Journal of Applied Genetics (2020) 61:367–377
https://doi.org/10.1007/s13353-020-00561-9

PLANT GENETICS • ORIGINAL PAPER

Genome-wide development of miRNA-based SSR markers in *Cleistogenes songorica* and analysis of their transferability to Gramineae/non-Gramineae species

Gisele Kanzana[1] · Yufei Zhang[1] · Tiantian Ma[1] · Wenxian Liu[1] · Fan Wu[1] · Qi Yan[1] · Xueyang Min[1] · Zhuanzhuan Yan[1] · Blaise Pascal Muvunyi[1] · Jie Li[1] · Zhengshe Zhang[1] · Yufeng Zhao[1] · Jiyu Zhang[1]

Received: 27 October 2019 / Revised: 28 April 2020 / Accepted: 7 May 2020 / Published online: 7 June 2020

Abstract

Simple sequence repeat (SSR) markers are commonly used for many genetic applications, such as map construction, fingerprinting, and genetic diversity analyses, due to their high reproducibility, polymorphism, and abundance. Endogenous miRNAs play essential roles in plant development and gene expression under diverse biotic and abiotic stress conditions. In the present study, we predicted 110 miRNA-SSR primer pairs from 287 precursor miRNAs. Among 110 primer pairs, 85 were successfully amplified and examined for transferability to other Gramineae and non-Gramineae species. The results showed that all 82 primer pairs yielded unambiguous and strong amplification, and across the 23 studied *Cleistogenes* accessions, a total of 385 alleles were polymorphic. The number of alleles produced per primer varied from 3 to 11, with an average of 4.69 per locus. The expected heterozygosity (He) ranged from 0.44 to 0.88, with an average of 0.74 per locus, and the PIC (Polymorphism Information Content) values ranged from 0.34 to 0.87, with an average of 0.69 per locus. Furthermore, 1422 miRNA target genes were predicted and analyzed using the GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) databases. In conclusion, the results showed that an miRNA-based microsatellite marker system can be applicable for genetic diversity and marker-assisted breeding studies.

Keywords *Cleistogenes songorica* · MicroRNA · Simple sequence repeat · Genetic diversity · Transferability · Target genes

Communicated by: Izabela Pawłowicz

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13353-020-00561-9) contains supplementary material, which is available to authorized users.

✉ Jiyu Zhang
zhangjy@lzu.edu.cn

Gisele Kanzana
kanzana16@lzu.edu.cn

Yufei Zhang
zhangyf16@lzu.edu.cn

Tiantian Ma
matt17@lzu.edu.cn

Wenxian Liu
liuwx@lzu.edu.cn

Fan Wu
wuf15@lzu.edu.cn

Qi Yan
yanq16@lzu.edu.cn

Xueyang Min
yanq16@lzu.edu.cn; minxy15@lzu.edu.cn

Zhuanzhuan Yan
yanzhzh16@lzu.edu.cn

Blaise Pascal Muvunyi
muvunyi14@lzu.edu.cn

Jie Li
lij17@lzu.edu.cn

Zhengshe Zhang
zhangzhsh14@lzu.edu.cn

Yufeng Zhao
zhaoyf17@lzu.edu.cn

[1] State Key Laboratory of Grassland Agro-Ecosystems, Key Laboratory of Grassland Livestock Industry Innovation, Ministry of Agriculture and Rural Affairs, Engineering Research Center of Grassland Industry, Ministry of Education, College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou 730020, People's Republic of China

Introduction

Cleistogenes songorica (*C. songorica*), which belongs to the Gramineae family, is an important perennial forage and ecological grass in Northwest China, including in Inner Mongolia, where the average annual rainfall is 110 mm (Yang et al. 2001). It exhibits high feeding value, cold resistance, and drought tolerance (Zhang et al. 2014) and has been domesticated as a turf grass cultivar, *C. songorica* (Roshev.) cv. Tenggeli. Thus, *C. songorica* has the potential to play a critical role in landscaping, vegetation, and desert ecosystem restoration. To study the drought tolerance mechanism of *C. songorica*, leaf and root expression sequence tag (EST) resources have been used to investigate its drought stress-responsive genes (Zhang et al. 2011). Some of these genes have been transferred into *Arabidopsis thaliana* and alfalfa to confirm their contribution to the stress tolerance of these plants (Duan et al. 2015; Zhang et al. 2016).

Simple sequence repeat (SSR) markers, also known as microsatellites, are one of the most variable types of short repetitive elements of 1–6 bases (Chen et al. 2009; Sujatha 2013) and are found in prokaryotic and all eukaryotic genomes (Schlötterer 2000). SSRs have many important biological functions, such as the regulation of chromatin organization, DNA metabolic processes, gene activity, and RNA structure (Haasl and Payseur 2012; Li et al. 2004; Li et al. 2002). SSR_S exhibit high polymorphism, abundance, and genetic diversity and tend to be codominant (Liu et al. 2016; Liu et al. 2010; Ni et al. 2002; Noli et al. 2008; Parveen et al. 2016). In plants, SSRs are widely used, for example, in genetic diversity analysis, germplasm identification, comparative genetics analysis, phylogenetic analysis, QTL analysis, linkage mapping, and marker-assisted selection (Cavagnaro et al. 2010; Liu et al. 2010; Mondal et al. 2015; Rakoczy-Trojanowska and Bolibok 2004; Saha et al. 2006). With the identification of increasing numbers of SSRs, SSRs are widely used to overcome the restrictions associated with other types of markers for genome mapping, fingerprinting, and population genetics studies as well as in molecular breeding.

MicroRNAs (miRNAs) are a class of noncoding RNA 18–24 nucleotides in length. Endogenous small RNAs play essential roles in plant development, phase transition (Khraiwesh et al. 2012), and gene expression under diverse biotic and abiotic stress conditions and in different developmental stages of the life cycle (Ganie and Mondal 2015; Singh et al. 2017). Like small nucleolar RNAs (snoRNAs), short interfering RNAs (siRNAs), and piwi-interacting RNAs (piRNAs), miRNAs are involved in the regulation of gene expression (Lin et al. 2008; Neilson and Sharp 2008). miRNAs are highly conserved in both plants and animals and have been found in plants, green algae, viruses, fungi, and older lineages of animals (Bartel and Bartel 2003; Saini et al. 2008). In addition, some species-specific miRNAs exist, regulating various developmental and biological processes (Fahlgren et al. 2010). Molecular markers or DNA markers have become efficient tools for identification of polymorphisms among different genotypes or genes (Jiang 2015) and are increasingly used in plant molecular research. For example, DNA fingerprinting can be used to detect polymorphisms among individuals and has become a fundamental tool for crop improvement via plant breeding methods (Ahmad et al. 2010). DNA markers can be categorized into two types: non-polymerase chain reaction (PCR)-based markers that identify restriction fragment length polymorphisms (RFLPs) and PCR-based markers. The latter type includes single nucleotide polymorphism (SNP) markers (Jin et al. 2003), intron length polymorphism (ILP) markers (Zhang et al. 2017), random amplified polymorphic DNA (RAPD) markers, amplified fragment length polymorphism (AFLP) markers (Bandelj et al. 2004), and SSR markers (Wu et al. 2016) (Baraket et al. 2011; Min et al. 2017; Zhang et al. 2012). All of these DNA-based markers have been used in various genetic studies. The selection of the appropriate markers depends on the study objectives. The abundance, low cost, high polymorphism, heritability, multiallelic nature, reproducibility, distribution throughout the genome, ease of use, and generally codominant nature of SSR markers make them highly suitable for genetic diversity studies (Cloutier et al. 2011; Kessuwan et al. 2016; Smulders and De Klerk 2011; Wassom et al. 2008; Wiesner et al. 2001). The potential significance of SSRs in the miRNA gene, i.e., miRNA-SSRs, has been considered in the development of miRNA-associated markers to study traits in *C. songorica* and other species. In the present study, we identified miRNA-SSRs in full genomic sequences of *C. songorica* pre-miRNAs. The previous report on *C. songorica* surveyed 44 late embryogenesis abundant (LEA) family genes in the *C. songorica* genome and analyzed the responses of four selected LEA genes to heat stress or abscisic acid (ABA) treatment in leaves and shoots (Muvunyi et al. 2018). Based on the morphological structure of the flower, *C. songorica* is a dimorphic plant that can produce both chasmogamous (CH) and cleistogamous (CL) flowers. This study examined miRNAs and target to illuminate the potential miRNA-mediated regulatory mechanism on CH and CL in *C. songorica* (Xu et al. 2020); in addition, miRNAs and mRNA were expressed in the flowering processes (Wu et al. 2018). Moreover, in *C. songorica*, a coexpression network based on the lncRNAs, miRNAs, protein-coding (PC) genes, and transcription factors under water stress and during recovery was constructed. lncRNAs, miRNAs, PC genes, and transcription factors constitute a complex transcriptional regulatory network in which lncRNAs can regulate PC genes and miRNAs under water stress and recovery (Yan et al. 2019).

The aim of this study is (i) the generation of an miRNA-derived PCR-based marker set carrying polymorphic SSR loci, (ii) validation of this set to a reference germplasm, and (iii)

analysis of its transferability to other related and unrelated germplasms, including reference Gramineae and non-Gramineae species. The development of miRNA-SSR markers based on *C. songorica* and their transferability to other species has not been reported so far, and they have great potential for studies on genetic diversity and transferability of miRNA-SSR to other Gramineae and non-Gramineae species. This study will be helpful for marker-assisted genetic improvement, for genotyping applications and in QTL analysis, and molecular-assisted selection studies for plant breeders and other researchers.

Materials and methods

Plant material and DNA extraction

The plant materials contained Gramineae and non-Gramineae species. The Gramineae species were *Cleistogenes* Keng, wheat (*Triticum aestivum)*, ryegrass (*Lolium perenne*), rice (*Oryza sativa*), and maize (*Zea mays*), while the non-Gramineae species were *Arabidopsis thaliana*, *Medicago truncatula*, alfalfa (*Medicago sativa*), yellow sweet clover (*Melilotus officinalis*), common vetch (*Vicia sativa*), and soybean (*Glycine max*). All 11 species were used to examine the transferability of Cs-miRNA-SSR primers, and the 23 *Cleistogenes* accessions were used to analyze genetic diversity. In the 23 *Cleistogenes* accessions, one was cultivated type and the other 22 were wild types. There were 19 accessions of *Cleistogenes songorica*, one accession of *Cleistogenes caespitosa*, one accession of *Cleistogenes hackelii*, one accession of *Cleistogenes hancei*, and one accession of *Cleistogenes squarrosa*. Twenty-two accessions were obtained from China (Gansu, Inner Mongolia, and Shandong), and one was obtained from Mongolia (Table S1).

Young leaves from one plant per accession (~ 1 g) were collected and bulked as one sample. The samples were packed in aluminum foil, frozen in liquid nitrogen, and stored at − 80 °C until DNA extraction. Genomic DNA was extracted from the young leaf tissues using the sodium dodecyl sulfate (SDS) method (Shan et al. 2011). The extracted DNA was detected by agarose gel electrophoresis; then, the DNA samples were diluted with ddH_2O and stored at − 20 °C. The DNA quality and quantity were checked by 2% agarose gel electrophoresis and spectrophotometric measurement using a NanoDrop ND 1000.

Identification of SSRs, miRNA-SSR primer design for *C. songorica*, and chromosome mapping

A total of 287 pre-miRNAs of *C. songorica* were extracted from the *Cleistogenes songorica* local genome and used for the identification and extraction of SSRs using a Perl 5 script (*MISA*, MIcroSAtellite identification tool). Among the 287 pre-miRNAs, 110 pre-miRNA sequences were selected and used as queries for designing primers flanking repeats. Primers were designed using Primer3 software and Perl 5 interface modules. To design the SSRs in this study, pre-miRNA sequences with 100% nucleotide identity to the *C. songorica* genome sequence were extracted from the genome assembly with 500 bp of flanking regions (Fig. S1). The minimum length criteria were as follows: 10 repeat units for mononucleotides, 6 for dinucleotides, and 5 for tri-, tetra-, penta-, and hexa-nucleotides (Table S2). The designed miRNA-SSR primers were synthesized by Shanghai Sangon Biological Engineering Technology (Shanghai, China). The primer design parameters were as follows: amplicon size, 100–350 bp; primer length, 18–27 bases with 20 as the optimum; annealing temperature, 57–63 °C with the optimum of 60 °C; GC content, 45–50% (Liu et al. 2015). *C. songorica* markers were mapped to the genome assembly using TBtools (https://github.com/CJ-Chen/TBtools).

Prediction of miRNA target genes and GO analysis

The pre-miRNA sequences and miRNA target genes were predicted in accordance with previous study methods (Wu et al. 2018). Small RNA tags were aligned to the *C. songorica* genome sequence using Bowtie software to find miRNA. Conserved miRNAs and pre-miRNA were identified through Blast against the miRNA database, miRBase 19.0 (http://www.mirbase.org/). The miRNA sequences found were used as queries by using plant small RNA target analysis server to predict miRNA target genes with the following values and parameters (http://plantgrn.noble.org/psRNATarget/): (1) translational inhibition range from 10–11 nucleotides, (2) maximum expectation value: 5, (3) maximum mismatches at the complementary site: ≤ 1, (4) multiplicity of target sites: 2, (5) length for complementarity scoring (HSP size): 19, and (6) maximum energy to unpair the target site (UPE): 25.0. These target genes were used for bioinformatics analysis, and the GO and KEGG pathway analyses were performed using Biocloud software (https://www.biocloud.net).

PCR amplification

PCR was performed in 10 μl reaction mixtures of 4.95 μl 2 × reaction mix (dNTPs at 500 μM each, 20 mM Tris-HCl, 100 mM $MgCl_2$, 100 mM KCl_2, and 3 mM $MgCl_2$), 2.0 μl ddH_2O, 1.0 μl template DNA, 1.0 μg forward primers, 1.0 μg reverse primers, and 0.05 μl Gold DNA polymerase. Amplifications were performed with pre-denaturation for 3 min at 94 °C, 35 cycles of 30 s at 94 °C, 30 s at 60 °C, 30 s at 72 °C, 30 s at 94 °C, 30 s at 56 °C, and a final elongation step of 7 min at 72 °C. The PCR products were visualized on 2% agarose gels using Gel Red staining on 6% non-denaturing polyacrylamide gel using silver staining.

Statistical analysis

The SSR marker profiles were scored in a binary format, where alleles were indicated as absent (0) or present (1) for the corresponding bands among different *Cleistogenes* accessions. Individual bands that could be clearly scored were used for the genetic diversity analysis. A dendrogram was constructed from a genetic identity matrix using NTSYS-pc V.2.1 software and the unweighted pair group method arithmetic mean analysis (UPGMA) method. The significance of each node was evaluated by bootstrapping data over a locus for 1000 replications of the original matrix. The genetic similarity analyses were performed using NTSYS-pc, and the pairwise similarities were obtained using Jaccard's coefficients. The matrices of the resemblance coefficients were subjected to UPGMA to estimate the genetic similarity among the accessions and arrange the dendrogram. The genetic structures of twenty-three accessions were analyzed using Cs-miRNA-SSR markers and STRUCTURE software. A principal components analysis (PCA) was used to compare the overall changes in the population structure of the accessions (Varshney et al. 2017). Ordination was performed with the "vegan" package and plotted with the "ggplot2" package of R statistical software. We transformed binary data from the amplified fragments of 23 accessions using Hellinger.

Validation of PCR products by sequencing

To confirm the purity and identity of the PCR amplification product, according to the presence of single bands and high amplification efficiency, we selected some PCR amplification products sequenced by the Shanghai Sangon Biotech Company. Chromas software (http://technelysium.com.au/wp/chromas/) was used for manual curation of the chromatograms.

Results

The development and frequencies of miRNA-SSRs in the *C. songorica* genome

A total of 287 pre-miRNAs were predicted from the *C. songorica* local genome database, and 110 pre-miRNAs were used as queries to identify SSRs, yielding 125 SSRs. The microsatellite search results revealed 0.88 kb distribution frequency of one SSR per locus. Examination of the SSR motifs in the SSRs containing pre-miRNA sequences revealed that 27 MIR genes contained more than one SSR motif. Among the 125 total SSRs, 110 contained simple repeat motifs, whereas the remaining 15 contained compound motifs. Among the simple repeat motifs, mononucleotide motifs were most abundant (61, 48.80%), followed by dinucleotide motifs (46, 36.80%) and trinucleotide motifs (3, 2.40%) (Table 1). No tetra-, penta-, or hexa-nucleotide repeats were found in any of the *C. songorica* pre-miRNA flanking sequences. In the 110 pre-miRNAs containing SSRs, mononucleotides occurred at the highest frequency (59%), being present in 58 miRNA genes, followed by dinucleotides (38%), present in 38 genes, and trinucleotides (3%), present in only 3 miRNA genes (Fig. 1a, b).

Table 1 A summary of SSR search results

Search items	Numbers
Total number of pre-miRNA examined	287
Number of pre-miRNAs containing SSRs	110
Number of pre-miRNAs containing more than 1 SSR	27
Total number of identified SSRs	125
Number of SSRs present in compound formation	15
Number of SSRs containing a simple repeat motif	110
Repeat type	
Mononucleotide	61
Dinucleotide	46
Trinucleotide	3
Frequency of SSRs	One per 0.88 kb

The mononucleotide repeats shown were the most common type of repeat; $(A)_{27}$ occurred at the highest frequency (24.5%) in miRNA genes, followed by $(T)_{25}$ (22.7%), $(C)_5$ (4.5%), and $(G)_4$ (3.6%). Among the dinucleotides, $(CT)_{16}$ occurred at the highest frequency (14.5%) in the SSR-containing miRNA genes, followed by $(TC)_7$ (6.3%), $(AG)_6$ (5.4%), $(TA)_4$ (3.6%), $(GA)_4$ (3.6), $(AT)_3$ (2.7%), $(CA)_2$ (1.8%), $(GT)_2$ (1.8%), and $(TG)_1$ and $(AC)_1$ (both 0.9%). Trinucleotide repeats occurred at low frequencies in the miRNA genes in the *C. songorica* genome, with $(CCT)_1$, $(TTG)_1$, and $(AAT)_1$ each occurring at a frequency of 0.9% (Fig. 1c). Our study showed that mononucleotides were the most common type of repeat motif, whereas trinucleotides were the least common.

Screening of Cs-miRNA-SSR primers, analysis of their transferability to other species, and sequencing of PCR amplification products

All 110 Cs-miRNA-SSR primer pairs were screened for PCR amplification from the genomic DNA of *C. songorica*. Among the 110 primer pairs, 85 were successfully amplified, and 25 failed to be amplified. The 85 successfully amplified pairs were assessed for transferability to eleven Gramineae or non-Gramineae species. More than 70% of them yielded amplification products of the expected size (150–300 bp); the PCR products generated by the other primer pairs were shorter or longer than predicted. PCR products were then selected according to the presence of single bands and high amplification efficiency. In addition, PCR amplicons of several Cs-miRNA-SSR primers were amplified and sequenced. Primer

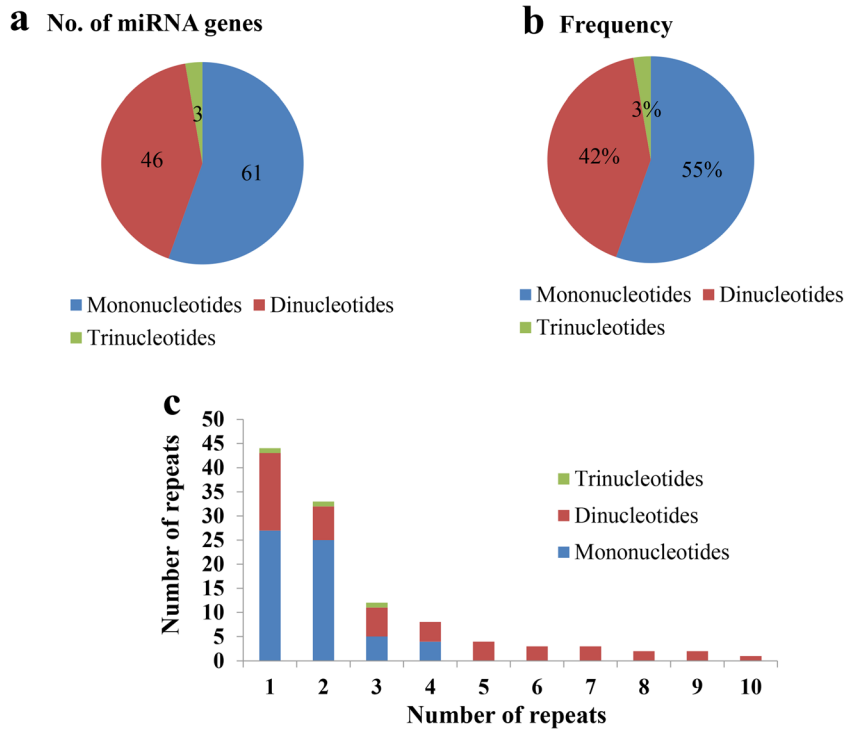

Fig. 1 Number of miRNA genes with SSR motifs and abundance of miRNA-SSR motifs in *Cleistogenes songorica*. **a** Number of miRNA genes possessing SSR motifs. **b** Frequency of different repeat motifs in miRNA genes. **c** Number of repeats having the most abundant miRNA-SSR motifs

pairs Cs-miRNA-SSR-27, Cs-miRNA-SSR-82, Cs-miRNA-SSR-108, and Cs-miRNA-SSR-67, which have trinucleotide repeat $(TTG)_5$, dinucleotide repeat $(TC)_{11}$, and mononucleotide repeats $(A)_9$ and $(T)_4$, respectively, were matched to the locus used for marker development (Fig. S2).

The highest amplification percentage (98%) was observed in *Cleistogenes songorica*, and the lowest (41.2%) was observed in *Glycine max*. The average of transferability (amplification percentages of Cs miRNA-SSR markers in the germplasms) across the various species was 55.93% (Table 2). The amplification in *Arabidopsis thaliana* and *Vicia sativa*, which are non-Gramineae species, was higher than in *Triticum aestivum* and *Zea mays*, two Gramineae species. *C. songorica* showed higher polymorphism of miRNA-SSR markers than the other species analyzed in this study. A total of 466 alleles were detected using 85 Cs-miRNA-SSR loci ranging in frequency from 2 to 10 per locus. The three primer pairs numbered 33, 43, and 92 yielded the highest number of alleles (10), and the lowest numbers of alleles were obtained from ten primer pairs (nos. 7, 8, 11, 12, 17, 20, 23, 24, 28, and 36). The expected heterozygosity (He) ranged from 0.44 to 0.88, with an average of 0.74 per locus, and the PIC values ranged from 0.34 to 0.87, with an average of 0.69 per locus (Table S3).

Table 2 Transferability of the 85 Cs-miRNA-SSR markers to the 11 Gramineae and non-Gramineae species

No.	Genus/species	Transferability (%)
1	*Cleistogenes songorica*	98.00
2	*Oryza sativa*	64.70
3	*Lolium perenne*	60.00
4	*Triticum aestivum*	51.76
5	*Zea mays*	49.41
6	*Arabidopsis thaliana*	55.30
7	*Vicia sativa*	52.94
8	*Medicago sativa*	48.23
9	*Medicago truncatula*	44.70
10	*Melilotus officinalis*	47.00
11	*Glycine max*	41.20
Mean		55.93

Genetic diversity, cluster, and structure analyses of *Cleistogenes* accessions

To evaluate the application potential of the Cs-miRNA-SSR markers to the study of genetic diversity among 23 *Cleistogenes* accessions, the 82 transferable primer pairs were analyzed. The results showed that all 82 primer pairs yielded unambiguous and strong amplification, and across the 23 accessions, a total of 385 alleles were polymorphic. The number of alleles produced per primer varied from 3 to 11, with an average of 4.69 per locus. Cs-miRNA-SSR 45 and Cs-miRNA-SSR 75 had the lowest PIC values (both 0.41), and Cs-miRNA-SSR 43 had the highest PIC value of 0.86. The average PIC value was 0.62 (Table S4), and values greater than 0.5 indicate a high level of polymorphism of these markers and suggest their potential for genetic diversity and genetic mapping analyses. Due to the genetic closeness of

Cleistogenes songorica with other *Cleistogenes* accessions, the genetic marker array between *Cleistogenes songorica* and other *Cleistogenes* accessions could be evaluated and further used for its genetic diversity analysis. The banding patterns of 23 *Cleistogenes* accessions obtained with Cs-miRNA-SSR-50 and Cs-miRNA-SSR-84 markers are portrayed in Fig. S3. As SSRs are highly polymorphic markers, miRNA-SSRs were studied, and they produced 8-11 alleles (Fig. S4).

In the cluster analysis, the twenty-three *Cleistogenes* accessions were grouped and divided into three clusters: cluster I had one accession (Acc01), cluster II contained 19 accessions (Acc02 to Acc20), and cluster III contained three accessions (Acc21, Acc22, and Acc23) (Fig. 2). In the population structure analysis, which examined $K = 1–10$, the optimal number of groups was three based on the maximum likelihood and delta K (ΔK) values, consistent with the cluster results that indicating that the 23 accessions belonged to three groups (Fig. 3). Among the three groups, group I includes eighteen accessions (Acc01 and Acc03-Acc19), group II includes one accession (Acc02), and group III contains 4 accessions (Acc20-Acc23).

The PCA showed that 5 *Cleistogenes* accessions, including Acc02 and Acc20, were outside the group of other *C. songorica* accessions. However, the remaining 18 clustered together into one group, and some of the accessions are densely displayed in Fig. 4 because of the closeness of their genetic relationships. Moreover, accession 1, which is *C. caespitosa*, was in the group (Fig. 4). Although *C. hackelii* and *C. hancei* clustered in the same group, their genetic distance was large.

Distribution of miRNA-SSRs on *Cleistogenes songorica* chromosomes

Among the 110 pre-miRNAs of the *C. songorica* genome, all 110 Cs-miRNA-SSRs were identified. Although the 110 Cs-miRNA-SSR markers were expected to physically map to 20 chromosomes, only 90 Cs-miRNA-SSRs were mapped to 19 chromosomes. The remaining 20 Cs-miRNA-SSR markers were located on scaffold regions, and no marker mapped to chromosome 3. Chromosome 12 contained the highest number of markers (12; 13.33%), whereas chromosomes 19 and 20 contained the lowest numbers of markers (1 each; 1.11%) (Fig. S5).

Gene ontology and KEGG pathway analysis

To examine the potential functions of the miRNA target genes, GO and KEGG analyses were performed. Annotations of the 1422 predicted target genes were analyzed, and 1235 genes were annotated and classified based on GO terms. The target genes were classified into three categories: biological processes (22 GO terms), cellular components (15 GO terms), and molecular functions (15 GO terms) (Fig. 5). Out of the 1422 miRNA target genes and 54,383 of all genes found in *Cleistogenes songorica* genome, 1235 miRNA target genes and 37,025 of all genes in the genome were annotated and classified into biological processes, cellular components, and molecular functions (Fig. 5). Generally, the number of miRNA genes has been represented with a high percentage of genes compared with the number of genes in the three

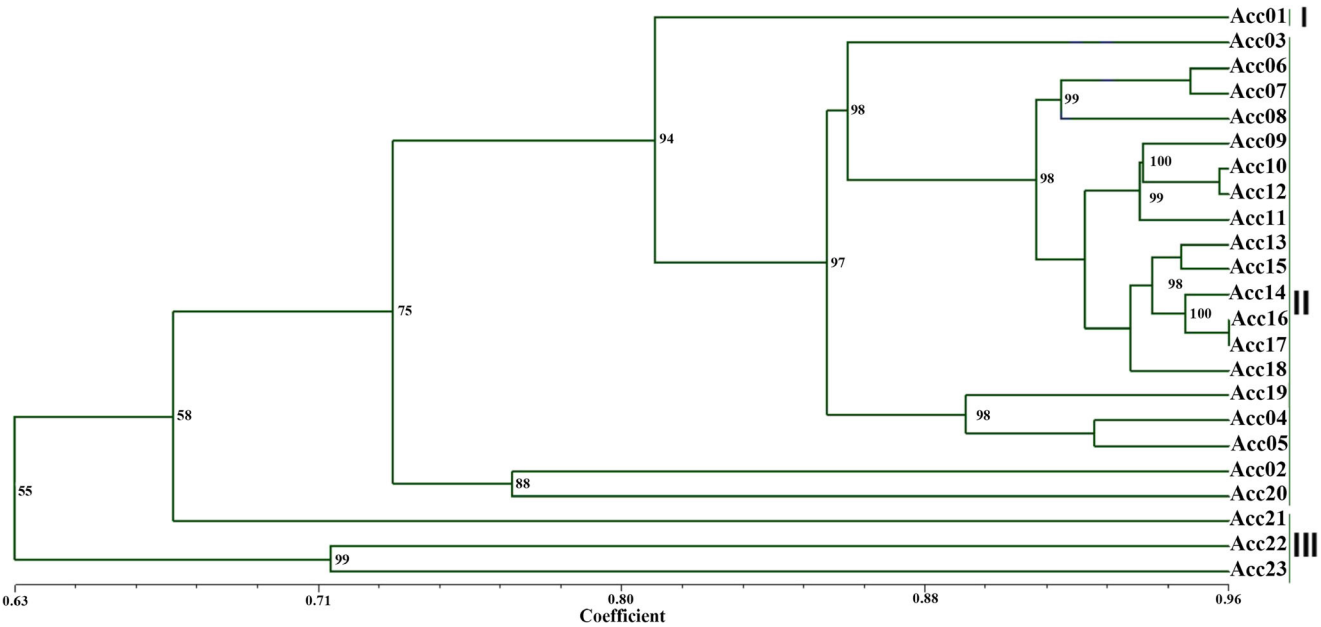

Fig. 2 Dendrogram of the unweighted pair group method arithmetic mean analysis (UPGMA). Acc01 = *C. caespitosa*, Acc03 to Acc20 = *C. songorica*, Acc21 = *C. hackelii*, Acc22 = *C. hancei*, and Acc23 = *C. squarrosa*. The numbers denote bootstrap values

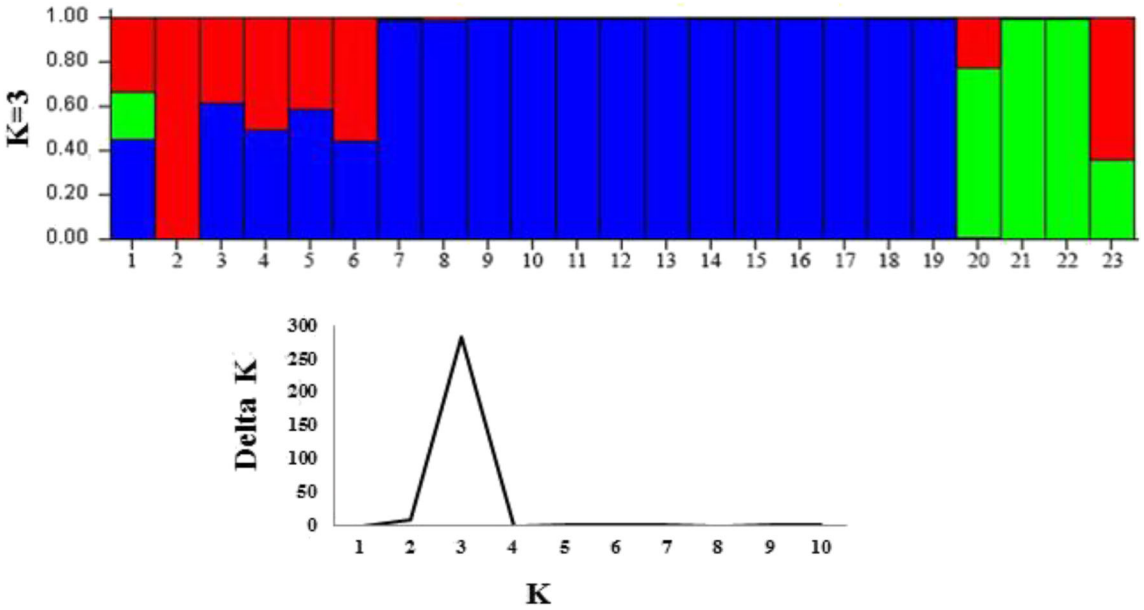

Fig. 3 Histogram structure of the Cs miRNA-SSR molecular markers data set for the model (showing the highest delta K). The colors represent the 3 groups clustered, the *x*-axis denotes 23 *Cleistogenes* accessions, and the *y*-axis $K = 3$

categories. However, in the biological processes category, detoxification, and biological adhesion where the GO terms that were represented with a higher number of genes than the miRNA genes of all genes in the genome. The organelle parts, cellular junction, and extracellular region GO terms represented a high percentage of all genes versus miRNA target genes contained in cellular components. In the molecular function category, all genes had a greater percentage than miRNA genes, as revealed in the antioxidant activity, molecular function regulator, signal transducer activity, molecular transducer activity, transcription factor activity protein binding, and nutrient reservoir activity. In the biological process category, the two most overrepresented GO terms were metabolic processes (648 genes) and cellular processes (634 genes), followed by single-organism processes (497 genes). In the cellular component category, the most overrepresented GO terms were cell

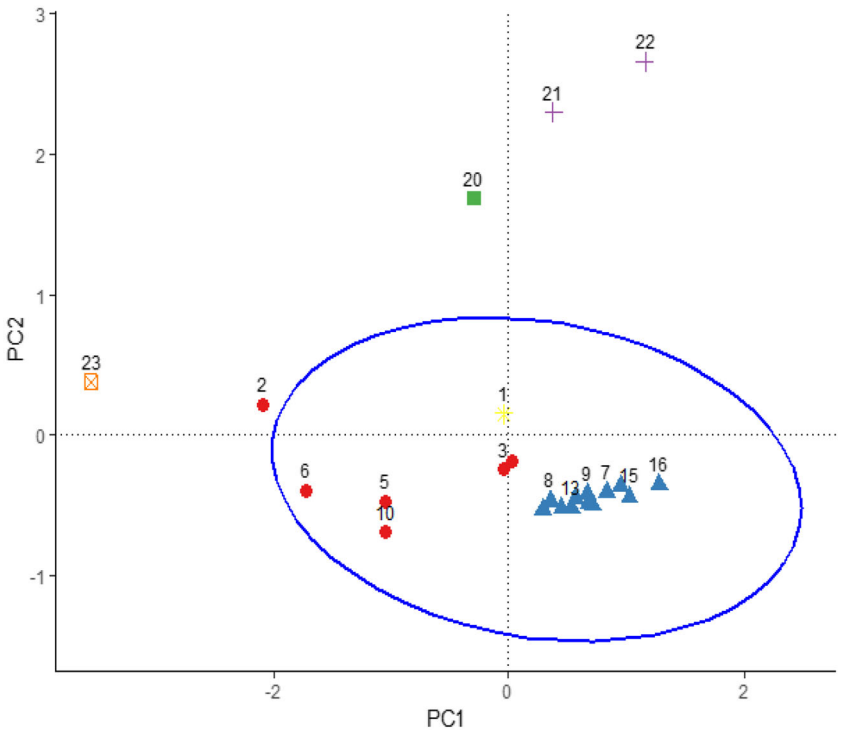

Fig. 4 Grouping of geographical origin and genetic distributions of 23 *Cleistogenes* accessions using 82 pairs of miRNA-SSR markers. In the principal component analysis (PCA) plot, each accession is represented by a vertical bar, and the length of each colored segment in each vertical bar represents the proportion contributed by ancestral populations. 1 = *C. caespitosa*, 2 to 20 = *C. songorica*, 21 = *C. hackelii*, 22 = *C. hancei*, and 23 = *C. squarrosa.* Red dots, blue triangles, and the green square denote *Cleistogenes songorica*; the yellow asterisk denotes *Cleistogenes caespitosa*; purple plus signs denote *Cleistogenes hackelii* and *Cleistogenes hancei*; and the orange box with an *x* denotes *Cleistogenes squarrosa*

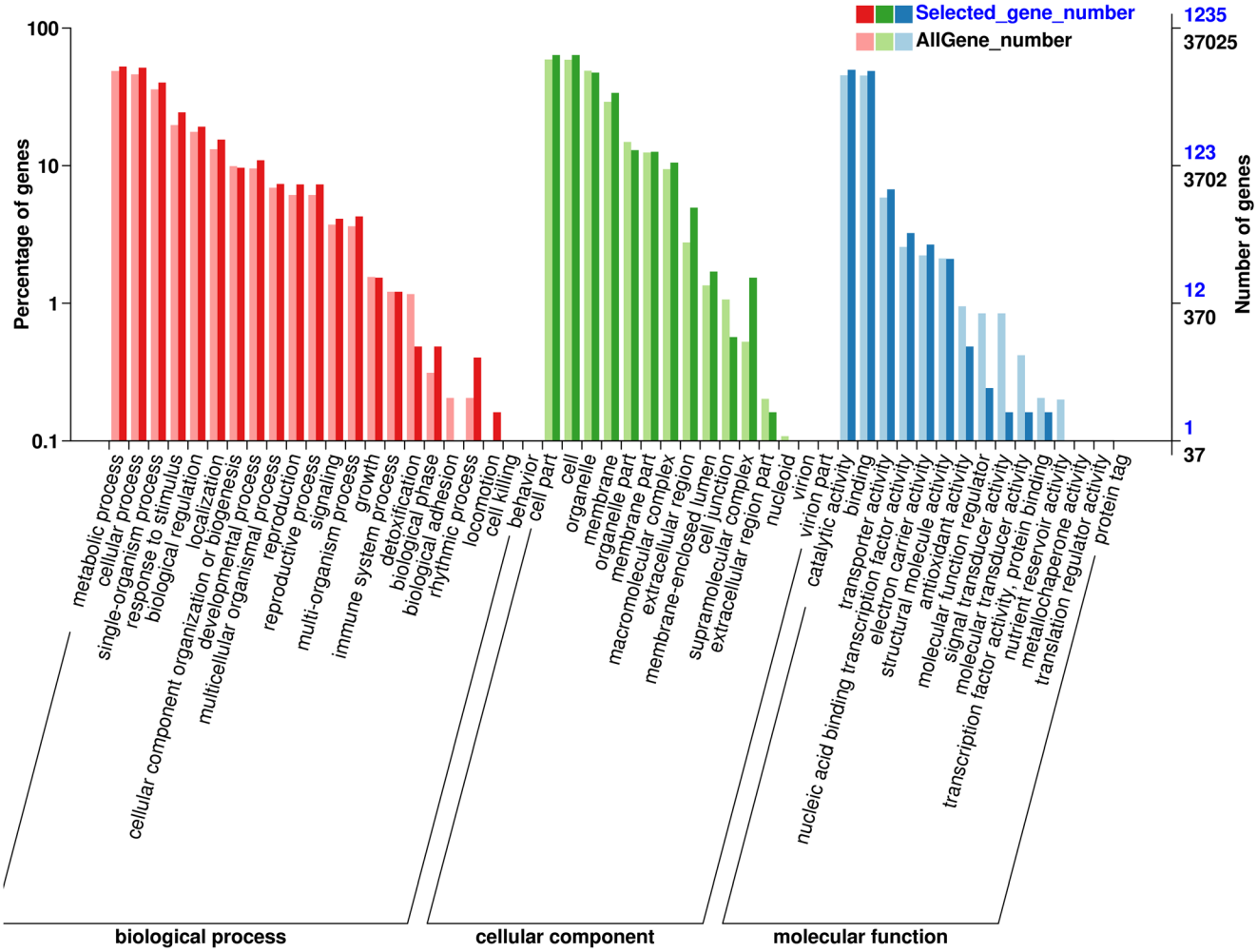

Fig. 5 The gene ontology analysis of identified and predicted miRNA target genes in *Cleistogenes songorica*. The relative frequencies of miRNA target genes assigned to the GO function in 3 categories. Red, green, and blue represent biological processes, cellular components, and molecular functions, respectively. The *x*-axis indicates the category name of the GO annotation. The right *y*-axis (blue) indicates the number of target genes in miRNA and (black) the number of all genes contained in the genome. The left *y*-axis indicates the percentage of a specific category of genes in that main category. The solid bars show the target genes of differentially expressed miRNAs while the hollow bars show genes from all genome

parts (786 genes) and cells (784 genes), followed by organelle (584 genes). In the molecular function category, the most frequent terms were catalytic activity (614 genes) and binding (600 genes) (Table S5).

The KEGG pathway analysis showed that 1422 miRNA target genes were annotated in the 20 most represented pathways. The enrichment analysis of the differentially expressed miRNAs identified five signaling pathways with q values < 1.00. Each enriched pathway is displayed in a scatter diagram, with points representing the enrichment level, and the color corresponds to the number of genes enriched for the given pathway. A small q value (red) indicates the high significance of the pathway. Many genes were assigned to carbon metabolism, with others assigned to pyruvate metabolism, phagosome, glycolysis/gluconeogenesis, and other pathways (Fig. S6).

Discussion

In several studies, the flanking regions of pre-miRNA sequences, ranging from 300 to 1 kb, have been used to analyze various parameters of miRNA genes. To understand the origin and evolution of miRNA genes of ten different plants, including rice, Nozawa et al. 2012 used the 300-bp flanking region at either end of the pre-miRNA sequences. Similarly, to identify the SNPs in rice miRNA genes, Liu et al. 2013 used the 1-kb flanking region from both ends of the pre-miRNA genes. The flanking length used in this study is similar to the lengths used in other studies targeting pre-miRNA sequences. To date, we identified a total of 110 miRNA-SSRs from the 287 pre-miRNAs, and 110 pre-miRNA sequences were selected and used as queries for designing primers flanking repeats.

The previous studies demonstrated successful transferability of microsatellites that were investigated in different species (González-Martínez et al. 2004; Kijas et al. 1995; Decroocq et al. 2003). For instance, research on the confamilial transferability of SSR markers from cotton (*Gossypium hirsutum L.*) and jute (*Corchorus olitorius L.*) to 22 Malvaceous species found that despite not finding good prospects for transferability of SSR markers in *Abelmoschus*, *Malvastrum*, *Sida*, or *Urena* from cotton and jute, the 30% transferability with a high number of SSR markers in cotton and jute may also be acceptable for building an initial SSR pool for genomic studies in species where sequence-based SSR development may be problematic due to polyploidy or large genome size (Satya et al. 2016). In the present study, all of the samples were bulked and packed in aluminum foil to maintain the quality of the sample and then stored at − 80 °C for further DNA extraction experiments. In recent studies, the EST-SSR marker was studied in transcriptome sequencing and cross-species transferability in *Arecaceae* species, where among 145 EST-SSRs, 144 were transferable in *Acrocomia intumescens*, 143 were transferable in *Acrocomia totai*, 117 were transferable in the African oil palm (*Elaeis guineensis*) and peach palm (*Bactris gasipaes*), 106 were transferable in the juçara palm, and 105 were transferable in the hat palm (*Sabal causiarum*), indicating that these newly developed EST-SSRs can be used in future population genetic studies (Bazzo et al. 2018). Furthermore, in a study of leguminous and non-leguminous plants, high cross-species transferability of the Mt-miRNA-SSR markers was reported in leguminous species; the transferability of 169 Mt-miRNA-SSR markers in leguminous species reached 77.51% in *Glycine max*, 89.35% in *Vicia sativa*, 88.76% in *Melilotus*, 90.53% in *Lotus corniculatus*, and 89.35 in *Sophora alopecuroides* (Min et al. 2017).

We have found good prospects for transferability in *C. songorica* and other Gramineae species with a high percentage (more than 50%) compared with non-Gramineae species. This finding showed that the transferability of miRNA-SSR markers to Gramineae and non-Gramineae species may facilitate molecular genetic studies, species conservation, ecological studies, and genetic improvement. The 82 transferable primer pairs were analyzed in 23 *Cleistogenes* accessions. All 82 primer pairs yielded unambiguous and high amplification across the 23 accessions, and a total of 385 alleles were polymorphic. The number of alleles per primer varied from 3 to 11, with an average of 4.69 per locus. The highest PIC value was 0.86, and the average PIC value was 0.62. PIC values greater than 0.5 are considered to indicate informative markers, and loci with PIC values greater than 0.7 are suitable for genetic mapping (Bandelj et al. 2004). Of the 110 SSR-containing miRNA genes, mononucleotide repeats occurred at higher frequencies than did di- and trinucleotides. $(A)_{27}$ mononucleotides occurred at the highest frequency (24.5%), and trinucleotides occurred at the lowest (all 0.9%). Our results indicate that the number of repeats is inversely related to repeat length such that as the repeat length increases, the number and relative percentage of SSRs decrease (Ganie and Mondal 2015; Rakoczy-Trojanowska and Bolibok 2004).

The comparative electropherogram analysis of the Cs miRNA-SSR loci revealed similarity to the original locus from which the marker was designed. In previous studies, including those in *Melilotus* (Yan et al. 2017), on the development EST-SSR of markers in Siberian wild rye (*Elymus sibiricus*) (Zhou et al. 2016) and in *Medicago truncatula* (Min et al. 2017), the sequenced alleles of SSR primers were similar to the original locus in cotton (Jena et al. 2012). The cluster analysis of the genetic relationships among the 23 *Cleistogenes* accessions, performed by UPGMA on the 82 polymorphic Cs-miRNA-SSR markers, clustered the 23 accessions into 3 clusters. There was no significant relationship between the clustering pattern of members of cluster 1 and the geographical location. This result may have been due to the small number of markers or accessions from each geographical location. The populations of the germplasm clustered together, and the genetic similarity coefficient ranged from 0.64 to 0.96, indicating close genetic relationships among the 23 *Cleistogenes* accessions (Wang et al. 2007). Similar results have been reported in other plant species (Singh et al. 2014; Verma and Rana 2011; Wang et al. 2013; Zhou et al. 2014).

This study helps elucidate the transferability of miRNA-SSR markers of *Cleistogenes songorica* in different species, which can now be explored for further genetic and breeding studies.

Acknowledgements The authors gratefully acknowledge the financial support from the National Natural Science Foundation of China (31572453), Gansu Provincial Science and Technology Major Projects (19ZD2NA002), Gansu Provincial Intellectual Property Program (19ZSCQ044), and the 111 Project (B12002).

Compliance with ethical standards

Conflict of interest The authors declare they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Ahmad F, Khan A, Awan F, Sadia B, Sadaqat H, Bahadur S (2010) Genetic diversity of chickpea (Cicer arietinum L.) germplasm in Pakistan as revealed by RAPD analysis. Genet Mol Res 9:1414–1420

Bandelj D, Jakše J, Javornik B (2004) Assessment of genetic variability of olive varieties by microsatellite and AFLP markers. Euphytica 136:93–102

Baraket G, Chatti K, Saddoud O, Abdelkarim AB, Mars M, Trifi M, Hannachi AS (2011) Comparative assessment of SSR and AFLP markers for evaluation of genetic diversity and conservation of fig,

Ficus carica L., genetic resources in Tunisia. Plant Mol Biol Report 29:171–184

Bartel B, Bartel DP (2003) MicroRNAs: at the root of plant development? Plant Physiol 132:709–717

Bazzo BR, de Carvalho LM, Carazzolle MF, Pereira GAG, Colombo CA (2018) Development of novel EST-SSR markers in the macaúba palm (*Acrocomia aculeata*) using transcriptome sequencing and cross-species transferability in Arecaceae species. BMC Plant Biol 18:276

Cavagnaro PF, Senalik DA, Yang L, Simon PW, Harkins TT, Kodira CD, Huang S, Weng Y (2010) Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). BMC Genomics 11:569

Chen M, Tan Z, Jiang J, Li M, Chen H, Shen G, Yu R (2009) Similar distribution of simple sequence repeats in diverse completed Human Immunodeficiency Virus Type 1 genomes. FEBS Lett 583:2959–2963

Cloutier S, Ragupathy R, Niu Z, Duguid S (2011) SSR-based linkage map of flax (*Linum usitatissimum* L.) and mapping of QTLs underlying fatty acid composition traits. Mol Breed 28:437–451

Decroocq V, Fave M, Hagen L, Bordenave L, Decroocq S (2003) Development and transferability of apricot and grape EST microsatellite markers across taxa. Theor Appl Genet 106:912–922

Duan Z, Zhang D, Zhang J, Di H, Wu F, Hu X, Meng X, Luo K, Zhang J, Wang Y (2015) Co-transforming bar and CsALDH genes enhanced resistance to herbicide and drought and salt stress in transgenic alfalfa (*Medicago sativa* L.). Front Plant Sci 6:1115

Fahlgren N, Jogdeo S, Kasschau KD, Sullivan CM, Chapman EJ, Laubinger S, Smith LM, Dasenko M, Givan SA, Weigel D (2010) MicroRNA gene evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. Plant Cell 22:1074–1089

Ganie SA, Mondal TK (2015) Genome-wide development of novel miRNA-based microsatellite markers of rice (*Oryza sativa*) for genotyping applications. Mol Breed 35:51

González-Martínez S, Robledo-Arnuncio J, Collada C, Díaz A, Williams C, Alía R, Cervera M (2004) Cross-amplification and sequence variation of microsatellite loci in Eurasian hard pines. Theor Appl Genet 109:103–111

Haasl RJ, Payseur BA (2012) Microsatellites as targets of natural selection. Mol Biol Evol 30:285–298

Jena SN, Srivastava A, Rai KM, Ranjan A, Singh SK, Nisar T, Srivastava M, Bag SK, Mantri S, Asif MH (2012) Development and characterization of genomic and expressed SSRs for levant cotton (*Gossypium herbaceum* L.). Theor Appl Genet 124:565–576

Jiang GL (2015) Molecular marker-assisted breeding: a plant breeder's review. Advances in plant breeding strategies: breeding, biotechnology and molecular tools. Springer, pp 431–472

Jin Q, Waters D, Cordeiro GM, Henry RJ, Reinke RF (2003) A single nucleotide polymorphism (SNP) marker linked to the fragrance gene in rice (*Oryza sativa* L.). Plant Sci 165:359–364

Kessuwan K, Kubota S, Liu Q, Sano M, Okamoto N, Sakamoto T, Yamashita H, Nakamura Y, Ozaki A (2016) Detection of growth-related quantitative trait loci and high-resolution genetic linkage maps using simple sequence repeat markers in the kelp grouper (*Epinephelus bruneus*). Mar Biotechnol 18:57–84

Khraiwesh B, Zhu J-K, Zhu J (2012) Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms 1819:137–148

Kijas J, Fowler J, Thomas M (1995) An evaluation of sequence tagged microsatellite site markers for genetic analysis within Citrus and related species. Genome 38:349–355

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 11:2453–2465

Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 21:991–1007

Lin SL, Wu DT, Ying S-Y (2008) Recent application of intronic microRNA agents in cosmetics. Current Perspectives in microRNAs (miRNA). Springer, pp 51–72

Liu WX, Jia X, Liu ZP, Zhang Z, Wang YR, Liu ZP, Xie WG (2015) Development and characterization of transcription factor gene-derived microsatellite (TFGM) markers in *Medicago truncatula* and their transferability in leguminous and non-leguminous species. Molecules 20:8759–8771

Liu WX, Liu WH, Jun W, Gao AN, Li LH (2010) Analysis of genetic diversity in natural populations of *Psathyrostachys huashanica* Keng using microsatellite (SSR) markers. Agric Sci China 9:463–471

Liu Q, Wang H, Zhu L, Hu H, Sun Y (2013) Genome-wide identification and analysis of miRNA-related single nucleotide polymorphisms (SNPs) in rice. Rice 6:10

Liu M, Xu Y, He J, Zhang S, Wang Y, Lu P (2016) Genetic diversity and population structure of broomcorn millet (*Panicum miliaceum* L.) cultivars and landraces in China based on microsatellite markers. Int J Mol Sci 17:370

Min XY, Zhang ZS, Liu Y, Wei X, Liu ZP, Wang YR, Liu WX (2017) Genome-wide development of microRNA-based SSR markers in *Medicago truncatula* with their transferability analysis and utilization in related legume species. Int J Mol Sci 18:2440

Mondal TK, Ganie SA, Debnath AB (2015) Identification of Novel and Conserved miRNAs from Extreme Halophyte,*Oryza coarctata*, a Wild Relative of Rice. PLoS One 10:e0140675

Muvunyi BP, Yan Q, Wu F, Min XY, Yan ZZ, Kanzana G, Wang YR, Zhang JY (2018) Mining late embryogenesis abundant (LEA) family genes in *Cleistogenes songorica*, a xerophyte perennial desert plant. Int J Mol Sci 19:3430

Neilson JR, Sharp PA (2008) Small RNA regulators of gene expression. Cell 134:899–902

Ni J, Colowit PM, Mackill DJ (2002) Evaluation of genetic diversity in rice subspecies using microsatellite markers. Crop Sci 42:601–607

Noli E, Teriaca M, Sanguineti M, Conti S (2008) Utilization of SSR and AFLP markers for the assessment of distinctness in durum wheat. Mol Breed 22:301–313

Nozawa M, Miura S, Nei M (2012) Origins and evolution of microRNA genes in plant species. Genome Biol Evol 4:230–239

Parveen S, Shahzad A, Yadav V (2016) Molecular markers and their application in plant biotechnology. Biotechnological strategies for the conservation of medicinal and ornamental climbers. Springer, pp 389–413

Rakoczy-Trojanowska M, Bolibok H (2004) Characteristics and a comparison of three classes of microsatellite-based markers and their application in plants. Cell Mol Biol Lett 9:221–238

Saha MC, Cooper JD, Mian MR, Chekhovskiy K, May GD (2006) Tall fescue genomic SSR markers: development and transferability across multiple grass species. Theor Appl Genet 113:1449–1458

Saini HK, Enright AJ, Griffiths-Jones S (2008) Annotation of mammalian primary microRNAs. BMC Genomics 9:564

Satya P, Paswan PK, Ghosh S, Majumdar S, Ali N (2016) Confamiliar transferability of simple sequence repeat (SSR) markers from cotton (*Gossypium hirsutum* L.) and jute (*Corchorus olitorius* L.) to twenty two *Malvaceous* species. 3. Biotech 6:65

Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. Chromosoma 109:365–371

Shan Z, Wu H, Li C, Chen H, Wu Q (2011) Improved SDS method for general plant genomic DNA extraction. Guangdong Agric Sci 8:113–115

Singh R, Narzary D, Bhardwaj J, Singh AK, Kumar S, Kumar A (2014) Molecular diversity and SSR transferability studies in Vetiver grass (*Vetiveria zizanioides* L. Nash). Ind Crop Prod 53:187–198

Singh I, Smita S, Mishra DC, Kumar S, Singh BK, Rai A (2017) Abiotic stress responsive miRNA-target network and related markers (SNP, SSR) in *Brassica juncea*. Front Plant Sci 8:1943

Smulders M, De Klerk G (2011) Epigenetics in plant tissue culture. Plant Growth Regul 63:137–146

Sujatha M (2013) Genetic diversity, molecular markers and marker assisted breeding in Jatropha. Challenges for a New Energy Crop. Springer, Jatropha, pp 395–422

Varshney RK, Shi C, Thudi M, Mariac C, Wallace J, Qi P, Zhang H, Zhao Y, Wang X, Rathore A (2017) Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. Nat Biotechnol 35:969

Verma S, Rana T (2011) Genetic diversity within and among the wild populations of *Murraya koenigii* (L.) Spreng., as revealed by ISSR analysis. Biochem Syst Ecol 39:139–144

Wang Z, Yan H, Fu X, Li X, Gao H (2013) Development of simple sequence repeat markers and diversity analysis in alfalfa (*Medicago sativa* L.). Mol Biol Rep 40:3291–3298

Wang G, Yuan Y, Li J (2007) SSR analysis of genetic diversity and phylogenetic relationships among different populations of Hyriopsis cumingii from the five lakes of China. J Fish China 2

Wassom JJ, Mikkelineni V, Bohn MO, Rocheford TR (2008) QTL for fatty acid composition of maize kernel oil in Illinois High Oil× B73 backcross-derived lines. Crop Sci 48:69–78

Wiesner I, Wiesnerova D, Tejklova E (2001) Effect of anchor and core sequence in microsatellite primers on flax fingerprinting patterns. J Agric Sci 137:37–44

Wu F, Zhang D, Ma J, Luo K, Di H, Liu ZP, Zhang JY, Wang YR (2016) Analysis of genetic diversity and population structure in accessions of the genus *Melilotus*. Ind Crop Prod 85:84–92

Wu F, Zhang D, Muvunyi BP, Yan Q, Zhang YF, Yan ZZ, Cao M, Wang YR, Zhang JY (2018) Analysis of microRNA reveals cleistogamous and chasmogamous floret divergence in dimorphic plant. Sci Rep 8: 1–13

Xu P, Wu F, Ma TT, Yan Q, Zong XF, Li J, Zhao YF, Kanzana G, Zhang JY (2020) Analysis of Six Transcription Factor Families Explores Transcript Divergence of Cleistogamous and Chasmogamous Flowers in *Cleistogenes songorica*. DNA Cell Biol

Yan Z, Wu F, Luo K, Zhao Y, Yan Q, Zhang Y, Wang YR, Zhang JY (2017) Cross-species transferability of EST-SSR markers developed from the transcriptome of *Melilotus* and their application to population genetics research. Sci Rep 7:17959

Yan Q, Wu F, Yan ZZ, Li J, Ma TT, Zhang YF, Zhao YF, Wang YR, Zhang JY (2019) Differential co-expression networks of long non-coding RNAs and mRNAs in *Cleistogenes songorica* under water stress and during recovery. BMC Plant Biol 19:1–19

Yang J, Zhu G, Gao G (2001) Effects of grazing systems on the reproductive feature of key plant population in Stipa breviflora steppe. J Arid Land Res Environ

Zhang J, Duan Z, Jahufer Z, An S, Wang Y (2014) Stress-inducible expression of a *Cleistogenes songorica* ALDH gene enhanced drought tolerance in transgenic *Arabislopsis thaliana*. Plant Omics 7:438

Zhang JY, Duan Z, Zhang D, Zhang J, Di H, Wu F, Wang YR (2016) Co-transforming bar and CsLEA enhanced tolerance to drought and salt stress in transgenic alfalfa (*Medicago sativa* L.). Biochem Biophys Res Commun 472:75–82

Zhang JY, John UP, Wang YR, Li X, Gunawardana D, Polotnianka RM, Spangenberg GC, Nan Z (2011) Targeted mining of drought stress-responsive genes from EST resources in *Cleistogenes songorica*. J Plant Physiol 168:1844–1851

Zhang Q, Ma B, Li H, Chang Y, Han Y, Li J, Wei G, Zhao S, Khan MA, Zhou Y (2012) Identification, characterization, and utilization of genome-wide simple sequence repeats to identify a QTL for acidity in apple. BMC Genomics 13:537

Zhang ZS, Min XY, Wang Z, Wang YR, Liu ZP, Liu WX (2017) Genome-wide development and utilization of novel intron-length polymorphic (ILP) markers in *Medicago sativa*. Mol Breed 37:87

Zhou Q, Chen T, Wang Y, Liu Z (2014) The development of 204 novel EST-SSRs and their use for genetic diversity analyses in cultivated alfalfa. Biochem Syst Ecol 57:227–230

Zhou Q, Luo D, Ma L, Xie WG, Wang YR, Wang Y, Liu ZP (2016) Development and cross-species transferability of EST-SSR markers in Siberian wildrye (*Elymus sibiricus* L.) using Illumina sequencing. Sci Rep 6:20549