



## OPEN ACCESS

## EDITED BY

Yi-Hong Wang,  
University of Louisiana at Lafayette,  
United States

## REVIEWED BY

Yupeng Pan,  
Northwest A&F University, China  
Bin Bai,  
Gansu Academy of Agricultural Sciences  
(CAAS), China

## \*CORRESPONDENCE

Siddanna Savadi  
✉ siddannasavadi@gmail.com

RECEIVED 18 June 2023

ACCEPTED 31 July 2023

PUBLISHED 21 August 2023

## CITATION

Savadi S, Muralidhara BM,  
Venkataravanappa V and Adiga JD (2023)  
Genome-wide survey and characterization  
of microsatellites in cashew and design  
of a web-based microsatellite  
database: CMDB.  
*Front. Plant Sci.* 14:1242025.  
doi: 10.3389/fpls.2023.1242025

## COPYRIGHT

© 2023 Savadi, Muralidhara,  
Venkataravanappa and Adiga. This is an  
open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that  
the original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Genome-wide survey and characterization of microsatellites in cashew and design of a web-based microsatellite database: CMDB

Siddanna Savadi<sup>1\*</sup>, B. M. Muralidhara<sup>2</sup>, V. Venkataravanappa<sup>2</sup>  
and J. D. Adiga<sup>1</sup>

<sup>1</sup>ICAR- Directorate of Cashew Research (DCR), Puttur, Karnataka, India, <sup>2</sup>ICAR-Indian Institute of Horticultural Research (IIHR), CHES, Madikeri, Karnataka, India

The cashew is an edible tree nut crop having a wide range of food and industrial applications. Despite great economic importance, the genome-wide characterization of microsatellites [simple sequence repeats (SSRs)] in cashew is lacking. In this study, we carried out the first comprehensive genome-wide microsatellites/SSRs characterization in cashew and developed polymorphic markers and a web-based microsatellite database. A total of 54526 SSRs were discovered in the cashew genome, with a mean frequency of 153 SSRs/Mb. Among the mined genome-wide SSRs (2-6 bp size motifs), the dinucleotide repeat motifs were dominant (68.98%) followed by the trinucleotides (24.56%). The Class I type of SSRs ( $\geq 20$  bp) were 45.10%, while Class II repeat motifs ( $\geq 12$ – $< 20$  bp) were 54.89% of the total genomic SSRs discovered here. Further, the AT-rich SSRs occurred more frequently in the cashew genome (84%) compared to the GC-rich SSRs. The validation of the *in silico*-mined genome-wide SSRs by PCR screening in cashew genotypes resulted in the development of 59 polymorphic SSR markers, and the polymorphism information content (PIC) of the polymorphic SSR markers ranged from 0.19 to 0.84. Further, a web-based database, “Cashew Microsatellite Database (CMDB),” was constructed to provide access to the genome-wide SSRs mined in this study as well as transcriptome-based SSRs from our previous study to the research community through a user-friendly searchable interface. Besides, CMDB provides information on experimentally validated SSRs. CMDB permits the retrieval of SSR markers information with the customized search options. Altogether, the genome-wide SSRs characterization, the polymorphic markers and CMDB database developed in this study would serve as valuable marker resources for DNA fingerprinting, germplasm characterization, genetic studies, and molecular breeding in cashew and related *Anacardium* species.

## KEYWORDS

microsatellite, genome wide, cashew nut, database, validation

## 1 Introduction

The cashew (*Anacardium occidentale* L.) is an edible tree nut crop grown in more than 30 countries in the tropical and subtropical regions (Sharma et al., 2020). Cashew is an evergreen tree with one or more vegetative and reproductive flushings occurring in an annual cycle (Adiga et al., 2019). It is a diploid species ( $2n = 42$ ) with an andromonoecy breeding system (Aliyu and Awopetu, 2007; Savadi et al., 2020a). It has its origins in North-Eastern Brazil and it was spread to different parts of the world, mainly by the Portuguese during the 16<sup>th</sup> century as the soil conservation tree (Samal et al., 2003; dos Santos et al., 2019).

Over time, cashew has become an economically important horticulture crop in many developing countries earning huge foreign exchange. Presently, the global raw cashew nut production is over 3.8 million metric tons, and the value of the global cashew nut market is worth nearly US\$ 6 billion (INC Statistical Yearbook, 2020–2021). Cashew nut kernels, cashew nut shell liquid (CNSL), kernel oil, and cashew apple (hypocarp or pseudofruit) are utilized in food and numerous other industries (Samal et al., 2003; Talbiersky et al., 2009; Das and Arora, 2017; Sharma et al., 2020). Kernels are used as a dessert nut and in confectioneries such as chocolates, cashew milk, cashew butter, etc. The kernel oil has excellent cooking quality and also widely used in the cosmetic industry due to its nourishing and nurturing effects on skin (Visalakshi et al., 2011). The cashew apple is also edible and is consumed raw or in processed form viz., jam, jelly, syrup, ready-to-serve juice (Akinwale, 2000; Sharma et al., 2020). The cashew apples used to prepare fermented alcoholic beverages (Das and Arora, 2017). CNSL, a reddish-brown oil present in a cashew nutshell, is a multipurpose byproduct of the cashew industry widely used in varnishes, lubricants, synthetic resins, molding compositions, and insulating coatings (Talbiersky et al., 2009; Furtado et al., 2019). Recently, anacardic acid, cardanol, and cardol, the major constituents of CNSL, have acquired great importance in the pharmaceutical industry as they have anticancer and many other great medicinal properties (Ashraf and Rathinasamy, 2018; Shi et al., 2019). Thus, the demand for cashew nuts and their products is increasing throughout the world.

Despite its great economic importance and demand, the genetic improvement of cashew using molecular breeding tools has lagged behind the other important fruit and nut crops due to the limited or lack of genetic and genomic investigations (Savadi et al., 2020a). Molecular markers are important molecular breeding tools extensively used in plant genetics and breeding (Collard and Mackill, 2008). Over the course of time, different types of molecular markers have been developed, which fall into mainly the dominant and co-dominant classes. Co-dominant markers such as Simple Sequence Repeats (SSRs)/Microsatellites and Single Nucleotide Polymorphisms (SNPs) are considered to be more informative and produce consistent results compared to the dominant markers such as RAPD, ISSR and/or AFLP and as a result, the dominant markers are becoming obsolete (Collard and Mackill, 2008; Grover and Sharma, 2016).

To date, in the majority of the genetic studies in cashew, the first generation or dominant markers, viz., RAPD, ISSR, and AFLP markers, have been used (Mnoney et al., 2001; Dhanaraj et al., 2003; Archak et al., 2003a; Archak et al., 2003b; Archak et al., 2009; Thimmappaiah et al., 2009; Aliyu, 2012; Jena et al., 2016; Borges et al., 2018; dos Santos

et al., 2019; da Costa Gomes et al., 2021) due to the limited availability of co-dominant markers viz., SSR, SNP and InDel (Insertion/Deletion) markers (Croxford et al., 2006; Savadi et al., 2022a; Savadi et al., 2023). Among the co-dominant markers, microsatellite or SSR markers have gained wide popularity and become markers of choice for genetic studies because of their multi-allelism, high abundance in the genome, ease of use, and amenability for high-throughput analysis (Taheri et al., 2018; Savadi et al., 2020b). Presently, only 21 genomic SSR (Croxford et al., 2006) and 36 genic SSR (Savadi et al., 2022a) markers are available in cashew, which is extremely low to represent the entire genome and does not meet the needs of comprehensive genetic research in cashew.

Previously, the development of microsatellite or SSR markers was an expensive and time-consuming process. But the rapid improvements in sequencing technologies and bioinformatics have reduced the cost and time required for the development of a large set of robust markers through genome sequencing and in silico mining of the genome sequences for potential markers (Luo et al., 2021; Wang et al., 2022; Savadi et al., 2023). To date, there have been no studies on the genome-wide characterization of microsatellites or SSRs in cashew. Further, the large set of SSRs mined from the genomic sequences cannot be efficiently utilized by the researchers without a user-friendly analytical search tool. In numerous crops such as *Cucumis melo* (CmMdb: Chaduvula et al., 2015), Sugar beet (SBMDb: Iquebal et al., 2015), Sesame (SisatBase: Dossa et al., 2017), and *Anemone* sp. (Martina et al., 2022), genome-wide SSRs have been discovered and web-based databases are designed for storage and easy accessibility of the large set of genome-wide SSRs to researchers. Recently, the first draft genome of cashew cv. Bhaskara (356 Mb size with 92% BUSCO value) was generated through hybrid assembly of Illumina mate-pair reads and Oxford Nanopore reads and reported by our group (Savadi et al., 2022b). The availability of draft genome sequence prompts the discovery of a large set of SSRs at the genome level and make them easily accessible to researchers.

In the present study, we carried out the comprehensive characterization of genome-wide microsatellites/SSRs for the first time in cashew and designed a user-friendly web-based microsatellite database for easy availability of genome-wide SSR information to cashew researchers. The possibility of *in silico*-discovered SSRs to detect polymorphism in cashew genotypes and cross-amplify in related *Anacardium* species was validated by PCR amplification and fragment separation using a subset of mined SSRs. Thus, microsatellites/SSRs resources generated here would be useful for accelerating genetic studies and crop improvement in cashew and related species.

## 2 Material and methods

### 2.1 Plant material and DNA extraction

In this study, 32 cashew genotypes (Table 1) and two *Anacardium* species, *A. microcarpum* and *A. othonianum*, were used for the validation of polymorphism and cross species amplification of SSRs, respectively. Leaves were harvested from

the field-grown plants, and genomic DNA was extracted following the Inglis et al. (2018) method. Initially, finely ground leaf tissues were pre-washed twice with the Sorbitol wash buffer [100 mM tris hydrochloride (Tris-HCl) pH 8.0, 0.35 M Sorbitol, 5 mM ethylenediaminetetraacetic acid (EDTA) pH 8.0, 1% (w/v) polyvinylpyrrolidone (PVP-40)] with 2-Mercaptoethanol (1% v/v) to remove the excessive phenolics. The prewashed samples were then used for DNA extraction with cetyl trimethylammonium bromide (CTAB) buffer [100 mM Tris-HCl pH 8.0, 3 M NaCl, 3% CTAB, 20 mM EDTA, and 1% (w/v) PVP-40]. The integrity of extracted DNA was checked by electrophoresis on a 1% agarose gel containing 0.1 mg/ml of ethidium bromide and quantified using a NanoPhotometer N60 (Implen, Munich, Germany).

## 2.2 Discovery, characterization, and validation of genome-wide SSRs

The whole-genome sequence (356 MB size, 92% BUSCO value) of cashew cv. Bhaskara generated by hybrid assembly of Illumina mate-pair reads and Oxford Nanopore reads at ICAR-DCR (Savadi et al., 2022b) and deposited in the National Centre for Biotechnology Information (NCBI) database (PRJNA766521) was used for the mining of genome-wide microsatellites/SSRs. The PolyMorphPredict software, which permits *in silico* mining of microsatellites and designs primers from genome and transcriptome data, was used to discover and characterize the composition and distributions of genome-wide SSRs (Das et al., 2019). PolyMorphPredict takes input sequence data in Fasta format, mines the SSRs with the MISA tool, and designs primers for the MISA mined SSRs using the Primer3 software with default parameters. The draft genome sequence was mined for SSRs with a minimum repeat number of 6, 5, 5, 4, 3 for di, tri, tetra, penta, and hexanucleotide SSRs, respectively, and a maximum difference of 100 bp between two SSRs. Further, a total of 100 primer pairs were synthesized from Eurofins Genomics, Bengaluru, India, for validation of *in silico* mined SSRs by polymerase chain reaction (PCR) amplification.

## 2.3 PCR amplification of SSRs

The annealing temperatures ( $T_a$ ) of the synthesized SSR primers were optimized using gradient PCR. The PCR reaction was performed in 15  $\mu$ l reaction mixtures containing 7.5  $\mu$ l EmeraldAmp<sup>®</sup> GT PCR Master Mix (Takara Bio Inc., Japan), 20 pM each of the forward and reverse primers, and 100 ng of genomic DNA in the Veriti<sup>™</sup> 96-Well Thermal Cycler (ThermoFisher Scientific USA) and volume makeup was done with Millipore water. The thermal profile conditions used for PCR amplification of SSRs included an initial denaturation step of 3 min at 95°C followed by 35 cycles of 40 s at 95°C, 40 s at primer-specific  $T_a$ , 45 s at 72°C and finally, 8 min at 72°C. The PCR products were resolved along with a 100 bp DNA ladder on 3.5% agarose gels containing 0.1 mg/ml of ethidium bromide in 1X TBE (Tris/borate/EDTA)

TABLE 1 List of genotypes used in this study with important characteristics.

Sl. No	Name	Characteristics
1	NRC-265	Short Flowering duration (<60 days)
2	NRC-346	Bold nut type (7- 12 g) (>7 g)
3	NRC-349	Bold nut type (7- 12 g) (>7 g)
4	NRC-366	Big apple type (>52 g)
5	NRC-383	Big apple type (>52 g)
6	NRC-385	Big apple type (>52 g)
7	NRC-386	Big apple type (>52 g)
8	NRC-458	Small apple type; High shelling % (>28%)
9	NRC-470	High shelling% (>28%)
10	NRC-478	High yielding (6 years cumulative harvests > 18 kg/tree)
11	NRC-20	Long flowering type (> 90 days)
12	NRC-38	High apple to nut ratio (> 12.0)
13	NRC-40	Early Flowering (Nov– Dec)
14	NRC-145	Big apple type (>52 g)
15	NRC-160	High yielding (6 years cumulative harvests > 18 kg/tree)
16	NRC-270	Big apple type (>52 g)
17	NRC-308	High shelling % (>28%) Smooth apple surface
18	NRC-318	High shelling % (>28%)
19	NRC-333	High shelling % (>28%); High kernel weight (> 2.5 g)
20	NRC-335	Low Shelling % (<18%)
21	H-130	Jumbo nut type (> 12 g)
22	Brazil dwarf	Dwarf type (<4 m height)
23	Purple mutant	Purple fruit and shoots
24	NRC-450 (Thaliparamba-1)	Compact type
25	Bhaskara	High yielding (6 years cumulative harvests > 18 kg/tree)
26	NRC-364 (V-7)	Bold nut type (7- 12 g)
27	NRC-188	CNSL free
28	NRC-281	CNSL free
29	NRC-285	CNSL free
30	NRC-473 (Kodippady-1)	Compact type
31	NRC-391	Bold nut type (> 7 g)
32	NRC-152	Lacks ridges on the cashew apple; unique nut shape

buffer by electrophoresis at 70 V for 3 h. The gels were visualized by exposing them to UV light in the Gel Doc system (Alpha Imager, USA). The primer pairs producing clear and distinct PCR bands in the expected size range on the gel were considered positive amplifications. The primer pairs showing positive amplifications were evaluated for polymorphism detection in 32 cashew genotypes at standardized PCR conditions (Table 1), and the primer pairs displaying different-sized bands among the genotypes were considered polymorphic. Further, SSR primer pairs were also tested for cross-species amplification in *A. microcarpum* and *A. othonianum* by PCR screening, and the analysis of PCR products was the same as described above for cashew. The SSR markers producing specific PCR bands in the expected size range were considered cross-transferable to the related *Anacardium* species.

## 2.4 Data analysis

The PCR bands in the gel photos of each SSR primer were scored in the allelic format (band sizes in bp). After scoring the data, genetic diversity, heterozygosity, allele frequencies, allele number, genotype frequency, and polymorphic information content (PIC) values were calculated for each SSR marker using PowerMarker V3 (Liu and Muse, 2005). Dice index-based dissimilarity matrix was calculated and used for clustering analysis of cashew genotypes by the Neighbor-Joining (NJ) method with the DARwin software V6 (Perrier and Jacquemond-Collet, 2006).

## 2.5 Development of the Cashew Microsatellite Database

Cashew Microsatellite Database (CMDB) was designed and implemented as a three-tier architecture website using the tech tools Node js, react js, and MongoDB. Users can create a custom query using the multiple filters available on submission of the search button on the website. An HTTP request is made to the Node JS

server where the request is processed, and data will be retrieved from MongoDB and sent back as an HTTP response to the client browser. The basic scheme of CMDB development involved the following steps: i) Genomic and Genic SSR datasets were experimented with, recorded, and consolidated; ii) entities and relationships among the entities in the relational database were created; iii) Database Normalization i.e., organization of data into tables in such a way that the results of using the database are always unambiguous and intended; iv) relationships between the tables were established according to rules designed both to protect the data and to make the database more flexible by eliminating redundancy and inconsistent dependency; and v) design and implementation of a three-tier architecture website using tech tools Node js, react js and MongoDB to make the SSR data accessible to the user.

## 3 Results

### 3.1 Composition and distribution of genome-wide microsatellite/SSRs

A total of 54,526 SSRs were mined from the 356 Mb draft genome sequence of *A. occidentale*, with mean marker density of 153 SSRs per Mb (Table 2). However, the primer pairs could be successfully designed for flanking sequences of 47,646 SSRs of the detected SSRs (Supplementary Table 1). Analysis of repeat motifs showed that 87.39% of mined SSRs were the perfect type of SSRs, i.e., repeat motifs are continuous without interruption by any nucleotide [e.g., (GC)<sub>20</sub>], while 12.61% were the imperfect or compound type of SSRs, i.e., SSRs with the stretches of repeat motifs interrupted by nucleotides that are not repeated [e.g., (AT)<sub>12</sub>GC(AT)<sub>8</sub>] (Table 2).

Analysis of the distribution of the five classes of perfect SSRs in the cashew genome revealed that the dinucleotide repeat types were most abundant, comprising 68.98%, followed by the trinucleotide repeat motifs (24.57%), the compound SSR repeats (12.62%), the tetra-nucleotide repeat motifs (4.70%), the pentanucleotide repeat

TABLE 2 Summary statistics and characteristics of genome-wide SSRs in cashew genome.

Microsatellite Mining	Number	Percentage	Density/Mb
Total number of sequences examined:	3268	-	-
Total size of examined sequences (bp):	356594228	-	-
Total number of identified SSRs:	54526	100	152.77
Number of SSR containing sequences:	2572	78.70	-
Number of sequences containing $\geq 1$ SSR:	2132	65.23	-
Di-nucleotide SSRs	37614	68.98	105.66
Tri-nucleotide SSRs	13395	24.57	37.63
Tetra-nucleotide SSRs	2560	4.70	7.19
Penta-nucleotide SSRs	523	0.96	1.47
Hexa-nucleotide SSRs	434	0.80	1.22
Compound SSRs	6880	12.62	19.33

motifs (0.96%), and the hexanucleotide repeat motifs (0.80%) (Table 2). The frequency distribution of different repeat motifs in the draft genome of cashew is presented in Figure 1A. The nucleotide composition of the identified SSRs showed that 84% were composed of A and/or T nucleotides, while 16% were composed of G and/or C nucleotides. The most dominant repeat sequences were AT (23.54%), followed by TA (16.71%) and AAT (4%) (Figure 1B). The detailed frequency distribution of repeat motifs and the repeat numbers in the di- and tri-nucleotide SSRs, which are dominant in the genome-wide SSRs, is presented in Table 3. In the di-nucleotide SSRs, AT/AT repeat motifs were most abundant (65.79%) and CG/CG repeats were the least abundant repeat motifs (0.30), and the remaining two types, AC/GT and AG/CT, were 17.38% and 16.49%, respectively. In the trinucleotide SSRs, AAT/ATT repeat motifs were most abundant (55.37%), followed by AAG/CTT with a frequency of 22.46%, and ATC/ATG repeat motifs were 8.94%, ACC/GGT motifs were 3.81%, AAC/GTT motifs were 3.10%, AGC/CTG motifs were 2.31%, AGG/CCT motifs were 2.61%, and other motifs (ACG/CGT, ACT/AGT, and CGG/CGG) together were 1.39%.

Based on the size of repeats motif, the mined SSRs were categorized into two classes, viz., Class I (hypervariable) SSRs, the SSRs with repeat motif size  $\geq 20$  bp, and Class II (variable) SSRs, the SSRs with repeat motif sizes of  $\geq 12$  and  $< 20$  bp. The frequency of Class I genomic SSRs was 45.10%, while that of Class II types was 54.89%. The Class I SSRs were dominated by the dinucleotide (46.71%) and the compound SSRs (24.66%), while the Class II SSRs were dominated by the dinucleotide (69.01%) and the tri-nucleotide SSRs (30.92%).

### 3.2 Polymorphism, genetic diversity, and transferability of genomic SSR markers

We validated the mined genomic SSR markers by synthesizing 100 primer pairs and testing them for PCR amplification in *A.*

*occidentale*. All the tested SSR primer pairs were successfully amplified in *A. occidentale*, indicating 100% accuracy in primer design. Further, fifty nine of 100 primer pairs screened in 32 germplasm accessions showed polymorphism (Table 4). The 59 polymorphic markers detected 294 alleles in 32 accessions. The number of alleles per SSR locus varied from 2 to 15, with a mean of 4.98 alleles per SSR locus (Table 4). PIC values of the assayed SSRs ranged from 0.19 to 0.84, with a mean of 0.59 (Table 4). Polymorphic SSR markers were grouped into three classes based on the PIC values. Of the 59 SSR markers developed, 47 were highly polymorphic (PIC value  $\geq 0.50$ ), 9 were moderately polymorphic (PIC value between 0.25-0.50), and 3 markers were least polymorphic (PIC value  $< 0.25$ ) (Table 4). Further, 39 of the polymorphic SSRs were of di-nucleotide type, 8 were of trinucleotide type, and 12 were compound SSR repeats (3 perfect and 9 imperfect type compound SSRs). The wide range in amplicon size difference ( $\sim 150$ – $250$  bp) was observed with the marker DCR SSR-22 amplifying a compound motif: (GT)<sub>6</sub>ct(GA)<sub>15</sub> (Figure 2A).

The dendrogram analysis of 32 genotypes using the 59 polymorphic SSR markers classified the assayed genotypes into three major clusters (Figure 2B). The first cluster consisted of 16 genotypes; the second cluster consisted of 14 genotypes; and the third cluster consisted of two genotypes. Pairwise dissimilarity was a maximum of 0.85 between NRC-335 and NRC-265, while a minimum of 0.33 was observed between NRC-385 and NRC-386. NRC-385 and NRC-386 are the two genotypes originating from a common parent, as depicted in the dendrogram analysis. Further, the NRC-335 and NRC-265 are from two geographically distinct regions, i.e., the NRC-335 is from the West Coast region of India, while the NRC-265 is from the East Coast region of India.

So far, there are no species-specific genetic markers designed for less studied species in the *Anacardium* genus. Cross-species transferability of cashew SSRs can be an alternative source of molecular markers for less studied *Anacardium* species. Testing of cross-species PCR amplification of the newly designed SSR primers in the two *Anacardium* species, viz., *A. microcarpum* and

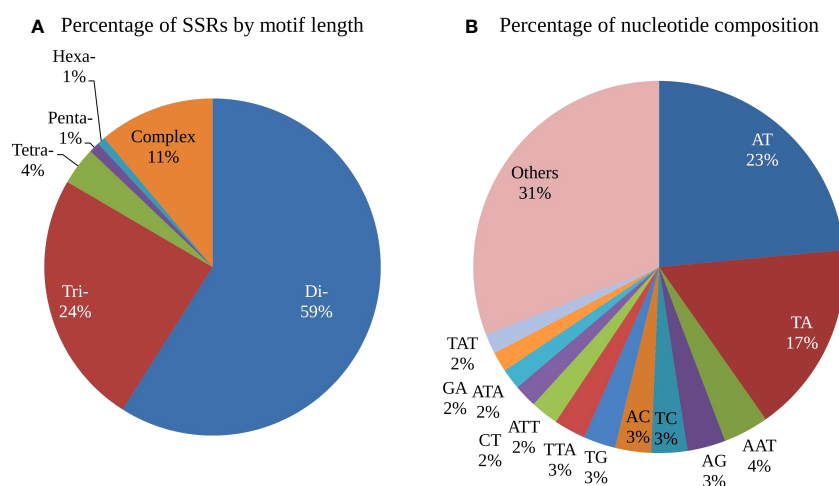


FIGURE 1

Frequency distribution of different types of SSR repeats in the Cashew genome (A) Frequency of motif types by unit length (K-mers) (B) Frequency of repeat motifs by nucleotide composition.



TABLE 3 Frequencies of different repeat motifs in di- and tri-nucleotide SSRs in cashew genome.

Repeat Motif	Repeat number								Total	%
	5	6	7	8	9	10	> 10			
AC/GT	–	1829	1064	727	660	454	1805	6539	17.38	
AG/CT	–	2035	967	611	570	418	1603	6204	16.49	
AT/AT	–	5776	3701	2964	2279	1782	8245	24747	65.79	
CG/CG	–	91	20	11	2	–	0	124	0.33	
Total	–	9731	5752	4313	3511	2654	11653	37614	–	
AAC/GTT	233	77	56	26	10	7	6	415	3.10	
AAG/CTT	1257	661	389	241	172	101	188	3009	22.46	
AAT/ATT	3291	1652	972	593	336	220	353	7417	55.37	
ACC/GGT	266	123	56	37	13	8	8	511	3.81	
ACG/CGT	29	12	8	2	1	1	0	53	0.40	
ACT/AGT	36	18	10	5	5	1	6	81	0.60	
AGC/CTG	210	68	24	8	–	–	0	310	2.31	
AGG/CCT	149	95	34	28	21	11	12	350	2.61	
ATC/ATG	631	309	124	58	33	18	24	1197	8.94	
CCG/CGG	39	7	4	–	1	1	0	52	0.39	
Total	6141	3022	1677	998	592	368	597	13395	–	

TABLE 4 Characteristics of the 59 novel polymorphic SSR markers developed in this study by scanning of the whole genome sequence of cashew for SSRs and validation.

Sl No.	Marker	SSR primer (F&R)	Ta (°C)	Repeat motif	Allele size range bp	Allele No	PIC
1.	DCR SSR-1	GAAACACCTGTCCACACGC CCTTGACCTCGTGCCAATCT	54	(AT)15	~260-325	5	0.66
2.	DCR SSR-3	AGTTTGCCGAAGCTCTCAACT GCGGAAGTAGTCTTTTGGCA	54	(TA)7	~275-320	4	0.69
3.	DCR SSR-6	ATACGTCCAACAAACGCCCT AGCGAACGATGTTGTTTGGT	53	(AT)10	~225-275	3	0.48
4.	DCR SSR-8	TGCTGCACAGAGAGACTTGG AGGAAGATTTGCCTGCAGCT	54	(TTC)5	~200-225	4	0.64
5.	DCR SSR-9	TGGCTATTTCTTGGGCAGG TTTTCTCTCAGCCGTCCC	53	(CT)7	~225-275	5	0.56
6.	DCR SSR-10	GGTTGTTGAGTAGCAGGGGT ACAACITGCAATTGTGGT	56	(TA)29	~250-300	4	0.61

(Continued)

TABLE 4 Continued

Sl No.	Marker	SSR primer (F&R)	Ta (°C)	Repeat motif	Allele size range bp	Allele No	PIC
7.	DCR SSR-11	GCGTACACACACACACACAC GCGAATGGGAAGTTGCCAAG	54	(AC)9	~275-300	3	0.51
8.	DCR SSR-12	TTTCCACGCCTACCACAGTC AGAGGGGAAAAGTGCAAGCA	54	(TA)25	~250-300	4	0.52
9.	DCR SSR-13	GAGGTTGATCCACAGCAGCT GCATTGCAGGCACAAGAACA	55	(TTG)8	~250-300	4	0.67
10.	DCR SSR-14	GATCACACGCACGATGAAGC CAAGCCCCCTTTTCTGTGC	56	(AT)6	~230-300	5	0.6
11.	DCR SSR-15	CGCGATGGGAATTCTACGGA CCCACACAGCGATCTCAGTT	54	(AT)20	~260-300	3	0.58
12.	DCR SSR-17	GGTGGTATGTTGGAAGCCCA TGGGAGCTAACCTAGAGCACA	52	(AT)16	~240-275	6	0.67
13.	DCR SSR-20	ATCGCCACCTACCCAACAAG GCAATGCAAAGTATGAGGGTGG	52	(AATA)5(AT)22	~250-275	15	0.82
14.	DCR SSR-21	TGGATTGAGAGGGTTCAAGGG ACACACGAGTCTGTGTCCAA	56	(AT)7tatatgtatgtgtatataatg(TA)11	~175-240	3	0.59
15.	DCR SSR-22	CCGTGTGTGTGTCTGAGA GCTGAGGCATCTCTTGGGT	54	(GT)6ct(GA)15	~150-250	5	0.66
16.	DCR SSR-23	AGTCATCGTCGCTGATTACG ACCTGCTGTGTTGATCAGACT	54	(AT)35	~175-250	4	0.57
17.	DCR SSR-28	AGATGTGTGTGGCTTCAGG TCCGCACTCTCAGCTTGTT	54	(AC)9ag(AT)18	~250-300	4	0.64
18.	DCR SSR-29	TGGTTGGATTTCCCTTGGA ATTAAGTGAACGCGAGCCCA	54	(AT)39	~175-250	4	0.57
19.	DCR SSR-30	ACACTTCCCATGAACAAGGACT CATTAGCTCGAGGCCTGACA	54	(AT)15	~240-300	4	0.63
20.	DCR SSR-32	CTGGCTAACGGGAGGTTTGT GAGAGAGGGAGAAGGGGAGG		(GT)6(AT)10	~245-290	6	0.76
21.	DCR SSR-33	ACCAATCCCACCAGCAACAA	54	(TG)10tctgtctctgtg(TGTGTC)5 tgtgtctgtgc(AT)7acaca cacacatatatgtgtctgtgtctgtgc(AT)	~200-250	6	0.67

(Continued)

TABLE 4 Continued

SI No.	Marker	SSR primer (F&R)	Ta (°C)	Repeat motif	Allele size range bp	Allele No	PIC
				7acacacacatata(TG)20 tctgtgc(AT)7acacacata(TG) 6tctgtgc (AT)11acacacacac(GT)9			
		GCACAGACACACACACACAC					
22.	DCR SSR-34	CTCTACCCACTCACCGAGGA	54	(AT)21	~220-275	5	0.53
		AGCACGTTCCACAAGGTTCA					
23.	DCR SSR-35	TGGAGGTGTTTGGGATGCAA	54	(AT)16	~250-300	6	0.7
		TCCCCATTGTGGTTGTGCA					
24.	DCR SSR-36	GTCTGATCAGCACAGCAGGT	54	(AT)6	~250-275	3	0.44
		CCCTCAAACCCAAGCAAGC					
25.	DCR SSR-37	GGGTGGGGTGAGTTCCATT	54	(TAT)8	~230-250	5	0.6
		GGCCAACCCAGCTTGAAATG					
26.	DCR SSR-38	GGCAACACGTCACCTGGATA	54	(AT)29	~200-250	11	0.84
		TAAAGAATGCCTGGGCCACA					
27.	DCR SSR-39	AATAAGAGCACCTCGGCAGC	54	(TA)14	~250-275	3	0.29
		GCACGTTGCACATGTTTTCG					
28.	DCR SSR-41	TGTCTACACCTGTTTCTCCGT	56	(TA)9	~250-300	6	0.56
		GGCAAGTAGTAGCTCCACCC					
29.	DCR SSR-42	CAAGAGGCCCAAAAACAGC	54	(AT)17	~250-325	6	0.72
		GCCTGCCACCCTCACAATAT					
30.	DCR SSR-43	TGTCCAGGGAGAGTACTGT	52	(AT)22	~250-300	6	0.7
		TTTCAGCTGGCATGCCATA					
31.	DCR SSR-44	TTTTGGGTTGGCAATGGCTG	54	(AC)7aa(AT)25(AC)9aa(AT)28	~225-250	5	0.66
		CTCATCGAGGTTGGTTGCCT					
32.	DCR SSR-45	CCCCTGCAATTTCCACGTG	54	(AT)16	~250-275	6	0.63
		AAACCCTCAGAGCTGATGCC					
33.	DCR SSR-46	GTTGGAACAGGCGACCTACA	54	(AT)46	~190-250	4	0.48
		ACCATTTCGAACTGGAGCCA					
34.	DCR SSR-47	ACCCACAGCTAGCCCAAATC	54	(AT)11	~225-275	5	0.71
		GGCAAGCCTAGGCCTACAAA					
35.	DCR SSR-49	GCGGGAAAACATGTGGTGTG	54	(TA)10aggttttccc(AT)10	~200-225	4	0.63

(Continued)



TABLE 4 Continued

Sl No.	Marker	SSR primer (F&R)	Ta (°C)	Repeat motif	Allele size range bp	Allele No	PIC
		CATGTTGTGGCTTGCATGCT					
36.	DCR SSR-51	GGGGT GAGTAGTTGGCCATC	54	(TAA)7tgcaatacctgattgaattcagaagctg Aatggttgaattgtcccacgggatgggggtgagta gttgccatc(ATA)6	~225-250	5	0.71
		AGGAGCTGTACGCAAAACCA					
37.	DCR SSR-52	GCTTGAGTTTGGCTTGGCTT	54	(TA)19	~200-225	2	0.38
		TCCACACAAGGCCACAAGTT					
38.	DCR SSR-53	TGGAGTTACCCACCTGTACCA	54	(TA)6	~225-275	4	0.62
		GGCTGTGAAGAAGTGTTCGC					
39.	DCR SSR-56	GGCTTGAGTGCTGAATCCCT	54	(TA)34	~225-250	2	0.3
		TCATAAACTTGCCTTGGGCA					
40.	DCR SSR-58	TGCTCCTGCCTTTGTGCTAA	54	(AC)7	~200-225	4	0.64
		ACACGAAAACCTCAACGTGGG					
41.	DCR SSR-59	ACCCAGTTTAACAGGCTGAA	54	(TAA)5	~200-250	4	0.7
		TTTTCTCCGCTTCTCGCAT					
42.	DCR SSR-60	GGGAATGGTGGGGTAAAGG	55	(TGG)5c(GGA)5ggtggaggtgctc gtactaatggtgagggtatggtcatg gaagtgtttcgggacaggagcggagcaggaa (GTG)6(GAG)5	~225-250	5	0.56
		TCTCATCAGACCACCTCCA					
43.	DCR SSR-62	TCCTTGCTGGCGACTTTGAT	54	(GT)12(AT)8	~175-200	4	0.58
		TTCGTCCCTTCTCTTCTCT					
44.	DCR SSR-63	TGGATTCTCCCTTCCCCCTT	54	(TC)16	~225-250	2	0.37
		AATAGGCCAGGAGGTGAGA					
45.	DCR SSR-67	TGATAGGCCGAATCAGCGAC	54	(CCA)7	~200-225	4	0.6
		ACTATTGTTCTCTGGCGGCC					
46.	DCR SSR-68	GGCAGTGTCTGGTAGGTCAC	54	(AG)9acggtaccaaacatttactattagtctcc ttatcatgcttgaatatgg(AT)9	~225-250	2	0.19
		TTGATCAAGAGGCCAGTCCG					
47.	DCR SSR-69	CAAGCAGCAAACAAGGCAT	54	(TA)7	~200-250	3	0.46
		AACACAGCAGGTACCTTCGG					
48.	DCR SSR-70	CGGTAGTTTTAGGGCTCCG	54	(AG)7	~200-250	2	0.36
		AACTCATGGGGCACTGTCTG					
49.	DCR SSR-71	GGAAATCCGGAAAGGGTCA	54	(TA)10	~200-225	2	0.19

(Continued)

TABLE 4 Continued

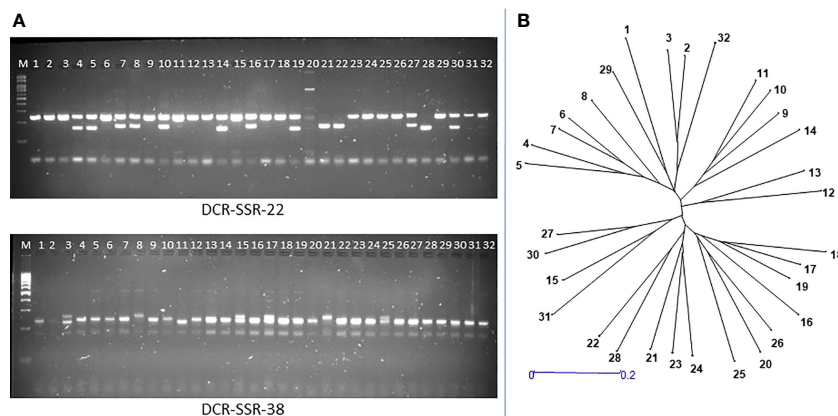
Sl No.	Marker	SSR primer (F&R)	Ta (°C)	Repeat motif	Allele size range bp	Allele No	PIC
		GGCCAGCCAACCTAATGACT					
50.	DCR SSR-72	GTTTGGTAGAGGGATCCTGCA	54	(TA)20	~175-200	2	0.19
		GAAGGTGTTACGGTGGCTCA					
51.	DCR SSR-73	CACAGAGGTGCATGATTTGGTG	56	(AT)8	~210-240	6	0.65
		GGTGCTTTCAACGATCATGGA					
52.	DCR SSR-74	CACCTCCCTCTACAAACGCA	56	(AT)8	~190-220	9	0.81
		AGAGGCTCGGGTAAGTGAGT					
53.	DCR SSR-75	AACCAACAAGTCGGGTGGTT	56	(AAG)8	~170-200	9	0.77
		GTCCCTCCAGGTAAGGAT					
54.	DCR SSR-76	GGACTCATATGTTGGTGTGCT	56	(AT)11	~190-220	6	0.71
		ATTGATCTGTGCTGGCTCC					
55.	DCR SSR-77	AACTGCAGGTCTCTGGGAGA	56	(ATA)10	~100-130	6	0.54
		CTCGCCCTCTTGTCAACA					
56.	DCR SSR-78	ACCCTGCGTTATGTGCACT	56	(AG)12	~240-260	7	0.76
		GCCAACACCAGCACTGAAA					
57.	DCR SSR-79	CATGGGAGCATCTGGACCAG	56	(TA)6	~185-210	10	0.82
		AGTCTGGCTCTGTTGTGCA					
58.	DCR SSR-80	TGGAAAGGCTGGAGAAGAGA	56	(TTA)11	~275-290	12	0.8
		GGTTGAATCAAGGTTGTTGCA					
59.	DCR SSR-81	TAGAAGGCTTGGGAGTGGA	56	(AT)13	~250-275	6	0.75
		ATCCACTCCCAGACCCTAGG					

*A. othonianum*, showed that 91% of the tested primers were successful in PCR amplification (Supplementary Table 2), suggesting a high rate of transferability of cashew genomic SSRs in the *Anacardium* genus.

### 3.3 CMDB: microsatellite database for cashew

The Cashew Microsatellite Database (CMDB) is an online relational database that stores the microsatellite repeats information mined from the recently sequenced cashew genome (Savadi et al., 2022b) and the shoot transcriptome (Savadi et al.,

2022a), as well as the experimentally validated SSR markers. CMDB is available at <https://www.cashewmicrosatellitesdatabase.in/>. CMDB is an interactive database that has been implemented as a 3-tier application architecture where we have a Client tier, an Application or Server tier, and a Database tier, as shown in Figure 3A. CMDB has a user-friendly interface developed using React js and a server designed and implemented using Node js that connects to MongoDB, where all the genomic and genic SSR data is stored. Users can access this responsive website using any browser on a desktop or mobile device connected to the internet. User-need-based customized queries can be generated from the web interface and allow users to search the Cashew microsatellite database in MongoDB.

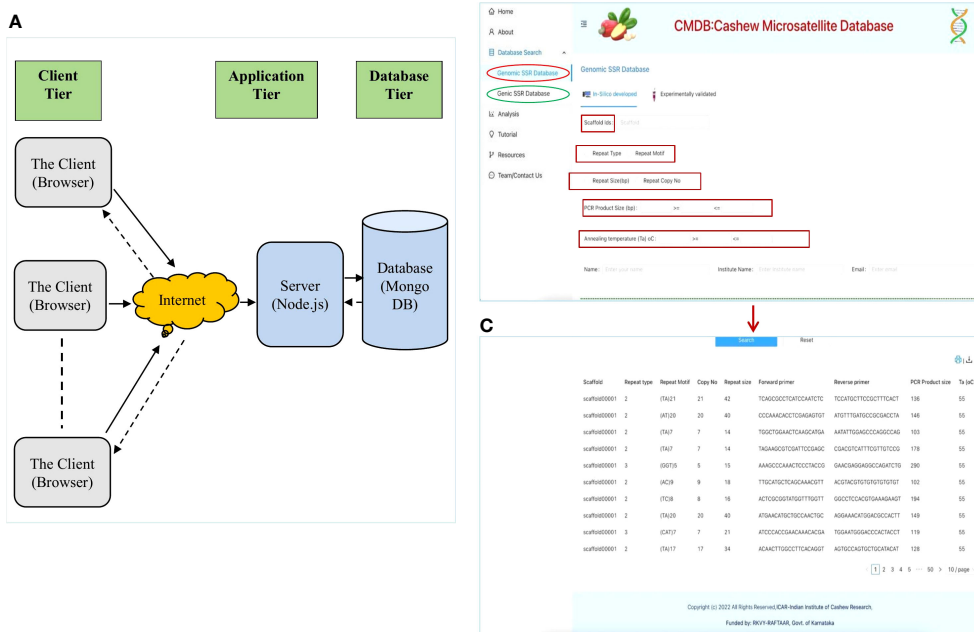


**FIGURE 2** Validation of genome-wide SSRs for amplification and polymorphism in 32 cashew genotypes: **(A)** Gel pictures showing polymorphism detection by the DCR SSR-22 and DCR SSR-38 markers in 32 cashew genotypes; **(B)** Neighbour-Joining dendrogram showing genetic relationships among 32 cashew accessions collected from different geographic regions. The dendrogram is constructed based on Nei's (D) genetic distance coefficient.

CMDB can be searched to extract genomic as well as genic microsatellites based on motif type (di, tri, tetra, penta, and hexa), repeat motif, copy number, repeat size, expected PCR product size and primer pair annealing temperatures (Ta) (Figure 3B). The microsatellites can be searched based on the choice of scaffolds/transcripts, where more than one scaffold/transcript can be selected using the dropdown option (Figure 3B) the results of the database search displays details of SSRs including the primer pairs for the displayed SSRs (Figure 3C). This is a novel approach and is helpful for breeders and biotechnologists to easily extract microsatellites based on their needs.

### 4 Discussion

Though the integration of molecular markers in breeding programs and genetic studies has substantially enhanced the speed and accuracy of crop improvement in important fruit and nut crops, molecular breeding and genetic analysis in cashew have lagged behind due to the scarcity of informative markers. Microsatellites or SSRs are highly informative markers and are widely used in genetic analyses and breeding of crops, including trees. In *A. occidentale*, very limited SSR markers are available for



**FIGURE 3** The interface and searching of the Cashew Microsatellite Database (CMDB) for SSRs: **(A)** The three-tier architecture of CMDB, **(B)** The database search page displaying different SSR search parameters; and **(C)** The database search results displaying the details of SSRs.

comprehensive genetic studies and molecular breeding (Croxford et al., 2006; Savadi et al., 2022a).

The NGS technologies allow rapid large-scale sequencings at a lesser cost, which permits discovery and development of SSR markers for less studied crops (Khodaeiaminjan et al., 2018; Taheri et al., 2018; Luo et al., 2021; Wang et al., 2022). The genome sequences generated using the NGS technology have been used to discover genome-wide SSRs and develop SSR markers in various tree species, such as pistachio (Ziya Motalebipour et al., 2016), hazelnut (Öztürk et al., 2018), avocado (Ge et al., 2019), fruit and forest species (Song et al., 2021), and *Grevillea* sp. (Dabral et al., 2021). Efficient utilization of the large set of SSRs mined through genome scanning is possible with a user-friendly web tool to search the SSR information from a database. So far, genome-wide SSRs discovery and development of a database for storage and retrieval of the discovered SSRs have not been reported in cashew. The present study, for the first time, reports the discovery and characterization of genome-wide SSRs and the development of a microsatellite database (CMDDB) for cashew.

In this study, a total of 54526 SSRs were discovered from the cashew genome, with a mean density of 153 SSRs/Mb. The density of markers found in this study was higher than that reported in apple (40.8 SSRs/Mb, 485 Mb) (Zhang et al., 2021b), Chinese spring wheat (36.68 SSRs/Mb, 9.93 Gb) (Han et al., 2015), and *Matthiola incana* (23.25 SSRs/Mb, 1977.48 Mb) (Tan et al., 2023), while it was less than mango (418.17 SSRs/Mb, 253.6 Mb) (Ravishankar et al., 2015), *Prunus mume* (794 SSRs/Mb, 237 Mb) (Sun et al., 2013), and Pomegranate (527.97 SSRs/Mb, 296 Mb) (Patil et al., 2020), suggesting that generally, the density of SSRs decrease with an increase in the genome size. Further, the frequency of perfect SSRs (87.39%) was much higher than the frequency of imperfect SSRs (12.61%) in the discovered genome-wide SSRs and is consistent with similar results in eggplant (Portis et al., 2018), *Anemone coronaria* (Martina et al., 2022), and *Aristolelia chilensis* (Bastías et al., 2016). In the perfect SSRs, dinucleotide repeat types were most dominant (68.98%), followed by trinucleotide repeat motifs (24.56%), which are consistent with similar results in other plant species investigations (Parida et al., 2015; Vieira et al., 2016; Ziya Motalebipour et al., 2016; Wang et al., 2018).

The most common dinucleotide was AT/AT repeat motifs (65.79%), while CG/CG repeats were the least abundant (0.30). This result is in agreement with previous findings that AT-rich SSRs are predominant in dicots, viz., apple (Zhang et al., 2012a), sweet orange (Biswas et al., 2014), and *Cucumis sativus* (Cavagnaro et al., 2010), while GC-rich dinucleotide repeats are dominant in monocots (Sonah et al., 2011; Qin et al., 2015). These differences in the SSR nucleotide among the dicots and monocots could be partially explained based on the relative nucleotide composition of the genomes. The average GC content of dicot genomes (34.6%) is lower than that of monocot genomes (43.7%) (Cavagnaro et al., 2010), and it is observed that the frequency of AT and TA in the genomes increased with the evolution of the plant kingdom (Qin et al., 2013).

In this study, Class I type of SSRs ( $\geq 20$  bp repeat motif) were 45.10%, while Class II types ( $\geq 12$  and  $< 20$  bp repeat motif) were 54.89%. The frequencies of two classes of SSRs are in agreement

with other studies in plants (Parida et al., 2015; Wang et al., 2018; Patil et al., 2021). Class I SSRs are observed to be highly polymorphic compared to Class II SSRs (Parida et al., 2015; Vieira et al., 2016; Patil et al., 2021) because shorter SSR sequences tend to have lower mutation rates (Vieira et al., 2016).

The validation of genome-wide SSRs was performed by the synthesis and screening of 100 randomly selected SSR primer pairs in cashew genotypes. Fifty nine of the 100 SSR primers screened in 32 germplasm accessions showed polymorphism. To date, 21 genomic and 36 genic SSRs have been reported in cashew (Croxford et al., 2006; Savadi et al., 2022a). Therefore, this study not only provides genome-wide SSR information but also experimentally validated polymorphic SSR markers for cashew. Further, SSR markers grouped based on the PIC values showed that 47 newly developed SSR markers were highly polymorphic (PIC value  $\geq 0.5$ ), 9 were moderately polymorphic (PIC value between 0.25 and 0.50), and 3 markers were least polymorphic (PIC value  $< 0.25$ ) according to the Botstein et al. (1980) classification of polymorphic markers. Furthermore, the grouping of polymorphic SSRs based on the five classes of perfect SSRs and compound SSRs showed that the dinucleotides were dominant. The higher tendency of dinucleotide SSRs to be polymorphic is consistent with other studies on genomic SSRs in apples (Silfverberg-Dilworth et al., 2006), peanuts (Zhao et al., 2012a), watermelon (Zhao et al., 2012b), and black pepper (Negi et al., 2022). However, in a previous study where a repeat-rich genomic library screening method was used to generate SSR markers, the compound SSRs were found to be more polymorphic in cashew (Croxford et al., 2006). This difference in the polymorphic SSRs motif size could be due to the biases caused by the repeat probe sequences (AC15, AG15, AAC8, AAG8, AAT8, ACC8, AGG8, ATC8, AAAC6, AAAG6, and ACAT6) used to screen the genomic libraries for SSRs. The genetic diversity analysis using the newly developed SSRs clustered the 32 genotypes into three major clusters. Further, the pairwise dissimilarity index revealed that the maximum distinction was observed between the accessions viz., NRC-335 and NRC-265 collected from different geographic regions, the West Coast and the East Coast of India, respectively, while the minimum pairwise dissimilarity index was observed between the two genotypes, viz., NRC-385 and NRC-386, which shared one of the parents, indicating that the newly developed genomic SSRs have high discriminating power and present a powerful molecular tool for investigating genetic diversity and genetic relationships in the cashew genotypes.

To our knowledge, there is no distinct set of microsatellites or SSR markers developed for other species in the *Anacardium* genus. It has been demonstrated that SSR markers have a high potential for cross-species amplification/transferability in related species of the same genus. Transferability of markers is considered a cost-effective approach for developing genetic markers for species lacking genomic resources (Ellis and Burke, 2007; Vieira et al., 2016). In this study, 91% of the SSR primer pairs showed cross-species amplifications in the two wild relatives of cashew, viz., *A. microcarpum* and *A. othonianum*. This transferability rate was comparable with the results observed in previous studies using genomic SSRs in cashew (Croxford et al., 2006; Soares et al., 2013),

but lower than the transferability with the genic SSRs (Savadi et al., 2022a). In *A. humile*, 85% of the 14 tested cashew SSRs amplified (Soares et al., 2013); in *A. microcarpum*, *A. pumilum*, and *A. nanum*, 92% of the 12 cashew SSRs amplified (Croxford et al., 2006); and in *A. microcarpum* and *A. othonianum*, 100% of 54 transcriptome-based SSRs amplified (Savadi et al., 2022a). The relatively lower rate of transferability of genomic SSRs compared to genic SSRs could be attributed to the higher conservation of genic sequences compared to sequences from the anonymous regions of the genomes (Ellis and Burke, 2007; Jiang et al., 2020). Thus, we contemplate that the SSR markers developed in this study can be a potential marker repository for not only cashew but also the related *Anacardium* species and could be employed for macro-syntenic comparisons, germplasm characterizations, genetic mapping, molecular breeding involving interspecies hybridizations, etc.

With the mining of genome-wide microsatellites/SSRs information, there is a need to develop a user-friendly web tool for easy access and efficient utilization of the mined SSRs in genetic studies and crop improvement. Several databases of genome-wide SSRs have been designed in different crop plants (Arora et al., 2013; Dossa et al., 2017; Jasrotia et al., 2019; Martina et al., 2022). However, in cashew, genome-wide SSRs information is not available. In this study, genome-wide SSR information generated in this study as well as in our previous study from transcriptome data (Savadi et al., 2022a) was integrated into CMDB, which permits the extraction of information related to both genomic and genic SSRs. Further, CMDB also provides the experimentally validated SSR markers in the cashew. Most of the microsatellite databases developed in other crops (Arora et al., 2013; Dossa et al., 2017; Jasrotia et al., 2019; Martina et al., 2022) provide only the *in silico* mined SSRs information of either the genomic SSRs or transcriptomic SSRs but not combined information like the CMDB. Thus, CMDB is the first comprehensive microsatellite database for cashew, and it will be of great use to cashew researchers, particularly the breeders, to develop novel markers from *in silico* mined SSRs and to directly use the experimentally validated markers in the research programs.

## 5 Conclusion

The limited availability of microsatellite/SSR markers in cashew has hindered genetic studies and crop improvement. In the current study, we mined and characterized genome-wide SSRs in the cashew genome and developed a cashew microsatellite database (CMDB), a comprehensive repository of microsatellites, which provides accessibility to genome-wide and transcriptome based SSRs information as well as the experimentally validated SSR markers to researchers and breeders. The large set of genome-wide SSRs and their free public accessibility will permit the development of a large set of new SSR markers for cashew, which are currently very scarce. Besides, we developed 59 highly informative SSR markers that are the first set of genomic SSRs developed in cashew through *in silico* mining of the cashew genome.

Thus, the knowledge of genome-wide SSRs distribution, the development of novel SSR markers, the cross-species transferable SSRs, and the comprehensive microsatellite database would significantly accelerate genetic studies and crop improvement in cashew and related *Anacardium* species.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

## Author contributions

SS conceived the idea, carried out the *in silico* analysis and experiments and development of database and wrote the manuscript. BM contributed in the experimentation and development of database. VV contributed towards sampling and experimentation. JA contributed to manuscript writing and proof editing and development of database.

## Acknowledgments

The authors acknowledge the financial support and encouragement of the Director, ICAR-Directorate of Cashew Research, Puttur, Karnataka, India, and the Indian Council of Agricultural Research (ICAR), New Delhi.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1242025/full#supplementary-material>



## References

- Adiga, J. D., Muralidhara, B. M., Preethi, P., and Savadi, S. (2019). Phenological growth stages of the cashew tree (*Anacardium occidentale* L.) according to the extended BBCH scale. *Ann. Appl. Biol.* 175, 246–252. doi: 10.1111/aab.12526
- Akinwale, T. O. (2000). Cashew apple juice: its use in fortifying the nutritional quality of some tropical fruits. *Eur. Food Res. Technol.* 211, 205–207. doi: 10.1007/s002170050024
- Aliyu, O. M. (2012). Genetic diversity of Nigerian cashew germplasm. *Genet. Diversity Plants* 498, 163–184.
- Aliyu, O. M., and Awopetu, J. A. (2007). Chromosome studies in cashew (*Anacardium occidentale* L.). *Afr. J. Biotechnol.* 6, 131–136.
- Archak, S., Gaikwad, A. B., Gautam, D., Rao, E. B., Swamy, K. R. M., and Karihaloo, J. L. (2003a). DNA fingerprinting of Indian cashew (*Anacardium occidentale* L.) varieties using RAPD and ISSR techniques. *Euphytica* 130, 397–404. doi: 10.1023/A:1023074617348
- Archak, S., Gaikwad, A. B., Gautam, D., Rao, E. V., Swamy, K. R., and Karihaloo, J. L. (2003b). Comparative assessment of DNA fingerprinting techniques (RAPD ISSR and AFLP) for genetic analysis of cashew (*Anacardium occidentale* L.) accessions of India. *Genome* 46, 362–369. doi: 10.1139/g03-016
- Archak, S., Gaikwad, A. B., Swamy, K. R. M., and Karihaloo, J. L. (2009). Genetic analysis and historical perspective of cashew (*Anacardium occidentale* L.) introduction into India. *Genome* 52, 222–230. doi: 10.1139/G08-119
- Arora, V., Iqbal, M. A., Rai, A., and Kumar, D. (2013). PIPEMicroDB: microsatellite database and primer generation tool for pigeonpea genome. *Database* 2013, bas054. doi: 10.1093/database/bas054
- Ashraf, M. S., and Rathinasamy, K. (2018). Antibacterial and anticancer activity of the purified cashew nut shell liquid: implications in cancer chemotherapy and wound healing. *Nat. Prod. Res.* 32, 2856–2860. doi: 10.1080/14786419.2017.1380022
- Bastias, A., Correa, F., Rojas, P., Almada, R., Munoz, C., and Sagredo, B. (2016). Identification and characterization of microsatellite loci in maqui (*Aristotelia Chilensis* [Molina] Stunz) using next-generation sequencing (NGS). *PLoS One* 11, e0159825. doi: 10.1371/journal.pone.0159825
- Biswas, M. K., Xu, Q., Mayer, C., and Deng, X. (2014). Genome wide characterization of short tandem repeat markers in sweet orange (*Citrus sinensis*). *PLoS One* 9, e104182. doi: 10.1371/journal.pone.0104182
- Borges, A. N. C., Lopes, A. C. A., Britto, F. B., Vasconcelos, L. F. L., and Lima, P. S. C. (2018). Genetic diversity in a cajui (*Anacardium* spp.) germplasm bank as determined by ISSR markers. *Genet. Mol. Res.* 17, 1–14. doi: 10.4238/gmr18212
- Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32, 314.
- Cavagnaro, P. F., Senalik, D. A., Yang, L., Simon, P. W., Harkins, T. T., Kodira, C. D., et al. (2010). Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC Genom.* 11, 1–18. doi: 10.1186/1471-2164-11-569
- Chaduvula, P. K., Bonthala, V. S., Manjusha, V., Siddiq, E. A., Polumeta, A. K., and Prasad, G. M. (2015). CmMDB: a versatile database for *Cucumis melo* microsatellite markers and other horticulture crop research. *PLoS One* 10, e0118630. doi: 10.1371/journal.pone.0118630
- Collard, B. C., and Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc Lond. B Biol. Sci.* 363, 557–572. doi: 10.1098/rstb.2007.2170
- Croxford, A. E., Robson, M., and Wilkinson, M. J. (2006). Characterization and PCR multiplexing of polymorphic microsatellite loci in cashew (*Anacardium occidentale* L.) and their cross-species utilization. *Mol. Ecol. Notes* 6, 249–251. doi: 10.1111/j.1471-8286.2005.01208.x
- Dabral, A., Shamoan, A., Meena, R. K., Kant, R., Pandey, S., Ginwal, H. S., et al. (2021). Genome skimming-based simple sequence repeat (SSR) marker discovery and characterization in *Grevillea robusta*. *Physiol. Mol. Biol. Plants* 27, 1623–1638. doi: 10.1007/s12298-021-01035-w
- da Costa Gomes, M. F., Borges, A. N. C., Batista, G. S. S., de Andrade, L. G., Oliveira, M. E. A., de Almeida Lopes, A. C., et al. (2021). Genetic diversity and structure in natural populations of Cajui from Brazilian Cerrado. *Biosci. J.* 37, e37080. doi: 10.14393/BJ-v37n0a2021-53974
- Das, R., Arora, V., Jaiswal, S., Iqbal, M. A., Angadi, U. B., Fatma, S., Kumar, D., et al. (2019). PolyMorphPredict: a universal web-tool for rapid polymorphic microsatellite marker discovery from whole genome and transcriptome data. *Front. Plant Sci.* 9, 1966.
- Das, I., and Arora, A. (2017). Post-harvest processing technology for cashew apple—A review. *J. Food Eng.* 194, 87–98. doi: 10.1016/j.jfoodeng.2016.09.011
- Dhanaraj, A. L., Rao, E. B., Swamy, K. R. M., Bhat, M. G., Prasad, D. T., and Sundur, S. N. (2003). Using RAPDs to assess the diversity in Indian cashew (*Anacardium occidentale* L.) germplasm. *J. Hortic. Sci. Biotechnol.* 77, 41–47. doi: 10.1080/14620316.2002.11511454
- Dossa, K., Yu, J., Liao, B., Cisse, N., and Zhang, X. (2017). Development of highly informative genome-wide single sequence repeat markers for breeding applications in sesame and construction of a web resource: SisaBase. *Front. Plant Sci.* 8, 1470. doi: 10.3389/fpls.2017.01470
- dos Santos, J. O., Mayo, S. J., Bittencourt, C. B., and de Andrade, I. M. (2019). Genetic diversity in wild populations of the restinga ecotype of the cashew (*Anacardium occidentale*) in coastal Piauí Brazil. *Plant Syst. Evol.* 305, 913–924. doi: 10.1007/s00606-019-01611-4
- Ellis, J. R., and Burke, J. M. (2007). EST-SSRs as a resource for population genetic analyses. *Heredity* 99 (2), 125–132. doi: 10.1038/sj.hdy.6801001
- Furtado, L. B., Nascimento, R. C., Seidl, P. R., Guimarães, M. J. O., Costa, L. M., Rocha, J. C., et al. (2019). Eco-friendly corrosion inhibitors based on Cashew nut shell liquid (CNSL) for acidizing fluids. *J. Mol. Liq.* 284, 393–404. doi: 10.1016/j.jmolloiq.2019.02.083
- Ge, Y., Tan, L., Wu, B., Wang, T., Zhang, T., Chen, H., et al. (2019). Transcriptome sequencing of different avocado ecotypes: *de novo* transcriptome assembly annotation identification and validation of EST-SSR markers. *Forests* 10, 411. doi: 10.3390/f10050411
- Grover, A., and Sharma, P. C. (2016). Development and use of molecular markers: past and present. *Crit. Rev. Biotechnol.* 36, 290–302. doi: 10.3109/07388551.2014.959891
- Han, B., Wang, C., Tang, Z., Ren, Y., Li, Y., Zhang, D., et al. (2015). Genome-wide analysis of microsatellite markers based on sequenced database in Chinese spring wheat (*Triticum aestivum* L.). *PLoS One* 10, e0141540. doi: 10.1371/journal.pone.0141540
- INC Statistical Yearbook (2020–2021) *International Nut and Dried Fruit Council (INC)*. Cashew. Available at: <https://www.nutfruit.org/>.
- Inglis, P. W., Pappas, M. D. C. R., Resende, L. V., and Grattapaglia, D. (2018). Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS One* 13, e0206085. doi: 10.1371/journal.pone.0206085
- Iqbal, M. A., Jaiswal, S., Angadi, U. B., Sablok, G., Arora, V., Kumar, S., et al. (2015). SBMDB: first whole genome putative microsatellite DNA marker database of sugarbeet for bioenergy and industrial applications. *Database* 2015, bav111. doi: 10.1093/database/bav111
- Jasrotia, R. S., Yadav, P. K., Iqbal, M. A., Bhatt, S. B., Arora, V., Angadi, U. B., et al. (2019). VigSatDB: Genome-wide microsatellite DNA marker database of three species of *Vigna* for germplasm characterization and improvement. *Database* 2019, baz055. doi: 10.1093/database/baz055
- Jena, R. C., Samal, K. C., Pal, A., Das, B. K., and Chand, P. K. (2016). Genetic diversity among some promising Indian local selections and hybrids of cashew nut based on morphometric and molecular markers. *Int. J. Fruit Sci.*, 16 (1), 69–93.
- Jiang, Y., Xu, S., Wang, R., Zhou, J., Dou, J., Yin, Q., et al. (2020). Characterization, validation, and cross-species transferability of EST-SSR markers developed from *Lycoris aurea* and their application in genetic evaluation of *Lycoris* species. *BMC Plant Biol.* 20, 1–14. doi: 10.1186/s12870-020-02727-3
- Khodaeiaminjan, M., Kafkas, S., Motalebipour, E. Z., and Coban, N. (2018). *In silico* polymorphic novel SSR marker development and the first SSR-based genetic linkage map in pistachio. *Tree Genet. Genomes* 14, 1–14. doi: 10.1007/s11295-018-1259-8
- Liu, K., and Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128–2129. doi: 10.1093/bioinformatics/bti282
- Luo, L., Yang, Y., Zhao, H., Leng, P., Hu, Z., Wu, J., et al. (2021). Development of EST-SSR markers and association analysis of floral scent in tree peony. *Sci. Hortic.* 289, 110409. doi: 10.1016/j.scienta.2021.110409
- Martina, M., Acquadro, A., Barchi, L., Gulino, D., Brusco, F., Rabaglio, M., et al. (2022). Genome-wide survey and development of the first microsatellite markers database (AnCorDB) in *Anemone coronaria* L. *Int. J. Mol. Sci.* 23, 3126. doi: 10.3390/ijms23063126
- Mnenedy, E., Mantell, S., and Bennett, M. (2001). Use of random amplified polymorphic DNA (RAPD) markers to reveal genetic diversity within and between populations of cashew (*Anacardium occidentale* L.). *J. Hortic. Sci. Biotechnol.* 76, 375–383. doi: 10.1080/14620316.2001.11511380
- Negi, A., Singh, K., Jaiswal, S., Kokkat, J. G., Angadi, U. B., Iqbal, M. A., et al. (2022). Rapid genome-wide location-specific polymorphic SSR marker discovery in black pepper by GBS approach. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.846937
- Öztürk, S. C., Göktay, M., Allmer, J., Doğanlar, S., and Frary, A. (2018). Development of simple sequence repeat markers in hazelnut (*Corylus avellana* L.) by next-generation sequencing and discrimination of Turkish hazelnut cultivars. *Plant Mol. Biol.* 36, 800–811. doi: 10.1007/s11105-018-1120-0
- Parida, S. K., Verma, M., Yadav, S. K., Ambawat, S., Das, S., Garg, R., et al. (2015). Development of genome-wide informative simple sequence repeat markers for large-scale genotyping applications in chickpea and development of web resource. *Front. Plant Sci.* 6, 645. doi: 10.3389/fpls.2015.00645
- Patil, P. G., Singh, N. V., Parashuram, S., Bohra, A., Mundewadikar, D. M., Sangnure, V. R., et al. (2020). Genome wide identification, characterization and



- validation of novel miRNA-based SSR markers in pomegranate (*Punica granatum* L.). *Plant Mol. Biol.* 26 (4), 683–696. doi: 10.1007/s12298-020-00790-6
- Patil, P. G., Singh, N. V., Bohra, A., Raghavendra, K. P., Mane, R., Mundewadikar, D. M., Sharma, J., et al. (2021). Comprehensive characterization and validation of chromosome-specific highly polymorphic SSR markers from pomegranate (*Punica granatum* L.) cv. *Tunisia genome*. *Frontiers in Plant Science* 12, 645055. doi: 10.3389/fpls.2021.645055
- Perrier, X., and Jacquemond-Collet, J. P. (2006) *DARwin software*. Available at: <http://darwin.cirad.fr/darwin/>.
- Portis, E., Lanteri, S., Barchi, L., Portis, F., Valente, L., Toppino, L., et al. (2018). Comprehensive characterization of simple sequence repeats in eggplant (*Solanum melongena* L.) genome and construction of a web resource. *Front. Plant Sci.* 9, 401. doi: 10.3389/fpls.2018.00401
- Qin, Y., Shi, G., and Sun, Y. (2013). Evaluation of genetic diversity in *Pampus argenteus* using SSR markers. *Genet. Mol. Res.* 12, 5833–5841. doi: 10.4238/2013.November.22.10
- Qin, Z., Wang, Y., Wang, Q., Li, A., Hou, F., and Zhang, L. (2015). Evolution analysis of simple sequence repeats in plant genome. *PLoS One* 10 (12), e0144108. doi: 10.1371/journal.pone.0144108
- Ravishankar, K. V., Bommesetty, P., Bajpai, A., Srivastava, N., Mani, B. H., Vasugi, C., et al. (2015). Genetic diversity and population structure analysis of mango (*Mangifera indica*) cultivars assessed by microsatellite markers. *Trees* 29, 775–783. doi: 10.1007/s00468-015-1155-x
- Samal, S., Rout, G. R., Nayak, S., Nanda, R. M., Lenka, P. C., and Das, P. (2003). Primer screening and optimization for RAPD analysis of cashew. *Biol. Plant* 46 (2), 301–304. doi: 10.1023/A:1022827416677
- Savadi, S., Muralidhara, B. M., Venkataravanappa, V., Adiga, J. D., Manjunatha, K., and Patil, B. (2022a). De novo transcriptome assembly and its utility in development and characterization of the first set of genic SSR markers in cashew. *Industrial Crops and Products* 189, 115734. doi: 10.1016/j.indcrop.2023.112233
- Savadi, S., Adiga, J. D., Muralidhara, B. M., Prasad, P., Manjunatha, K., Ashwitha, K., Manoj, K., et al. (2023). Discovery of genome-wide genetic variations and development of first set of InDel markers for genetics research in cashew. *Sci. Hortic.* 320, 112233. doi: 10.1016/j.scienta.2023.112233
- Savadi, S., Muralidhara, B. M., Godwin, J., Adiga, J. D., Mohana, G. S., Eradasappa, E., et al. (2022b). De novo assembly and characterization of the draft genome of the cashew (*Anacardium occidentale* L.). *Sci. Rep.* 12, 18187. doi: 10.1038/s41598-022-22600-7
- Savadi, S., Muralidhara, B. M., and Preethi, P. (2020a). Advances in genomics of cashew tree: molecular tools and strategies for accelerated breeding. *Tree Genet. Genomes* 16, 1–15. doi: 10.1007/s11295-020-01453-z
- Savadi, S., Prasad, P., Sharma, K., Rathore, R., Bhardwaj, S. C., Gangwar, O. P., et al. (2020b). Development of novel transcriptome-based SSR markers in *Puccinia triticina* and their potential application in genetic diversity studies. *Trop. Plant Pathol.* 45, 499–510. doi: 10.1007/s40858-020-00347-8
- Sharma, P., Gaur, V. K., Sirohi, R., Larroche, C., Kim, S. H., and Pandey, A. (2020). Valorization of cashew nut processing residues for industrial applications. *Ind. Crops Prod.* 152, 112550. doi: 10.1016/j.indcrop.2020.112550
- Shi, Y., Kamer, P. C., and Cole-Hamilton, D. J. (2019). Synthesis of pharmaceutical drugs from cardanol derived from cashew nut shell liquid. *Green Chem.* 21, 1043–1053. doi: 10.1039/C8GC03823F
- Silfverberg-Dilworth, E., Matasci, C. L., Van de Weg, W. E., Van Kaauwen, M. P. W., Walser, M., Kodde, L. P., et al. (2006). Microsatellite markers spanning the apple (*Malus x domestica* Borkh.) genome. *Tree Genet. Genomes* 2, 202–224. doi: 10.1007/s11295-006-0045-1
- Soares, T. N., Sant'ana, L. D. L., Oliveira, L. K. D., Telles, M. P. D. C., and Collevatti, R. G. (2013). Transferability and characterization of microsatellite loci in *Anacardium humile* A. St. Hil. (Anacardiaceae). *Genet. Mol. Res.* 12, 3146–3149. doi: 10.4238/2013.January.4.24
- Sonah, H., Deshmukh, R. K., Sharma, A., Singh, V. P., Gupta, D. K., Gacche, R. N., et al. (2011). Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in *Brachypodium*. *PLoS One* 6, e21298. doi: 10.1371/journal.pone.0021298
- Song, X., Li, N., Guo, Y., Bai, Y., Wu, T., Yu, T., et al. (2021). Comprehensive identification and characterization of simple sequence repeats based on the whole-genome sequences of 14 forest and fruit trees. *For. Res.* 1, 1–10. doi: 10.48130/FR-2021-0007
- Sun, L., Yang, W., Zhang, Q., Cheng, T., Pan, H., Xu, Z., et al. (2013). Genome-wide characterization and linkage mapping of simple sequence repeats in *mei* (*Prunus mume* Sieb. et Zucc.). *PLoS One* 8, e59562. doi: 10.1371/journal.pone.0059562
- Taheri, S., Lee, A. T., Yusop, M. R., Hanafi, M. M., Sahebi, M., Azizi, P., et al. (2018). Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants. *Molecules* 23, 399. doi: 10.3390/molecules23020399
- Talibersky, J., Polaczek, J., Ramamoorthy, R., and Shishlov, O. (2009). Phenols from cashew nut shell oil as a feedstock for making resins and chemicals. *Oil Gas Eur. Mag.* 35, 33–39.
- Tan, C., Zhang, H., Chen, H., Guan, M., Zhu, Z., Cao, X., et al. (2023). First report on development of genome-wide microsatellite markers for stock (*Matthiola incana* L.). *Plants* 12, 748. doi: 10.3390/plants12040748
- Thimmappaiah, Santhosh, W. G., Shobha, D., and Melwyn, G. S. (2009). Assessment of genetic diversity in cashew germplasm using RAPD and ISSR markers. *Sci. Hortic.* 120, 411–417. doi: 10.1016/j.scienta.2008.11.022
- Vieira, M. L. C., Santini, L., Diniz, A. L., and Munhoz, C. D. F. (2016). Microsatellite markers: what they mean and why they are so useful. *Genet. Mol. Biol.* 39, 312–328. doi: 10.1590/1678-4685-GMB-2016-0027
- Visalakshi, M., Jawaharlal, M., and Ganga, M. (2011). “Cashing in on cashew,” in *I International symposium on cashewnut*, Vol. 1080. 103–107.
- Wang, H., Gao, S., Liu, Y., Wang, P., Zhang, Z., and Chen, D. (2022). A pipeline for effectively developing highly polymorphic simple sequence repeats markers based on multi-sample genomic data. *Ecol. Evol.* 12, e8705. doi: 10.1002/ece3.8705
- Wang, X., Yang, S., Chen, Y., Zhang, S., Zhao, Q., Li, M., et al. (2018). Comparative genome-wide characterization leading to simple sequence repeat marker development for *Nicotiana*. *BMC Genom.* 19, 500. doi: 10.1186/s12864-018-4878-4
- Zhang, Q., Li, J., Zhao, Y., Korban, S. S., and Han, Y. (2012a). Evaluation of genetic diversity in Chinese wild apple species along with apple cultivars using SSR markers. *Plant Mol. Biol.* 30, 539–546. doi: 10.1007/s11105-011-0366-6
- Zhang, Q., Ma, B., Li, H., Chang, Y., Han, Y., Li, J., et al. (2012b). Identification characterization and utilization of genome-wide simple sequence repeats to identify a QTL for acidity in apple. *BMC Genom.* 13, 1–12. doi: 10.1186/1471-2164-13-537
- Zhao, S., Guan, L., Yan, Z., He, N., Lu, X., and Liu, W. (2012b). Fingerprinting of Zhengkang triploid seedless watermelon varieties using SSR markers. *China Cucurbits Vegetables* 25, 1–4.
- Zhao, Y., Prakash, C. S., and He, G. (2012a). Characterization and compilation of polymorphic simple sequence repeat (SSR) markers of peanut from public database. *BMC Res. Notes* 5, 1–7. doi: 10.1186/1756-0500-5-362
- Ziya Motalebipour, E., Kafkas, S., Khodaeiaminjan, M., Coban, N., and Gozel, H. (2016). Genome survey of pistachio (*Pistacia vera* L.) by next generation sequencing: Development of novel SSR markers and genetic diversity in *Pistacia* species. *BMC Genom.* 17, 998. doi: 10.1186/s12864-016-3359-x