

## RESEARCH

## Open Access



# PineElm\_SSRdb: a microsatellite marker database identified from genomic, chloroplast, mitochondrial and EST sequences of pineapple (*Ananas comosus* (L.) Merrill)

Sakshi Chaudhary<sup>†</sup>, Bharat Kumar Mishra, Thiruvettai Vivek, Santoshkumar Magadam and Jeshima Khan Yasin<sup>\*†</sup> 

## Abstract

**Background:** Simple Sequence Repeats or microsatellites are resourceful molecular genetic markers. There are only few reports of SSR identification and development in pineapple. Complete genome sequence of pineapple available in the public domain can be used to develop numerous novel SSRs. Therefore, an attempt was made to identify SSRs from genomic, chloroplast, mitochondrial and EST sequences of pineapple which will help in deciphering genetic makeup of its germplasm resources.

**Results:** A total of 359511 SSRs were identified in pineapple (356385 from genome sequence, 45 from chloroplast sequence, 249 in mitochondrial sequence and 2832 from EST sequences). The list of EST-SSR markers and their details are available in the database.

**Conclusions:** PineElm\_SSRdb is an open source database available for non-commercial academic purpose at <http://app.bioelm.com/> with a mapping tool which can develop circular maps of selected marker set. This database will be of immense use to breeders, researchers and graduates working on *Ananas* spp. and to others working on cross-species transferability of markers, investigating diversity, mapping and DNA fingerprinting.

**Keywords:** Ananas, Genome wide marker analysis, Organelle, Pineapple, Simple sequence repeats

## Background

The extremely surprising flavour and fragrance of pineapple (*Ananas comosus* L.) delighted mankind at that time of its discovery by Christopher Columbus and even today. Pineapple, a perennial monocot plant belongs to Bromeliales order, Bromelioideae subfamily and Bromeliaceae family. Pineapple is a tropical plant native to South America, domesticated more than 6000 years ago [1]. At the end of the sixteenth century, pineapple had become pantropical and is the third most economically important tropical fruit crop after banana and mango. Pineapple has become

industrial crop during 20<sup>th</sup> century [2,3]. In addition to fresh fruit consumption, pineapple is used for canned slices, juice and juice concentrate, extraction of bromelain (a meat-tenderizing enzyme), high-quality fibre, animal feed and medicines [2]. At present, gross production value of pineapple is approaching \$9 billion due to its cultivation on 1.02 million hectares of land in over 80 countries and annual production of 24.8 million metric tonnes of fruit [4]. Wild varieties of pineapple are self-compatible, whereas cultivated pineapple, *A. comosus* (L.) Merr., is self-incompatible [5], which provides an opportunity to scrutinize the molecular basis of self-incompatibility in monocots.

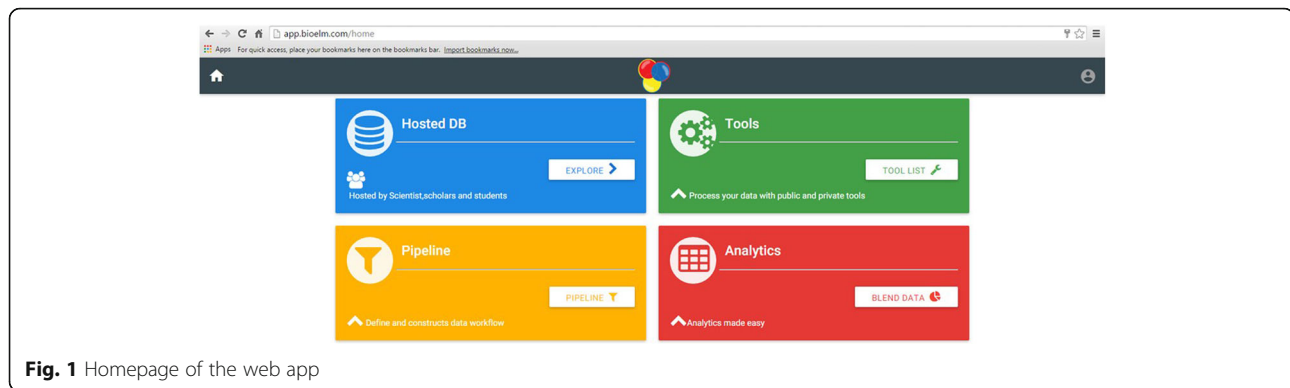
Over the last few decades, a wide range of molecular markers have been developed and used in crop improvement as molecular markers are helpful in assessing

\* Correspondence: [Yasin.Jeshima@icar.gov.in](mailto:Yasin.Jeshima@icar.gov.in); [yasinlab.icar@gmail.com](mailto:yasinlab.icar@gmail.com)

<sup>†</sup>Equal contributors

Division of Genomic Resources, ICAR- National Bureau of Plant Genomic Resources, PUSA campus, 110012 New Delhi, India





**Fig. 1** Homepage of the web app

germplasm diversity, testing of hybridity, trait mapping, marker assisted selection etc. [6]. Among all the markers till date, Simple Sequence Repeats (SSRs) are the most ideal, powerful and reliable markers for molecular plant breeding applications because of their high abundance, co-dominant inheritance and multiple alleles [7]. In addition, BES-SSR markers serve a useful resource for integrating genetic and physical maps [8,9].

SSRs consists of 2–7 base pair tandem repeat motifs of mono-, di-, tri-, tetra and penta-nucleotides (A, T, AT, GA, AGG, AAAG etc.) with different lengths of repeat motifs. These repeats are extensively distributed throughout plants and animal genomes. A high level of genetic variation is observed between and within species due to differences in the number of tandem repeating units at a locus which produces a highly polymorphic banding pattern [10] and is detected by the Polymerase Chain Reaction (PCR) using locus specific flanking primers [11]. Molecular markers are widely recognized as a tool in generating linkage maps [12] as they define specific locations in the genome unambiguously [13,14].

There are few valuable software and tools available for SSRs identification and *in-silico* marker development. Important sources for SSR identification are with benefits from the advanced next generation sequencing technology such as TROLL [15], MISA [16], SciRoko [17], SSR Locator [18] and GMATo [19]. MISA is the most common tool used for SSR identification. Generation of SSR markers have been exhaustive due to the time-

consumption, expensive process for generation of genomic libraries and sequencing of large number of clones later to find the SSR-containing DNA regions [20] and labour-intensive. To expedite this task, the traditional methods of SSR markers generation from genomic libraries [21] have been recouped briskly by *in-silico* mining of SSRs from DNA sequences available in biological databases [22,23] and from expressed sequence tags (ESTs) that represent only the coding region of the genome [24–26].

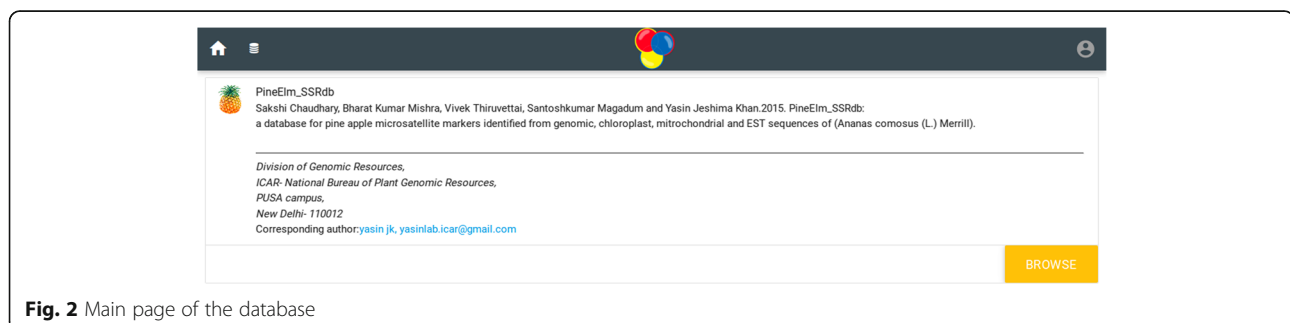
## Methods

### Retrieval of genome sequences

The complete genome sequence of pineapple (*Ananas comosus* (L.) Merrill) was retrieved from the CoGe Genome (Genome ID- 25734) page (<https://genomevolution.org/coge/GenomeInfo.pl?gid=25734>) in FASTA format. The chloroplast genome (Genome ID- 25280) and mitochondrial genome (Genome ID- 25281) of pineapple were also downloaded from CoGe Genome info respectively (<https://genomevolution.org/coge/GenomeInfo.pl?gid=25280&81>) in FASTA format. Total 5978 EST sequences of pineapple were downloaded from NCBI <http://www.ncbi.nlm.nih.gov/nucest/?term=ananas+comosus> in FASTA format.

### SSRs identification

MISA tool allows the identification and localization of perfect microsatellite as well as compound microsatellite



**Fig. 2** Main page of the database

SSR	Size	Start	End	SSR Type	Group
(CGC)7(CTC)4	33	392	424	Complex	EST_SSR
(CTG)5(TTG)9	42	824	865	Complex	EST_SSR
(TCT)7(GCT)4(GCG)4	45	82	126	Complex	EST_SSR
(CTCCA)3(CGCCA)3	30	20	49	Complex	EST_SSR
(TTC)6(TCC)6	36	40	75	Complex	EST_SSR
(TC)15(AC)6	42	76	117	Complex	EST_SSR
(CTG)5(TTG)9	42	809	850	Complex	EST_SSR
(GCT)7(GGT)7	42	512	553	Complex	EST_SSR
(GAG)6(GAC)5	33	56	88	Complex	EST_SSR

**Fig. 3** EST-SSRs list

which are interrupted by a certain number of bases. MISA uses Perl script for SSRs analysis. It requires a set of sequences in FASTA format and a parameter file that defines unit size and minimum repeat number of each SSR. MISA is available at <http://pgrc.ipk-gatersleben.de/misa/>. MISA tool provides two result files; misa file and statistical file. MISA file provides the information about SSR repeat types like simple, interrupted or compound, size of SSR and SSR position in genome sequence. Statistical file contains the statistical information like the frequency chart of SSR motif and distribution of SSR to differently repeat type classes. Classification of SSRs was done manually on the basis of their presence in coding region and non-coding region of the genome sequences.

**Database development**

An open, non-commercial database PineElm\_SSRdb is designed for educational purpose. PineElm\_SSRdb is available at <http://app.bioelm.com/>.

**Results and discussion**

The home page of the database provides the complete access of the database (Figs. 1 and 2). Genomic sequence of *Ananas comosus* (L.) Merrill is available as 3133 scaffolds of 381,905,120 bp length. Of these scaffolds only 2726 contain SSRs. From genomic sequence 356385 SSRs were identified. Two thousand four hundred eighty-six sequences contain more than one SSR and 19086 compound SSRs are also exists. From NCBI 5978 EST sequences of 4,294,909 bp were downloaded of which 1886 sequences yielded 2832 SSRs (Fig. 3). Of these 1886 sequences 638 were found to contain more than one SSR region and 83 with compound repeat regions. The access to the database can be obtained from <http://app.bioelm.com/> by creating user account. This will allow the users to generate images and search results. Genomic markers can be viewed as listed in Fig. 4 and as a circular map of the selected markers using incorporated tools (Additional file 1).

There is only one chloroplast genome sequence of 159,600 bp available for *Ananas comosus* (L.) Merrill in

SSR	Size	Start	End	SSR Type	Group
(AT)33	66	16026580	16026625	Di	Genomic Marker
(AT)33	66	2814402	2814467	Di	Genomic Marker
(AT)33	66	3352285	3352350	Di	Genomic Marker
(AT)33	66	6676138	6676203	Di	Genomic Marker
(AT)33	66	15027310	15027375	Di	Genomic Marker
(AT)33	66	1691070	1691135	Di	Genomic Marker
(AT)33	66	395743	395808	Di	Genomic Marker
(AT)33	66	4003158	4003223	Di	Genomic Marker
(AT)33	66	10108588	10108653	Di	Genomic Marker

**Fig. 4** Genomic SSR marker list

which there are 45 SSRs exists and only one complex repeat (Additional file 2). There are 13 sequences of 881,399 bp *Ananas comosus* (L.) Merrill mitochondrial genome yielded 249 SSRs of which 13 sequences contain more than one SSR and 8 SSRs were found in compound formation (Fig. 5 and Additional file 3). The EST-SSR statistics were represented in Additional file 4. SSRs identified from chloroplast and mitochondrial sequences are highly specific and unique to pineapple. These SSRs are not present in any other NCBI database and also not present in the nuclei genome of pineapple.

Databases support instant availability of curated data for individual users in facilitating further effective use of the generated data. In that path, we have developed a database to support pineapple breeders to effectively use SSR markers in their breeding program. These SSR markers or microsatellites are of 1–6 nucleotide tandem repeated motifs present in all prokaryotic and eukaryotic genomes [27]. Amid different classes of available molecular markers, SSR markers are effective for a variety of applications in plant genetics and breeding [28, 29].

Although being a commercially important plant, only few studies for SSR development were available for Pineapple. Wohrmann and Weising [30], identified 696 EST-SSR markers in pineapple; Feng et al. [31] developed genomic and EST-SSR library to identify 94 and 1110 SSRs loci respectively. Complete Pineapple genome [4] opens new direction to focus our research towards pineapple. In addition, bioinformatics tools also add-on prevailing methods by automating the assignment of SSR identification from existing DNA sequences. A recent study reported 320,207 SSRs in genomic and ESTs sequences of pineapple [32]. Whereas, we have identified 356385 SSRs from genomic sequences of pineapple which may play a major role in diversity analysis of genetic stocks. Diversity analyses of pineapple genetic stocks were reported earlier with few markers which were insufficient in to distinguish them. Developing fingerprints

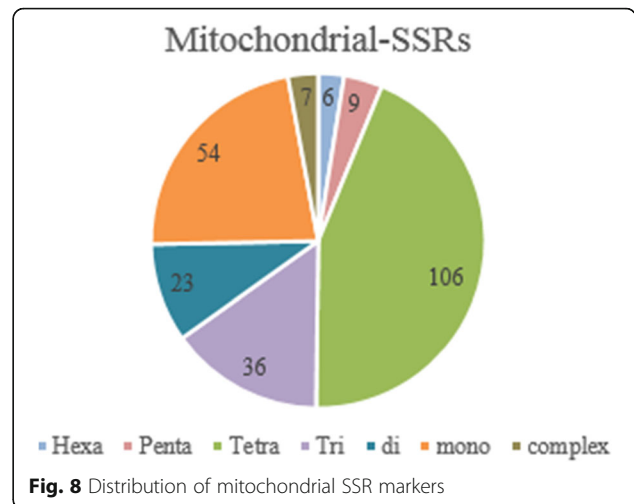
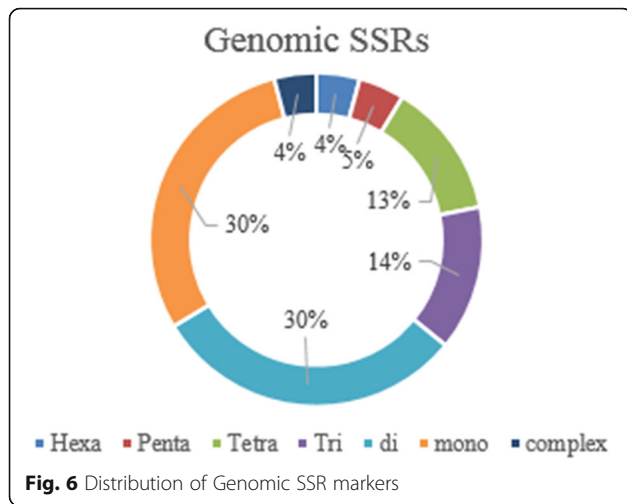
of cultivars may be required to protect the breeders right. Genome wide identification of markers can serve this purpose as SSR markers have been handy for integration the physical, genetic and sequence-based physical maps in plant species, and concurrently equipped breeders and geneticists with an effective tool to bridge phenotypic and genotypic variation. SSR markers have been handy for integration the physical, genetic and sequence-based physical maps in plant species, and concurrently equipped breeders and geneticists with an effective tool to bridge phenotypic and genotypic variation [33].

SSR markers were classified based on number of repeats (Figs. 6, 7 and 8). The proportion of mono and di repeats are likely to be equal for genomic SSRs contributing to the total 60% of genomic SSR markers. Likewise, hexa and complex repeats are approximately equal at 4% of genomic SSRs (Fig. 6). Recent studies of plant markers are more focused towards gene-specific markers rather than arbitrary DNA markers, and microsatellite markers are of great significance in identification of genes and QTLs [34]. EST-SSRs are highly efficient in differentiating genotypes differing for a specific trait. We have identified 2832 SSR markers from the validated EST sequences, where tri repeats followed by di repeats contributes more to the markers (Fig. 7). Tetra and mono repeats contribute maximum number of mitochondrial – SSRs (Fig. 8). In all sets of data analysed, we could find least number of complex repeats and only one complex SSR exists in chloroplast genome of pineapple.

Molecular basis of polymorphism and their distribution across the genome is quite different for SNP and SSR markers. Both SSR and SNP are neutral, multi-allelic and co-dominant markers. SSR marker in genetic diversity analyses have been a powerful, handy, cost effective tool and can reveal the amplicon size polymorphism as they vary in sequence, whereas SNP haplotypes vary within a sequence. SNP markers display

SSR	Size	Start	End	SSR Type	Group
(AT)7	14	13752	13765	Di	Mitochondria_Ananas_misa
(TC)6	12	3436	3447	Di	Mitochondria_Ananas_misa
(TA)6	12	14325	14336	Di	Mitochondria_Ananas_misa
(AT)10	20	3543	3562	Di	Mitochondria_Ananas_misa
(AG)7	14	33533	33546	Di	Mitochondria_Ananas_misa
(AG)6	12	130606	130617	Di	Mitochondria_Ananas_misa
(AT)7	14	146641	146654	Di	Mitochondria_Ananas_misa
(TA)7	14	155398	155411	Di	Mitochondria_Ananas_misa
(AT)6	12	194554	194565	Di	Mitochondria_Ananas_misa

**Fig. 5** Mitochondrial SSR marker list



population structure better with bigger population whereas, for diversity analyses, SSR unveils better grouping of accessions even at trait level. Further, it has been demonstrated that haplotypes at combinations of SSR loci may be very powerful in detecting association of QTLs (Quantitative Trait Loci) in their proximity [35]. Henceforth, the utility of SSR/SNP marker in crop improvement will depend on the quality of information required with respect to parameters for genetic diversity and population structure. Overall, to assess genetic relatedness, SSR markers are more informative and highly effective [36].

**Conclusion**

The main outcome of this study; identified SSRs markers in genomic, chloroplast, mitochondrial and EST sequences of Pineapple will be of immense use to breeders and molecular biologists to assess marker frequency and distribution in both coding and non-coding regions, to study transferability across genera

and to carry out phylogenetic analysis based on SSRs. PineElm\_SSRdb is an open source database developed for easy handling and availability for the scientific community.

**Additional files**

- Additional file 1:** Genomic SSR markers of *Ananas comosus* (L.) Merrill. (TXT 5249 kb)
- Additional file 2:** Chloroplast SSR markers of *Ananas comosus* (L.) Merrill. (TXT 1 kb)
- Additional file 3:** Mitochondrial SSR markers of *Ananas comosus* (L.) Merrill. (TXT 7 kb)
- Additional file 4:** EST SSR markers of *Ananas comosus* (L.) Merrill. (TXT 42 kb)

**Acknowledgments**

We appreciate Mr. Sakubar Satik, founder of ArivElm for his help in database preparation and for providing open access to the data.

**Funding**

This work was supported through in-house grants of Indian Council of Agricultural Research-National Bureau of Plant Genetic Resources.

**Availability of data and materials**

PineElm\_SSRdb is an open source database available for non-commercial academic purpose at <http://app.bioelm.com/>.

**Authors' contributions**

SC and YJK - conceived and designed the research; SC and BKM- conducted identification; VT- contributed to functional annotation; YJK, SM, SC, VT and BKM wrote, read, reviewed and approved the manuscript.

**Authors' information**

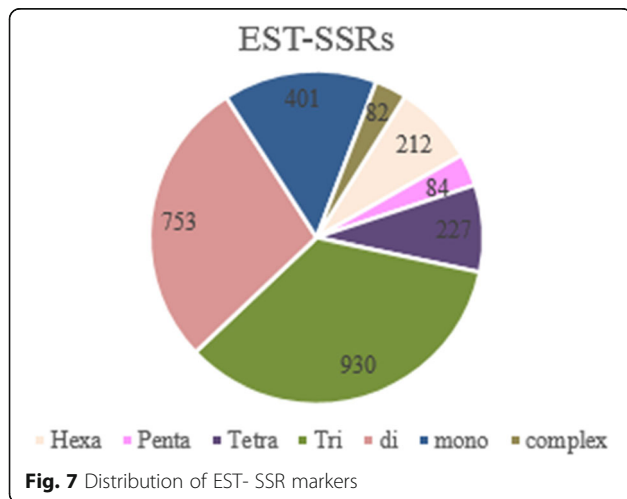
SC is a Ph. D scholar, BKM is a Junior Research Fellow, VT is a post graduate, Dr. SM is working as Research Associate and Dr. YJK is a scientist at Division of Genomic Resources, ICAR-National Bureau of Plant Genetic Resources, Pusa Campus, New Delhi, India.

**Competing interests**

The authors declare that they have no competing interests.

**Consent for publication**

Not applicable.





**Ethics approval and consent to participate**

Not applicable.

Received: 11 August 2016 Accepted: 2 November 2016

Published online: 24 November 2016

**References**

- Clement CR, de Cristo-Araújo M, Coppens D'Eeckenbrugge G, Alves Pereira A, Picanço Rodrigues D. Origin and domestication of native Amazonian crops. *Diversity*. 2010;2:72–106.
- Bartholomew DP, Paull RE, Rohrbach KG. *The Pineapple: Botany, Production and Uses*. UK: CAB; 2002.
- Beauman F. *The Pineapple: King of Fruits*. UK: Random House; 2006.
- Ming R, VanBuren R, Wai CM, et al. The pineapple genome and the evolution of CAM photosynthesis. *Nat Genet*. 2015;47:1435–42.
- Brewbaker JL, Gorrez DD. Genetics of self-incompatibility in the monocot genera, *Ananas* (pineapple) and *Gasteria*. *Am J Bot*. 1967;54:611–6.
- Jones N, Ougham H, Thomas H, Pasakinskiene I. Markers and mapping revisited: finding your gene. *New Phytol*. 2009;183:935–66.
- Dutta S, Mahato AK, Sharma P, Raje RS, Sharma TR, Singh NK. Highly variable 'Arhar' simple sequence repeat markers for molecular diversity and phylogenetic studies in pigeonpea [*Cajanus cajan* (L.) Millsp.]. *Plant Breed*. 2013;132:191–6.
- Mun JH, Kim DJ, Choi HK, Gish J, Debelle F, Mudge J, Denny R, Endre G, Saurat O, Dudez AM, Kiss GB, Roe B, Young ND, Cook D. Distribution of microsatellites in the genome of *Medicago truncatula*: a resource of genetic markers that integrate genetic and physical maps. *Genetics*. 2006;172:2541–55.
- Shultz JL, Samreen K, Rabia B, Jawaad AA, Lightfoot DA. The development of BAC-end sequence-based microsatellite markers and placement in the physical and genetic maps of soybean. *Theor Appl Genet*. 2007;114:1081–90.
- Farooq S, Azam F. Molecular markers in plant breeding–1: concepts and characterization. *Pak J Biol Sci*. 2002;5(10):1135–40.
- Jonah PM, Bello LL, Lucky O, Midau A, Moruppa SM. The importance of molecular markers in plant breeding programmes. *Global J Sci Front Res C Biol Sci*. 2011;11(5):5–12.
- Walunjkar B, Parihar A, Chaurasia P, Pachchigar K, Chauhan RM. Genetic analysis of wild and cultivated germplasm of pigeonpea using random amplified polymorphic DNA (RAPD) and simple sequence repeats (SSR) markers. *Afr J Biotechnol*. 2013;12(40):5823–32.
- Akkaya MS, Shoemaker RC, Specht JE, Bhagwat AA, Cregan PB. Integration of simple sequence repeat DNA markers into a soybean linkage map. *Crop Sci*. 1995;35:1439–45.
- Bell CJ, Ecker JR. Assignments of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics*. 1994;19:137–44.
- Castelo AT, Martins W, Gao GR. TROLL - tandem repeat occurrence locator. *Bioinformatics*. 2002;18(4):634–6.
- Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet*. 2003;106(3):411–22.
- Kofler R, Schlotterer C, Lelley T. SciRoKo: a new tool for whole genome microsatellite search and investigation. *Bioinformatics*. 2007;23:1683.
- Da-Maia LC, Palmieri DA, de Souza VQ, Kopp MM, de Carvalho FI, Costa de Oliveira A (2008) SSR Locator: Tool for simple sequence repeat discovery integrated with primer design and PCR simulation. *Int J Plant Genomics*. doi: 10.1155/2008/412696
- Wang X, Lu P, Luo Z. GMATo: A novel tool for the identification and analysis of microsatellites in large genomes. *Bioinformation*. 2013;9(10):541–4.
- Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC, Mian M. *Medicago truncatula* EST-SSRs reveal cross species genetic markers for *Medicago* spp. *Theor Appl Genet*. 2004;108:414–22.
- Weising K, Nybom H, Wolff K, Kahl G. DNA fingerprinting in plants: principles, methods, and applications. *Ann Bot*. 2006;97:476–7.
- Shanker A. Computationally mined microsatellites in chloroplast genome of *Pellia endiviifolia*. *Arch Bryol*. 2014;199:1–5.
- Shanker A, Singh A, Sharma V. In silico mining in expressed sequences of *Neurospora crassa* for identification and abundance of microsatellites. *Microbiol Res*. 2007;162:250–6.
- Kantety RV, La Rota M, Matthews DE, Sorrells ME. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol*. 2002;48:501–10.
- Scotti I, Magni F, Fink R, Powell W, Binelli G, Hedley PE. Microsatellite repeats are not randomly distributed within Norway spruce (*Picea abies* K.) expressed sequences. *Genome*. 2000;43:41–6.
- Varshney RK, Thiel T, Stein N, Langridge P, Graner A. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell Mol Biol Lett*. 2002;7:537–46.
- Zane L, Bargelloni L, Patarnello T. Strategies for microsatellite isolation: a review. *Mol Ecol*. 2002;11:1–16.
- Liu ZJ, Cordes JF. DNA marker technologies and their applications in aquaculture genetics. *Aquaculture*. 2004;238:1–37.
- Powell W, Machray GC, Provan J. Polymorphism revealed by simple sequence repeats. *Trends Plant Sci*. 1996;1:215–22.
- Wohrmann T, Weising K. In silico mining for simple sequence repeat loci in a pineapple expressed sequence tag database and cross-species amplification of EST-SSR markers across Bromeliaceae. *Theor Appl Genet*. 2011;123:635–47.
- Feng S, Helin T, Chen Y, et al. (2013) Development of Pineapple Microsatellite Markers and Germplasm Genetic Diversity Analysis. *BioMed Research International* 317912. doi:10.1155/2013/317912
- Fang J, Miao C, Chen R, Ming R (2016) Genome-Wide Comparative Analysis of Microsatellites in Pineapple. *Tropical Plant Biology* doi: 10.1007/s12042-016-9163-6
- Gupta PK, Varshney RK. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica*. 2000;113:163–85.
- Zhao Y, Williams R, Prakash CS, He G. Identification and characterization of gene-based SSR markers in date palm (*Phoenix dactylifera* L.). *BMC Plant Biol*. 2012;12:237.
- Koch HG, McClay J, Loh EW, Higuchi S, Zhao JH, Sham P, Ball D, Craig IW. Allele association studies with SSR and SNP markers at known physical distances within a 1 Mb region embracing the *ALDH2* locus in the Japanese, demonstrates linkage disequilibrium extending up to 400 kb. *Hum Mol Genet*. 2000;9:2993–9.
- Lapitan VC, Brar DS, Abe T, Redona ED. Assessment of genetic diversity of Philippine rice carrying good quality traits using SSR markers. *Breed Sci*. 2007;57:263–70.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
www.biomedcentral.com/submit

