ratio (Powell *et al*., 1996).

The third advantage has to do with the co-dominant nature of SSR polymorphisms. Although homoplasious bands can be misleading in scoring SSR profiles, the SSR bands produced from the same set of primers are intuitively orthologous (a more detailed discussion of homoplasy is provided in the ensuing section). The multiple bands generated from RAPD and AFLP analyses do not permit their designation as allelic or orthologous bands until they are converted into STS markers after sequencing. The co-dominant nature of SSRPs is suitable for genetical analysis in segregating $F_2$ populations or parentage analysis in hybrids (Scott *et al*., 2000 and Slavov *et al*., 2005).

The fourth advantage of the SSR marker system is their abundance and distribution in genomes. As more and more genomic sequences are being identified in various eukaryotic species, it is becoming increasingly evident that SSRs are truly abundant in almost all species, and are well distributed throughout their genomes (Wang *et al*., 1994; Varshney *et al*., 2005 and Tóth *et al*., 2000). Genetic analysis is often frustrated by the fact that large numbers of anonymous RAPD or AFLP markers are clustered in specific locations of chromosomes or linkage maps (Vuylsteke *et al*., 1999 and Kwon *et al*., 2006). In search of SSRs longer than 12 bp in a 57.8 Mb, publicly available rice (*Oryza sativa* L.) sequence, Temnykh *et al*. (2001) showed that many kinds of SSRs are present every 16 kb. In another survey of SSRs in different eukaryotic genomes (Tóth *et al*., 2000) reported that coding and non-coding regions differed significantly in SSR distribution and characteristic differences also existed between intergenic regions and introns in eukaryotes from yeasts to mammals to plants. Like the early findings in plants by Wang *et al*. (1994) tri- or hexa-nucleotide SSRs were predominantly present in coding regions, in the study by (Tóth *et al*., 2000).

A fifth advantage of the SSR marker system is that SSRs are preferentially associated with non-repetitive DNA (Varshney *et al*., 2005; Morgante *et al*., 2002 and Andersen and Liberstedt, 2003). Genomic sites of SSR markers, derived from genomic libraries, fall into either the transcribed region (genic SSRs) or the non-transcribed region (genomic SSRs). The SSRs, derived from ESTs or cDNAs, are mostly genic SSRs, which have the potential for application in such areas as gene function characterization (Ronning *et al*., 2003), association analysis for gene tagging (Szalma *et al*., 2005; Shin *et al*., 2006 and Crossa *et al*., 2007) and QTL analysis (Buerstmayr *et al*., 2002; Breseghello and Sorrels, 2006 and Zeng *et al*., 2009). However, this last advantage can only be applied to those species with large amounts of EST or cDNA sequences that are freely accessible to public.

**Tools for SSR/ microsatellite detection :**

*The tool simple sequence repeat identification tool* (*SSRIT*) (Table 1 and 2) :

(*http://www.gramene.org/db/searches/ssrtool*, Temnykh *et al*., 2001) uses Perl script to find perfect SSR repeats (2 to 10 bp in length) within a sequence. Kantety *et al*. (2001) used SSRIT to mine SSR in ESTs from barley, maize, rice, sorghum and wheat. Singh *et al*. (2011) used SSRIT to mine SSRs in wheat rust *Puccinia* sp.

*Another SSR identification tool is TROLL* :

(Tandem repeat occurrence locator, Castelo *et al*., 2002) which draws a keyword tree and matches it with a technique adapted from bibliographic searches, based on the *Aho-Corasick* algorithm. One of the major disadvantages of TROLL is that it cannot handle very large sequences and cannot process large batches of sequences as the tree takes up large amounts of memory.

*The microsatellite* (*MISA*) *tool* :

(*http://pgrc.ipk-gatersleben.de/misa/*) identifies

perfect, compound and interrupted SSRs. It requires a set of sequences in FASTA format and a parameter file that defines unit size and minimum repeat number of each SSR. The output includes a file containing the tables of repeat found, and a summary file. MISA can also design PCR amplification primers either side of SSR. The tool is written in Perl and is therefore platform independent, but it requires as installation of Primer3 for primer search (Thiel *et al*., 2003). MISA has been applied for SSR identification in coffee (Aggarwal *et al*., 2007), barley (Thiel *et al*., 2003 and Kota *et al*., 2001), wheat (Yu *et al*., 2004), rye (Khlestkina *et al*., 2004) and peanut (Liang *et al*., 2009).

*Another SSR search tool called as 'Repeat Finder'* (Volfovsky *et al*., 2001) (*http://www.cbcb.umd.edu/software/Repeat Finder/*) finds SSRs in four steps. The first step is to find all exact repeats using Repeat Match or REPuter. The second step merges repeats together into repeat classes and the third step includes merging all of the other repeats that match those already merged, into the same classes. Finally, step four matches all repeats and classes against each other in a non-exact manner using BLAST. The input is a genome or set of sequences, and the output is a file containing the repeat classes and number of merged repeats found in each class. Repeat Finder can finds repeats of any length. Also it finds perfect, imperfect and compound repeats and runs on Unix or Linux.

*SSRPrimer combines Sputnik and the PCR primer design software Primer3 to find SSRs and associated amplification primers* (*Robinson et al., 2004 and Jewell et al., 2006*) :

It takes multiple sequences in FASTA format as input and produce lists of SSRs and associated PCR primers in tabular format. SSRPrimer has been applied to a wide range of species including shrimp (Perez *et al*., 2005), citrus (Chen *et al*., 2006), mint (Lindqvist *et al*., 2006), strawberry (Keniry *et al*., 2006), *Brassica* (Batley *et al*., 2007; Burgess *et al*., 2006; Hopkins *et al*., 2007; Ling *et al*., 2007), *Sclerotinia* (Winton *et al*., 2007) and *Eragrostis curvula* (Cervigni *et al*., 2008).

*Maia et al.* (*2008*) *came with an interesting tool for SSR discovery integrated with primer design and PCR simulation called SSR Locator (http://www.ufpel.edu.br/*).

SSR locator detects SSR and minisatellite motifs between 1 and 10 bp, design primer for each locus found, amplify fragments with different primer pairs from a given set of FASTA files, produce global alignment between amplicons generated by the same primer pair and estimates alignment scores and identities between amplicons thus generating information on primer specificity and redundancy

*CID* :

Cid is a tool based on the collection of input data and parameters that are defined by the user, for the external tools. Both processes are submitted to the model layer and after processing, the final results are presented through the interface to the user. It is freely available on *http://www.shrimp.ufscar.br/cid/index.php* and constitutes a simple tool of easy manipulation that responds in a very efficient manner to its proposed use: processing cloned sequences, identifying microsatellites and automating the establishment of flanking primer pairs.

*SAT* :

SAT searches for SSRs and keeps the sequences containing an SSR motif. Users may choose between two different programmes to perform the SSR search: the SSRIT programme (Amador *et al*., 2004), which allows the choice of the minimal number of repeats for each pattern of di, tri or tetra nucleotide; or the Sputnik software [9], which is offered as an alternative because it presents the advantage of reporting imperfect SSR.

*GMATo* :

The soft GMATo was written in Perl and Java language. Java was used for developing graphic interface. Perl was used to discover the microsatellite and perform statistical analyzing. In GMATo DNA sequences are formatted first and the long DNA sequence is chunked to small segments at several Mb for easy processing. All microsatellite motifs consisting of A, T, G and C nucleotide of DNA at user controlled length are generated using Perl met characters and regular expression pattern. All motifs are searched greedily through each DNA chunk using Perl powerful pattern matching function. The returned values are used to generate SSR loci information at each chunk and the final SSR loci data at a chromosome after merging data from chunks. This method allows microsatellite discover efficiently in any genome with any size theoretically. Statistical

classification and summarization were performed at four levels *i.e.* motif length, motif composition, grouped complementary motifs and chromosome/ scaffold.

*SciRoKo* :

The SciRoKo SSR-search module is based on a scoring system, which considers the length of a microsatellite. Since a previous study identified this characteristic as to be the most informative variable describing microsatellites (Dieringer and Schlötterer, 2003). Five search modes are available, three for perfect and two for mismatched SSR search.

*G-IMEx* :

The algorithmic details of IMEx have been reported elsewhere (Crossa *et al.*, 2007). For the sake of continuity we reiterate the method. IMEx scans the input sequence and looks for two consecutive exact repeat units or two alternate exact repeat units and considers them as the 'candidate' microsatellite repeat tract. The

**Table 1 : Tools for SSR/ microsatellite detection**

| Sr. No. | Programmes/ tools | Website |
|---|---|---|
| 1. | Microsatellite (MISA) | *http://pgrc.ipk-gatersleben.de/misa/* |
| 2. | SSR locator | *http://www.ufpel.edu.br* |
| 3. | CID | *http://www.shrimp.ufscar.br/cid/index.php.* |
| 4. | SAT | *http://sat.cirad.fr/sat* |
| 5. | Sputnik | *http://espressosoftware.com/pages/sputnik.jsp* |
| 6. | Tandem repeat occurrence locator (TROLL) | *http://wsmartins.net/webtroll/troll.html* |
| 7. | SSR identification tool (SSRIT) | *http://www.gramene.org/db/searches/ssrtool* |
| 8. | SSR poly | *http://acpfg.imb.uq.edu.au/ssrpoly.php* |
| 9. | Tandem repeat finder (TRF) | *http://tandem.bu.edu/trf/trf.html* |
| 10. | Repeat finder | *http://www.cbcb.umd.edu/software/RepeatFinder/* |
| 11. | SciRoKo | *www.kofler.or.at/Bioinformatics* |
| 12. | GMATo | *http://sourceforge.net/projects/gmato/files/?source=navbar* |
| 13. | G-IMEx | *http://www.cdfd.org.in/imex* |

**Table 2 : Databases for SSR/microsatellite**

| Sr. No. | Database name | Organism name | Web link |
|---|---|---|---|
| 1. | MMDBJ | Mouse | *http://www.shigen.nig.ac.jp/mouse/mmdbj/top.jsp* |
| 2. | Satellyptus | Eucalyptus | *https://doaj.org/article/e03bfa80e0854b95a9fc729979d16be1* |
| 3. | SilkSatDb | Silkworm | *http://210.212.212.7:9999/PHP/SILKSAT/index.php* |
| 4. | InsatDb | Fruit-fly, honeybee, malarial mosquito, red-flour beetle and silkworm | *http://cdfd.org.in/INSATDB/home.php* |
| 5. | SNPSTR | Human, mouse, rat, dog and chicken | *http://www.sbg.bio.ic.ac.uk/~ino/SNPSTRdatabase.html* |
| 6. | TPMD | Human | *http://www.ncbi.nlm.nih.gov/pubmed/15608171* |
| 7. | CMD | Cotton | *http://www.mainlab.clemson.edu/cmd/* |
| 8. | FishMicrosat | Fishes | |
| 9. | Microsatellite database | Human | *http://www.microsatellites.org/db_search.php* |
| 10. | MICAS | Prokaryotes | *http://210.212.212.7/MIC/index.html* |
| 11. | Microsatellites Repeats(MRD) | Prokaryotes and Eukaryotes | *http://www.ccmb.res.in/mrd/* |
| 12. | Pipemicrodb | Pigeonpea | *http://cabindb.iasri.res.in/pigeonpea* |
| 13. | VMD | Viral genome | *http://www.mcr.org.in/vmd.* |
| 14. | CicArMiSatDB | Chickpea | *http://cicarmisatdb.icrisat.org/* |
| 15. | CmMDb | *Cucumis melo* L. | *http://65.181.125.102/cmmdb2/index.html* |
| 16. | EuMicroSatdb | Eukaryotes | *http://ipu.ac.in/usbt/EuMicroSatdb.htm* |

'candidate' tract is expanded on both sides by allowing few mismatches in each individual repeat unit ('k' - imperfection limit/ repeat unit) such that the percentage of imperfection of the entire tract does not cross the threshold set by the user. The expansion is also terminated if a repeat unit with more than 'k' mismatches is encountered. The program further collates and clusters equivalent microsatellite repeats into families. It also has an option to identify compound microsatellites, which are regions containing more than one microsatellite tract separated by a certain distance as defined by the user.

*Tandem repeats finder* :

Tandem Repeats Finder (TRF) (Benson, 1999) (*http://tandem.bu.edu/trf/trf.html*) can find very large SSR repeats, up to a length of 2000 bp. It uses a set of statistical tests for reporting SSRs, which is based on four distributions of pattern length, the matching probability, the indel probability and the tuple size. TRF finds perfect, imperfect and compound SSRs and is available for Linux. TRF has been used for SSR identification in cowpea (Chen *et al*., 2007).

*SSRPoly* :

SSRPoly (*http://acpfg.imb.uq.edu.au/ssrpoly.php*) is currently the only tool which is capable of identifying polymorphic SSRs from DNA sequence data. The input is a file of FASTA format sequences. SSRPoly includes a set of Perl scripts and MySQl tables than can be implemented on UNIX, Linux and window platforms.

Authors' affiliations :

**N.K. SINGH, V.K. SINGH, D.P. SINGH AND N.P. SINGH,** Krishi Vigyan Kendra (NDUAT), MAU (U.P.) INDIA

## REFERENCES

**Aggarwal, R.K.,** Hendre, P.S., Varshney, R.K., Bhat, P.R., Krishnakumar, V. and Singh, L. (2007). Identification, characterization and utilization of ESTderived genic microsatellite markers for genome analyses of coffee and related species. *Theor. Appl. Genet.*, **114** : 359-372.

**Amador, M.L.,** Oppenheimer, D., Perea, S., Maitra, A., Cusatis, G., Iacobuzio-Donahue, C., Baker, S.D., Ashfaq, R., Takimoto, C., Forastiere, A. and Hidalgo, M. (2004). An epidermal growth factor receptor intron polymorphism mediates response to epidermal growth factor receptor inhibitors. *Cancer Res*., **64** (24) : 9139 - 9143.

**Andersen, J. R.** and Liberstedt, T. (2003). Functional markers in plants. *Trends Plant Sci*., **8**: 554–560.

**Bacolla, A.** and Wells, R. D. (2009). Non-B DNA conformations as determinants of mutagenesis and human disease. *Mol. Carcinog*, **48** (4): 273-285.

**Batley, J.,** Hopkins, C.J., Cogan, N.O.I., Hand, M., Jewell, E., Kaur, J., Kaur, S., Li, X., Ling, A.E., Love, C., Mountford, H., Todorovic, M., Vardy, M., Walkiewicz, M., Spangenberg and Edwards, D. (2007). Identification and characterization of simple sequence repeat markers from *Brassica napus* expressed sequences. *Mol. Ecol. Notes*., **7** : 886-889.

**Bekessy, S. A.,** Ennos, R. A., Burgman, M. A., Newton, A. C. and Ades, P. K. (2003). Neutral DNA markers fail to detect genetic divergence in an ecologically important trait. *Biol. Conserv*., **110**: 267–275.

**Benson, G. (1999).** Tandem repeats finder: A programme to analyze DNA sequences. *Nucleic Acids Res*., **27**: 573–580.

**Boder, P.,** Deak, T., Bacso, R., Velich, I., Bisztray, G. D., Fascar, G. and Gyulai, P. (2006). Morphological and genetic investigation of medieval grape seeds. *Acta Hort*. (*ISHS*), 713–718.

**Breseghello, F.** and Sorrels, M. E. (2006). Association analysis as a strategy for improvement of quantitative traits in plants. *Crop Sci*., **46**: 1323–1330.

**Buerstmayr, H.,** Lemmens, M., Hartl, L., Doldi, L., Steiner, B., Stierschneider, M. and Ruckenbauer, P. (2002). Molecular mapping of QTLs for Fusarium head blight resistance in spring wheat. I. Resistance to fungal spread (Type II resistance). *Theor. Appl. Genet*., **104**: 84–91.

**Burgess, B.,** Mountford, H., Hopkins, C.J., Love, C., Ling, A.E., Spangenberg, G.C., Edwards, D. and Batley, J. (2006). Identification and characterization of simple sequence repeat (SSR) markers derived in silico from *Brassica oleracea* genome shotgun sequences. *Mol. Ecol. Notes*., 1191-1194.

**Castelo, A.T.,** Martins, W. and Gao, G.R. (2002). Troll- Tandem. *Bioinformatics*, **18**: 634-636.

**Cervigni, G.D.,** Paniego, N., Díaz, M., Selva, J.P., Zappacosta, D., Zanazzi, D., Landerreche, I., Martelotto, L., Felitti, S., Pessino, S., Spangenberg, G. and Echenique, V. (2008). Expressed sequence tag analysis and development of gene associated markers in a near-isogenic plant system of *Eragrostis curvula.* Plant. *Mol Biol.*, **67** : 1-10.

**Chen, C.X.,** Zhou, P., Choi, Y.A., Huang, S. and Gmitter, F.G. (2006). Mining and characterizing microsatellites from citrus ESTs. *Theor. Appl. Genet*., **112** : 1248-1257.

**Chen, X.,** Cho, Y. and Mc Couch, S. (2002). Sequence divergence of rice microsatellites in *Oryza* and other plant species. *Mol. Genet. Genomics*, **268** : 331–343.

**Chen, X.F.,** Laudeman, T.W., Rushton, P.J., Spraggins, T.A.

and Timko, M.P. (2007). CGKB: An annotation knowledge base for cowpea (*Vigna unguiculata* L.) methylation filtered genomic genespace sequences. *BMC Bioinf.,* **8**: 112-116.

**Coil, D. A.,** Vandersmissen, L., Ginevra, C., Jarraud, S., Lammertyn, E. and Anné, J. (2008). Intragenic tandem repeat variation between *Legionella pneumophila* strains. *BMC Microbiol*., **8**: 218.

**Crossa, J.,** Burgueno, J., Dreisigacker, S., Vargas, M., Herrera-Foessel, S.A., Lillemo, M., Singh, R.P., Trethowan, R., Warburton, M. and Franco, J. (2007). Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. *Genetics*, **177**: 1889–1913.

**Cruz, F.,** Pérez, M. and Presa, P. (2005). Distribution and abun dance of microsatellites in the genome of bivalves. *Gene.*, **346** : 241-247.

**Cummings, C. J.** and Zoghbi, H. Y. (2000). Fourteen and counting: Unraveling trinucleotide repeat diseases. *Hum. Mol. Genet*., **9**: 909–916.

**Dieringer, D.** and Schlotterer, C. (2003). Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. *Genome. Res.,* **13** (10): 2242-2251.

**Eckert, K. A.** and Hile, S. E. (2009). Every microsatellite is different: Intrinsic DNA features dictate mutagenesis of common microsatellites present in the human genome. *Mol. Carcinog*, **48**(4): 379-388.

**Ennos, R. A.** (1996). Utilizing genetic information in plant conservation programmes. In: *Aspects of the genesis and maintenance of biological diversity*; Hochberg, M.E., Clbert, J., Barbault, R., Eds.; Oxford University Press: Oxford, UK, pp. 278–291.

**Fitz, Simmons N. N.,** Moritz, C. and Moore, S. S. (1995). Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Mol. Biol. Evol*., **12**: 432–440.

**Gupta, P. K.,** Rustgi, S., Sharma, S., Singh, R., Kumar, N. and Balyan, H. S. (2003). Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol. Genet. Genomics*, **270**: 315–323.

**Hancock, J.M.** and Santibanez-Koref, M.F. (1998). Trinucleotide expansion diseases in the context of micro- and minisatellite evolution. *EMBO J*. **17**: 5521-5524.

**Hancock, J. M.,** Worthey, E.A. and Santibáñez-Koref, M.F. (2001). A role for selection in regulating the evolutionary emergence of disease-causing and other coding CAG repeats in humans and mice. *Mol. Biol. Evol*., **18** (6): 1014-1023.

**Holderegger, R.,** Kamm, U. and Gugerli, F. (2006). Adaptive vs. neutral genetic diversity: Implications for landscape genetics. *Landsci. Ecol*., **21**: 797–807.

**Hopkins, C.J.,** Cogan, N.O.I., Hand, M., Jewell, E., Kaur, J., Li, X., Lim, G.A.C., Ling, A., Love, C., Mountford, H., Todorovic, M., Vardy, M., Spangenberg, G.C., Edwards, D. and Batley, J. (2007). Sixteen new simple sequence repeat markers from *Brassica juncea* expressed sequences and their crossspecies amplification. *Mol. Ecol. Notes*., **7**: 697-700.

**Jacob, K.D.** and Eckert, K. A. (2007). *Escherichia coli* DNA polymerase IV contributes to spontaneous mutagenesis at coding sequences but not microsatellite alleles. *Mutat. Res*., **619** (1-2) : 93-103.

**Jewell, E.,** Robinson, A., Savage, D., Erwin, T., Love, C.G., Lim, G.A.C., Li, X., Batley, J., Spangenberg, G.C. and Edwards, D.( 2006) .SSR Primer and SSR Taxonomy Tree: Biome SSR discovery. *Nucleic Acids Res.*, **34** : 656- 659.

**Kantety, R.V.,** Rota, M.L., Matthews, D.E. and Sorrells, M.E. (2002). Data mining for simple sequence repeats in expressed sequence tags from barely, maize, rice, sorghum and wheat. *Plant Mol. Biol.*, **48**: 501-510.

**Kashi, Y.** and King, D.G. (2006). Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.*, **22**(5): 253-259.

**Kehrer-Sawatzki, H.** and Cooper, D. N. (2008). Molecular mechanisms of chromosomal rearrangement during primate evolution. *Chromosome Res*., **16**(1): 41-56.

**Keniry, A.,** Hopkins, C.J., Jewell, E., Morrison, B., Spangenberg, G.C., Edwards, D. and Batley, J. (2006). Identification and characterization of simple sequence repeat (SSR) markers from Fragaria x ananassa expressed sequences. *Mol. Ecol. Notes*., **6** : 319-322.

**Khlestkina, E.K.,** Than, M.H.M., Pestsova, E.G., Röder, M.S., Malyshev, S.V., Korzun, V. and Börner, A. (2004). Mapping of 99 new microsatellitederived loci in rye (*Secale cereale* L.) including 39 expressed sequence tags. *Theor. Appl. Genet.*, **109** : 725-732.

**Kostia, S.,** Varvio, S. L., Vakkari, P. and Pulkkinen, P. (1995). Microsatellite sequences in a conifer, *Pinus sylvestris*. *Genome*, **38** : 1244–1248.

**Kota, R.,** Varshney, R.K., Thiel, T., Dehmer, K.J. and Graner, A. (2001). Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). *Hereditas,* **135** : 145-151.

**Kwon, S.,** Hong, S., Son, J., Lee, J. K., Cha, Y., Eun, M. and Kim, N. (2006). CACTA and MITE transposon distributions on a genetic map of rice using F15 RILs derived from Milyang 23

and Gihobyeo hybrids. *Mol. Cells,* **21**: 360–366.

**Lee, J. R.,** Hong, G. Y., Dixit, A., Chung, J. W., Ma, K. H., Lee, J. H., Kang, H. K., Cho, Y. H., Gwag, J. G. and Park, Y. J. (2008). Characterization of microsatellite loci developed for *Amaranthus hypochondriacus* and their cross-amplifications in wild species. *Conserv. Genet*. **9**: 243–246.

**Liang, X.,** Chen, X., Hong, Y., Liu, H., Zhou, G., Li, S. and Guo, B. (2009). Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and Arachis wild species. *BMC Plant Biol.,* **9** : 35.

**Lindqvist, C.,** Scheen, A.C., Yoo, M.J., Grey, P., Oppenheimer, D.G., LeebensMack, J.H., Soltis, D.E., Soltis, P.S. and Albert, V.A. (2006). An expressed sequence tag (EST) library from developing fruits of an Hawaiian endemic mint (Stenogynerugosa, Lamiaceae): characterization and microsatellite markers. *BMC Plant Biol*., **6** : 16.

**Ling, A.E.,** Kaur, J., Burgess, B., Hand, M., Hopkins, C.J., Li, X., Love, C.G., Vardy, M., Walkiewicz, M., Spangenberg, G., Edwards, D. and Batley, J. (2007). Characterization of simple sequence repeat markers derived in silico from *Brassica rapa* bacterial artificial chromosome sequences and their application in Brassica napus. *Mol. Ecol. Notes.,***7**: 273-277.

**Liquori, C. L.,** Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., Day, J. W. and Ranum, L. P. W. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science,* **293**: 864–867.

**Li, W. H.** and Graur, D. (1991). *Fundamentals of molecular evolution*. Sinauer Pub.: Sunderland, MA, USA.

**Li, Y. C.,** Korol, A.B., Fahima, T., Beiles, A. and Nevo, E. (2002).Microsatellites:Genomic distribution, putative functions and mutational mechanisms: A review. *Mol. Ecol*., **11**: 2453–2465.

**Li, Y.C.,** Korol, A.B., Fahima, T. and Nevo, E. (2004). Microsatellites within genes: Structure, function, and evolution. *Mol. Biol. Evol*., **21**: 991–1007.

**Lukusa, T.** and Fryns, J. P. (2008). Human chromosome fragility. *Biochem. Biophys. Acta*., **1779** (1): 3-16.

**Ma, K.H.,** Kim, N.S., Lee, G.A., Lee, S.Y., Lee, J.K., Yi, J.Y., Park, Y.J., Kim, T.S., Gwag, J.G. and Kwon, S.J. (2009). Development of SSR markers for studies of diversity in common buckwheat. *Theor. Appl. Genet.,* **119**: 1247–1254.

**Maia, L.C.,** Palmieri, D.A., Souza, V.Q., Kopp, M.M., Carvalho, F.I. and Oliveira, A.C. (2008). SSR Locator: Tool for Simple Sequence Repeat Discovery Integrated with Primer Design and PCR Simulation. *Internat. J. Plant Genomics*,1–9. ( doi:10.1155/2008/412696).

**Manen, J.F.,** Bouby, L., Dalnoki, O., Marinval, P., Turgay, M. and Schlumbaum, A. (2003). Microsatellites from archaeological *Vitis vinifera* seeds allow a tentative assignment of the geographical origin of ancient cultivars. *J. Archaeol. Sci.,* **30**: 721–729.

**Marcotte, E.M.,** Pellegrini, M., Yeates, T.O. and Eisenberg, D. (1999). A census of protein repeats. *J. Mol. Biol.,* **293**: 151–160.

**Morgante, M.,** Hanafey, M. and Powell, W. (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat. Genet.,* **30**: 194–200.

**Morgante, M.,** Rafalski, A., Biddle, P., Tingey, S. and Olivieri, A.M. (1994). Genetic mapping and variability of seven soybean simple sequence repeat loci. *Genome,* **37**: 763–769.

**Mrázek, J.** (2006). Analysis of distribution indicates diverse functions of simple sequence repeats in Mycoplasma genomes. *Mol. Biol. Evol.,* **23**(7): 1370-1385.

**Peakall, R.,** Gilmore, S., Keys, W., Morgante, M. and Rafalski, A. (1998). Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: Implications for the transferability of SSRs in plants. *Mol. Biol. Evol.,* **15**: 1275–1287.

**Pérez, M.,** Cruz, F. and Presa, P. (2005). Distribution properties of poly mononucleotide repeat in molluscan genomes. *J. Hered.* **96** (1) : 40-51.

**Powell, W.,** Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S. and Rafalski, A. (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed*., **2**: 225–238.

**Rico, C.,** Rico, I. and Hewitt, G. (1996). 470 million years of conservation of microsatellite loci among fish species. *Proc. R. Soc. Lond., B., Biol. Sci.,* **263** : 549–557.

**Robinson, A.J.,** Love, C.G., Batley, J., Barker, G. and Edwards, D. (2004). Simple sequence repeat marker loci discovery using SSR primer. *Bioinfor,* **20** : 1475-1476.

**Ronning, C. M.,** Stegalkina, S.S., Ascenzi, R.A., Bougri, O., Hart, A.L., Utterbach, T. R., Vanaken, S. E., Riedmuller, S. B., White, J. A. and Cho, J. (2003). Comparative analyses of potato expressed sequence tag libraries. *Plant Physiol.,* **131**: 419–429.

**Scott, K. D.,** Eggler, P., Seaton, G., Rossetto, M., Ablett, E. M., Lee, L.S. and Henry, R. J. (2000). Analysis of SSRs derived from grape ESTs. *Theor. Appl. Genet.,* **100** : 723–726.

**Shin, J. H.,** Kwon, S. J., Lee, J. K., Min, H. K. and Kim, N.S. (2006). Genetic diversity of maize kernel starch-synthesis genes with SNAPs. *Genome,* **49** : 1287–1296.

**Singhal, D.,** Gupta, P., Sharma, P., Kashyap, N., Anand, S. and Sharma, H. (2011). In-silico single nucleotide polymorphisms (SNP) mining of sorghum bicolor genome. *African J. Biotechnol.,* **10** (4) : 580-583.

**Slavov, G. T.,** Howe, G. T., Gyaourova, A.V., Birkes, D. S. and Adams, W. T. (2005). Estimating pollen flow using SSR markers and paternity exclusion: Accounting for mistyping. *Mol. Ecol.,* **14** : 3109–3121.

**Sreenu, V. B.,** Kumar, P., Nagaraju, J. and Nagarajam, H. A. (2007). Simple sequence repeats in mycobacterial genomes. *J. Biosci.,* **32**(1): 3-15.

**Streelman, J.T.** and Kocher, T.D. (2002). Microsatellite variation associated with prolactin expression and growth of salt-challenged tilapia. *Physiol. Genomics,* **9**: 1–4.

**Szalma, S.J.,** Buckler, E.S., Snook, M.E. and McMullen, M.D. (2005). Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. *Theor. Appl. Genet.,* **110**: 1324–1333.

**Temnykh, S.,** DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S. and McCouch, S. (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.) : Frequency, length variation, transposon associations and genetic marker potential. *Genome Res.,* **11**: 1441–1452.

**Thiel, T.,** Michalek, W., Varshney, R.K. and Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.,***106** : 411-422.

**Tóth, G.,** Gáspári, Z. and Jurka, J. (2000). Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Res.,* **10**: 967–981.

**Van, Tienderen P. H.,** De Haan, A. A., Van der Linden, C. G. and Vosman, B. (2002). Biodiversity assessment using markers for ecologically important traits. *Trends Ecol. Evol.,* **17**: 577–582.

**Varshney, R.K.,** Graner, A. and Sorrells, M.E. (2005). Genic microsatellite markers in plants: Features and applications. *Trends Biotechnol.,* **23**: 48–55.

**Volfovsky, N.,** Haas, B.J. and Salzberg, S.L. (2001). A clustering method for repeat analysis in DNA sequences. *Genome Biol.,* **2** (8) RESEARCH0027 .

**Vuylsteke, M.,** Mank, R., Antonise, R., Bastiaans, E., Senior, M. L., Stuber, C. W., Melchinger, A. E., Lbberstedt, T., Xia, X. C. and Stam, P. (1999). Two high-density AFLP® linkage maps of *Zea mays* L.: Analysis of distribution of AFLP markers. *Theor. Appl. Genet.,* **99**: 921–935.

**Wang, Z.,** Weber, J. L., Zhong, G. and Tanksley, S. D. (1994). Survey of plant short tandem DNA repeats. *Theor. Appl. Genet.,* **88**: 1–6.

**White, G.** and Powell, W. (1997). Cross-species amplification of SSR loci in the Meliaceae family. *Mol. Ecol.,* **6**: 1195–1197.

**Winton, L.M.,** Krohn, A.L. and Leiner, R.H. (2007). Microsatellite markers for Sclerotiniasubarctica nom. prov., a new vegetable pathogen of the High North. *Mol. Ecol. Notes.,* **7**:1077-1079.

**Wren, J. D.,** Forgacs, E., Fondon, J. W., Pertsemlidis, A., Cheng, S. Y., Gallardo, T., Williams, R. S., Shohet, R. V., Minna, J. D. and Garner, H. R. (2000). Repeat polymorphisms within gene regions: Phenotypic and evolutionary implications. *Am. J. Hum. Genet.,* **67**: 345–356.

**Yamakoshi, K.,** Shishido, Y. and Shimoda, N. (2005). Generation of aberrant transcripts of and free DNA ends in zebra fish no tail gene. *Mar Biotechnol*., **7** (3) : 163-172.

**Yu, J.K.,** Dake, T.M., Singh, S., Benscher, D., Li, W., Gill, B.S. and Sorrells, M.E. (2004). Development and mapping of EST-derived simple sequence repeat (SSR) markers for hexaploid wheat. *Genome,* **47** : 805 - 818.

**Zeng, Y.,** Yang, S., Cui, H., Yang, X., Xu, L., Du, J., Pu, X., Li, Z., Cheng, Z. and Huang, X. (2009). QTLs of cold-related traits at the booting stage for NIL-RILsin rice revealed by SSR. *Genes Genom.,* **31**: 143–145.

**Zhang, X.,** Yue, B., Jiang, W. and Song, Z. (2009). The complete mitochondrial genome of rock carp *Procypris rabaudi* (Cypriniformes: Cyprinidae) and phylogenetic implications. *Mol. Biol. Rep.,* **36** (5) : 981-191.

## WEBLIOGRAPHY

*http://www.gramene.org/db/searches/ssrtool*, Temnykh *et al*., 2001).

(*http://pgrc.ipk-gatersleben.de/misa/*).

(*http://www.cbcb.umd.edu/software/Repeat Finder/*) .

*(http://www.ufpel. edu.br/*).

http://www.shrimp.ufscar.br/cid/index.php.

(*http://tandem.bu.edu/trf/trf.html*) .

(*http://acpfg.imb.uq.edu.au/ssrpoly. php*).

http://www.shrimp.ufscar.br/cid/index.php