

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

Gisele Kanzana¹.Yufei Zhang¹.Tiantian Ma¹.Wenxian Liu¹. Fan Wu¹.Qi Yan¹.Xueyang Min¹.Zhuanzhuan Yan¹.Blaise Pascal Muvunyi¹.Jie Li¹.Zhengshe Zhang¹.Yufeng Zhao¹.Jiyu Zhang^{1,*}

kanzana16@lzu.edu.cn; zhangyf16@lzu.edu.cn ; matt17@lzu.edu.cn; liuwx@lzu.edu.cn; wuf15@lzu.edu.cn; yanq16@lzu.edu.cn; minxy15@lzu.edu.cn; yanzhzh16@lzu.edu.cn; muvunyi14@lzu.edu.cn; lij17@lzu.edu.cn; zhangzhsh14@lzu.edu.cn; zhaoyf17@lzu.edu.cn; zhangjy@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

***Correspondence: zhangjy@lzu.edu.cn.**

Tel. /Fax: +86-931-891-4051

Abstract

SSR markers are commonly used for many genetic applications, such as map construction, fingerprinting and genetic diversity analysis due to their high reproducibility, levels of polymorphism and abundance. As endogenous, small RNAs, miRNAs have essential roles in plant development and gene expression under diverse stress conditions, including various biotic and abiotic stress conditions. In the present study, we predicted 110 pre-miRNAs sequences from 287 precursor miRNAs and used them as queries for SSR marker development. 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 (H_e) 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. In this study, 1422 miRNA target genes were predicted and analyzed using the GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) databases. The results showed that this miRNA-based microsatellite marker system can be very useful for genetic diversity and marker-assisted breeding studies.

Keywords *Cleistogenes songorica*; microRNA; simple sequence repeat; genetic diversity; transferability; target genes.

Cleistogenes songorica (*C. songorica*), which belongs to the family Gramineae, 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. To study the drought tolerance mechanism of *C. songorica*, leaf and root expression sequence tag (EST) resources have been used to investigate drought stress-responsive genes (Zhang et al. 2011). Some of these genes have been transferred into *Arabidopsis thaliana* and alfalfa to confirm and enhance stress tolerance in these plants (Duan et al. 2015; Zhang et al. 2016).

Small RNAs or microRNAs (miRNAs) are a class of noncoding RNAs 18-24 nucleotides in length. Endogenous small RNAs have essential roles in plant development, phase transition (Khraiweh et al. 2012) and gene expression under diverse stress conditions, including 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 not translated into proteins but are often 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, which regulate various developmental and biological processes (Fahlgren et al. 2010). Molecular markers or DNA markers have become efficient tools for identify 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 in two types: non-polymerase chain reaction-based markers that identify restriction fragment length polymorphisms (RFLPs) and polymerase chain reaction-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 simple sequence repeat (SSR) markers (Baraket et al. 2011; Zhang et al. 2012; Min et al. 2017). 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, multi-allelic nature, distribution, reproducibility, distribution throughout the genome, ease of use and generally codominant nature of SSR markers make them highly suitable for genetic diversity studies (Wiesner et al. 2001; Wassom et al. 2008; Cloutier et al. 2011; Smulders and De Klerk 2011; Kessuwan et al. 2016).

SSR markers, 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 (Li et al. 2004; Li et al. 2002; Haasl and Payseur 2012). SSRs exhibit high polymorphism, abundance, and genetic diversity and tend to be codominant (Ni et al. 2002; Noli et al. 2008; Liu et al. 2010; Liu et al. 2016; Parveen et al. 2016). SSRs are widely used for genetic diversity analysis, germplasm identification, comparative genetics analysis, phylogenetic analysis, QTL analysis, linkage mapping and marker-assisted selection (Rakoczy-Trojanowska and Bolibok 2004; Saha et al. 2006; Cavagnaro et al. 2010; Liu et al. 2010; Mondal et al. 2015). 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. Despite the advances in miRNA and SSR development, there is a need for the development of miRNA-associated markers, i.e., miRNA-SSRs, 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 present study describes the first genome-wide development of miRNA-SSR markers based on *C. songorica* and their transferability to other species.

Methods and materials

Plant material and DNA extraction

Plant materials contained gramineae and non- gramineae species. Gramineae species are *Cleistogenes keng*, wheat (*Triticum aestivum*), ryegrass (*Lolium perenne*), rice (*Oryza sativa*) and maize (*Zea mays*), while non- gramineae are *Arabidopsis thaliana*, *Medicago truncatula*,

alfalfa (*Medicago sativa*), yellow sweet clover (*Melilotus officinalis*), common vetch (*Vicia sativa*) and soybean (*Glycine max*). All these 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 23 *Cleistogenes* accessions one is cultivated type and the other 22 are wild types. The 22 *Cleistogenes* accessions were obtained from China (in Gansu, Inner Mongolia and Shandong provinces) and only one accession is from Mongolia (**Table S1 in File S1**).

The young leaves of the plants were collected separately and were bulked as a single sample and used for genomic DNA isolation. *Cleistogenes songorica* samples were frozen in nitrogen liquid 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 samples were diluted with ddH₂O to 50 ng/μl and stored at -20 °C. 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 by using a Perl 5 script (*MISA*, MicroSATellite identification tool). Among the 287 pre-miRNAs, 110 pre-miRNAs sequences were selected and used as queries for designing primers flanking repeats. Flanking primers to the SSRs were designed using Primer3 software and Perl 5 interface modules. To design the SSRs in this study, pre-miRNA sequences with 100% matching (E value = 10⁻¹⁰) to the *C. songorica* genome sequence (local database in a webserver, www.biocloud.net) along with the 500 bp on each end (5' and 3') were used (**Figure S1 in File S2**). The minimum length criteria were 10 and six repeat units for mononucleotide and dinucleotide repeats, respectively, and five repeat units for trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide repeats (**Table S2 in File S1**). The miRNA-SSR primers were designed using BatchPrimer3, and 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 on 19 chromosomes using TBtools (<https://github.com/CJ-Chen/TBtools>) where 90 and 20 markers were successfully mapped and scaffolds respectively.

Prediction of miRNA target genes and GO analysis

In the present study, we extracted pre-miRNAs from local genome database of *C. songorica* and after removing all the redundant, the corresponding pre-miRNA sequences of non-redundant pre-miRNAs were used as queries for a BlastN search using miRbase against the rice genome in grameneae (Jaiswal et al. 2002) to obtain the small mature miRNAs corresponding to the pre-miRNAs that were used as queries by using psRNATarget Server to identify miRNA target genes. Target genes were used for bioinformatics analysis, GO and KEGG pathway were analyzed using BioCloud software (<https://www.biocloud.net>).

PCR amplification

PCR was carried out in a 10µl reaction mixtures containing 4.95 µl 2 × reaction Mix (dNTPs at 500 µM each, 20 mM Tris-HCL, 100 mM MgCl₂, 100 mM KCl₂, 3 mM MgCl₂), 2.0 µl double distilled water, 1.0 µl Template DNA, 1.0 µl forward primers and 1.0 µl reverse primers, 0.05 µl Golden DNA polymerase. Amplifications were performed with pre-denaturation of 3min at 94 °C, 30sec at 94 °C, 30sec at 60 °C, 30sec at 72 °C, 30sec at 94 °C, 30sec at 56 °C and final elongation step of 7min at 72 °C. PCR products were visualized on 2% agarose gels using Gel Red staining or 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 for absent (0) or present (1) of the corresponding bands among different *Cleistogenes* accessions. Individual bands that could be clearly scored were used for genetic diversity analysis. A dendrogram was constructed from a genetic identity matrix by using NTSYS-pc V.2.1 software and the UPGMA method. The significance of each node was evaluated by bootstrapping data over a locus for 1,000 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 by using Cs-miRNA-SSR markers and STRUCTURE software. 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 truly and exactitude of PCR amplification product and according to the presence of single bands and high amplification efficiency, we select some PCR amplification products and sequenced by the Shanghai Sangon Biotech Company. We used chromas software (<http://technelysium.com.au/wp/chromas/>) in order to obtain good and clear sequencing chromatograms nucleotide peaks results of PCR product sequences.

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 MISA 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 genes revealed that 27 pre-miRNA genes contained more than one SSR. 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 hexanucleotide repeats were found in any of the *C. songorica* pre-miRNA flanking sequences. In the 99 pre-miRNAs containing SSRs, mononucleotides occurred at the highest frequency (58%), being present in 58 miRNA genes, followed by dinucleotides (38%), present in 38 genes, and trinucleotides (3%), present in only 3 miRNA genes (**Figure 1A and B**).

Table 1 A summary of SSR search results.

Search items	Numbers
Total number of pre-miRNA examined	287
Total number of identified SSRs	125
Number of SSR containing sequences	99
Number of pre-miRNA containing more than 1 SSR	27
Number of SSRs present in compound formation	15
Repeat type	
Mononucleotide	61
Dinucleotide	46
Trinucleotide	3
Total length of sequences searched (kb)	110
Frequency of SSRs	One per 0.88kb

In the 110 SSR-containing miRNA genes, mononucleotide repeats were the most common type of repeat. Among the mononucleotides, (A)₂₇ was 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% (**Figure 1C**). Our study showed that mononucleotides were the most common type of repeat motif, whereas trinucleotides were the least common.

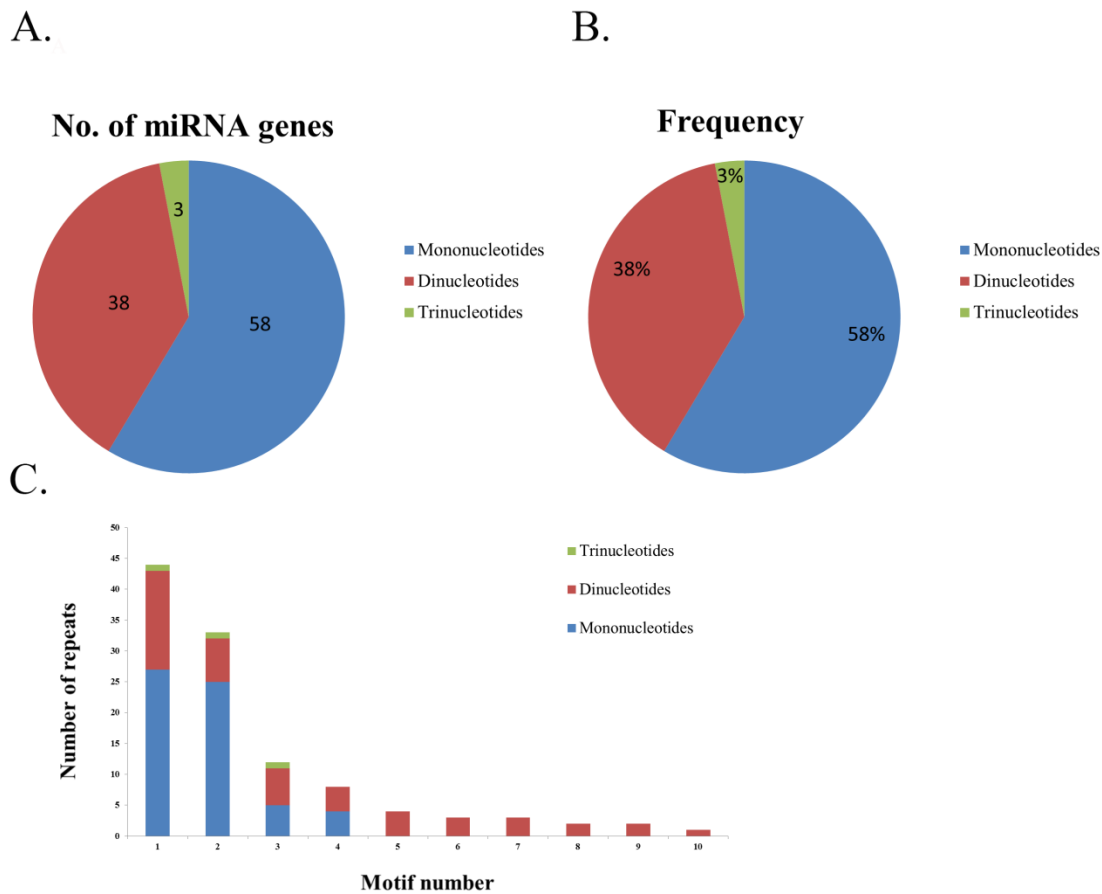


Figure 1 Number of miRNA genes with SSR motifs and abundance of miRNA-SSR motifs of *Cleistogenes songorica*. **A.** Number of miRNA genes possessing SSR motifs; **B.** Frequency of different repeat motifs in miRNA genes; **C.** Number of repeats having the most abundant miRNA-SSR motifs.

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

All of the 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 smaller or larger than predicated. PCR products were then selected according to the presence of single bands and high amplification efficiency. For verification, PCR amplicons of several Cs-miRNA-SSR primers were amplified and sequenced. Among the sequenced alleles from the different miRNA-SSR primers, only some were homologous to the original locus from which the marker was developed. Primer pair Cs-miRNA-SSR-27, which contained trinucleotide repeats (TTG)₅, and Cs-miRNA-SSR-82, which had dinucleotide repeats (TC)₁₁, were matched to the locus used for marker development. Primer Cs-miRNA-SSR-108 and Cs-miRNA-SSR-67, which contained (A)₉ and (T)₄, respectively, were not homologous to the marker locus (**Figure S2 in File S2**). The highest amplification percentage (98%) was observed in *Cleistogenes songorica*, and the lowest (41.2%) was observed in *Glycine max*. The average amplification 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 *Triticuma estivum* and *Zea mays*, two gramineae species. The 85 primer pairs were successfully amplified C. songorica were amplified in 11 species. C. songorica showed higher polymorphism than did the other species. A total of 466 alleles were detected by 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 S3a in File S1**).

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

No.	Genus/species	Transferability (%)
1	<i>Cleistogenes songorica</i>	98
2	<i>Oryza sativa</i>	64.7
3	<i>Lolium perenne</i>	60

4	<i>Triticum aestivum</i>	51.76
5	<i>Zea mays</i>	49.41
6	<i>Arabidopsis thaliana</i>	55.3
7	<i>Vicia sativa</i>	52.94
8	<i>Medicago sativa</i>	48.23
9	<i>Medicago truncatula</i>	44.7
10	<i>Melilotus officinalis</i>	47
11	<i>Glycine max</i>	41.2
Mean		55.93

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

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 S3b in File S1**), which, being greater than 0.5, indicates the high level of polymorphism of these markers and suggests their potential for genetic diversity and genetic mapping analyses. To evaluate the application potential of the Cs-miRNA-SSR markers to the study of genetic diversity in *Cleistogenes songorica*, the 82 transferable primer pairs were analyzed in 23 *Cleistogenes* accessions. Due to the genetic closeness of *Cleistogenes songorica* with other *Cleistogenes* accessions, genetic marker array between *Cleistogenes songorica* and other *Cleistogenes* accessions could be evaluated and further use for its genetic diversity analysis. The banding patterns of 23 *Cleistogenes* accessions obtained with Cs-miRNA-SSR-31 and Cs-miRNA-SSR-82 markers are portrayed in (**Figure S3 in File S2**).

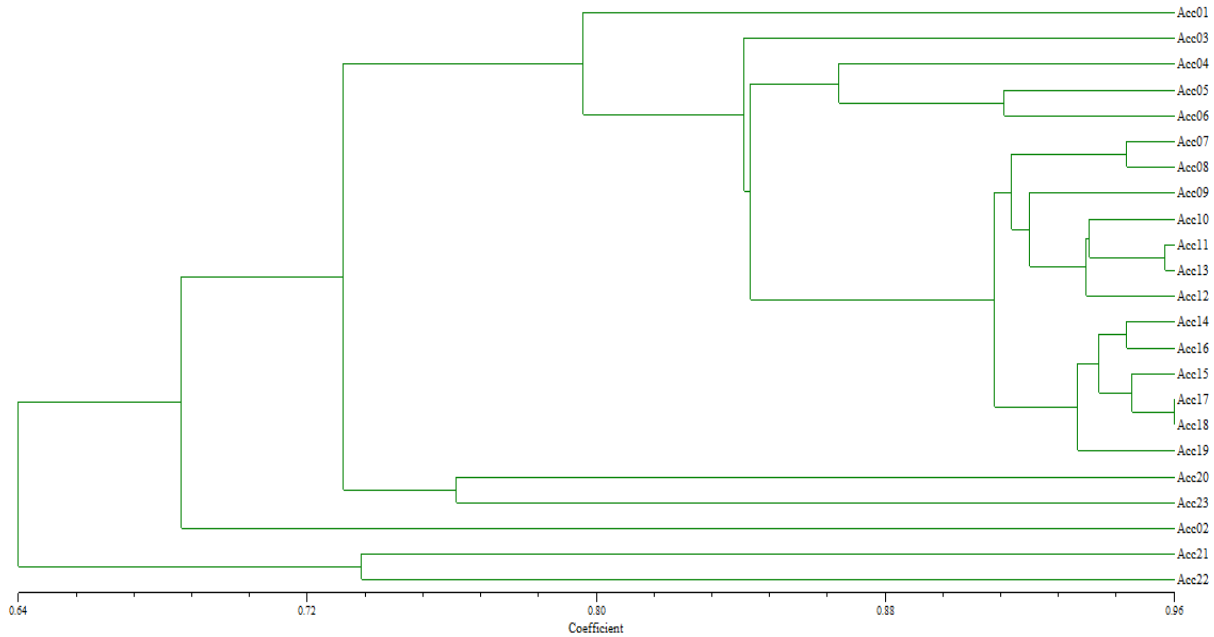


Figure 2 A dendrogram of unweighted pair group method arithmetic mean analysis (UPGMA)

Analysis of the 23 *Cleistogenes* accessions grouped 17 of them together. In the cluster analysis, the twenty-three *Cleistogenes* accessions were divided into three clusters: cluster I, containing 20 accessions (Acc 01, Acc 03, Acc 04, Acc 05, Acc 06, Acc 07, Acc 08, Acc 09, Acc 10, Acc 11, Acc 12, Acc 13, Acc 14, Acc 15, Acc 16, Acc 17, Acc 18, Acc 19, Acc 20 and Acc 23); cluster II, containing one accession (Acc 02); and cluster III, containing two accessions (Acc 21 and Acc 22) (**Figure 2**). In the analysis of population structure, 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 (**Figure 3**).

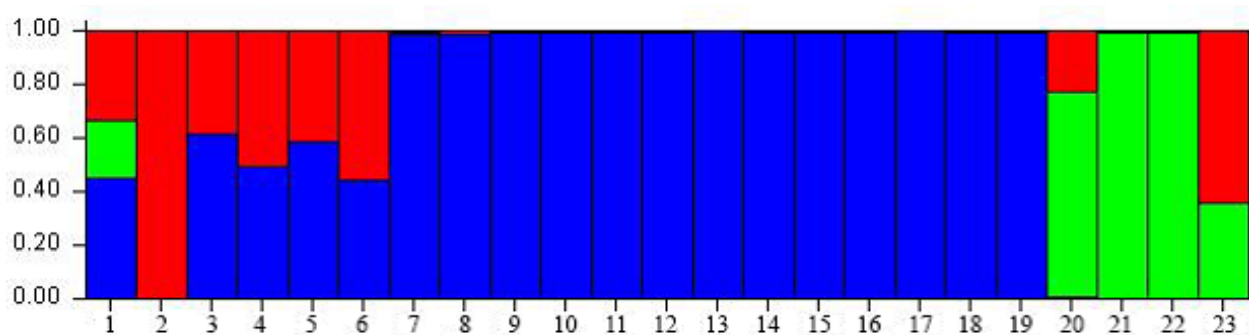


Figure 3 Histogram structure of the miRNA-SSR molecular markers data set for the model with $K = 3$ (showing the highest ΔK)

Among the three groups, group I contained eighteen accessions, group II contained one accession and group III contained 4 accessions. The PCA showed that 5 *Cleistogenes* accessions, including Acc 02 and Acc 20, which were *C. songorica*, did not cluster into any group, whereas the remaining 18 clustered together into one group. Most of these 18 were *C. songorica*, and one was *C. caespitosa* (**Figure 4**). Although *C. hackelii* and *C. hancei* clustered in the same group, their genetic distance was large.

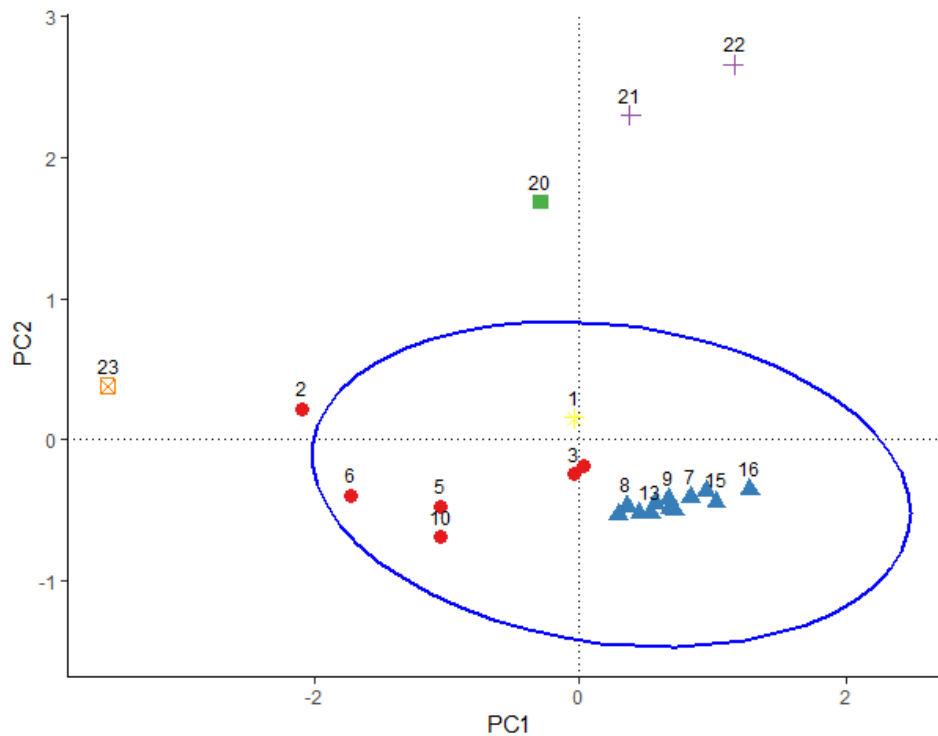


Figure 4 Grouping of geographical origin and genetic distributions of 23 *Cleistogenes* accessions using 82 pairs of miRNA-SSR markers. The principal components analysis (PCA) plot, each accession is represented by a vertical bar and the length of each colored segment in each vertical bar represents the proportion contributed by ancestral populations.

Distribution of miRNA-SSRs on *Cleistogenes songorica* chromosomes

Among the 110 pre-miRNAs of the *C. songorica* genome, all 110 Cs-miRNA-SSRs were discovered. Although the 110 Cs-miRNA-SSRs markers were expected to physically map to 20 chromosomes, only 90 Cs-miRNA-SSRs were mapped, to 19 chromosomes. The remaining 20 Cs-miRNA-SSRs 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%) (**Figure S4 in File S2**).

Gene ontology and KEGG pathway analysis

To evaluate 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 process (19 GO terms), cellular component (12 GO terms) and molecular function (11 GO terms) (**Figure 5**). In the biological process category, the two most overrepresented GO terms were metabolic process (648 genes) and cellular process (634 genes), followed by single-organism process (497 genes). In the cellular component category, the most overrepresented GO terms were cell part (786 genes) and cell (784 genes), followed by membrane (419 genes). In the molecular function category, the most overrepresented terms were catalytic activity (614 genes) and binding (600 genes).

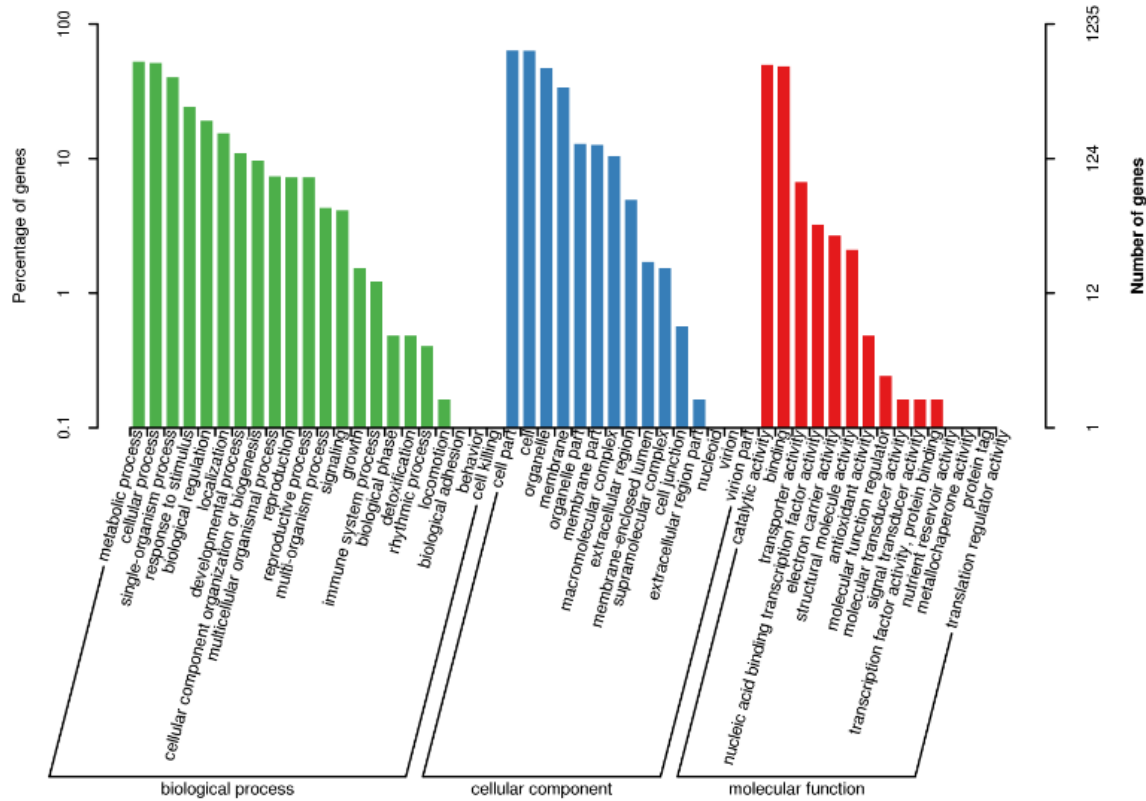


Figure 5 The relative frequencies of miRNA target genes assigned to the GO functional in 3 categories; biological process, cellular component and molecular function.

KEGG pathway analysis showed that 1422 miRNA target genes were annotated in the 20 most represented pathways. 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 enrichment level, and the color corresponds to the number of genes enriched for the given pathway. A small q -value (red) indicates high significance of the pathway. Many genes were assigned to carbon metabolism, with others assigned to pyruvate metabolism, phagosome, glycolysis/gluconeogenesis and other pathways (**Figure S5 in File S2**).

Discussion

In the present study, a total of 287 pre-miRNAs of *C. songorica* were predicted from the *Cleistogenes songorica* local genome. We discovered 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. In several studies, the flanking regions of pre-miRNA sequences, ranging from 300 bp 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 pre-miRNA sequences. Similarly, to identify the single nucleotide polymorphisms in rice miRNA genes, (Liu et al. 2013) used the 1-kb flanking region from both ends of the pre-miRNA genes. Thus, we believe that the 500-bp regions on either side of the pre-miRNA sequences in this study form part of the miRNA genes that were used to design the SSRs in this study. Our study found no tetra-, penta- or hexanucleotides in the *C. songorica* pre-miRNA sequences. Among the repeat types, mononucleotide miRNA-SSRs occurred at the highest frequency, whereas trinucleotide miRNA-SSRs occurred at the lowest frequency. Among the 110 miRNA-SSR primer pairs, 85 were successfully amplified in 11 gramineae and non-gramineae species, and *C. songorica* showed the highest levels of polymorphism. The 82 transferable primer pairs were analyzed in 23 *Cleistogenes* accessions. All 82 primer pairs yielded unambiguous, high amplification and 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 tri-nucleotides. (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 repeat length increases, number and relative percentage of SSRs decreases (Rakoczy-Trojanowska and Bolibok 2004; Ganie and Mondal 2015). Among the three groups, group I contained eighteen accessions, group II contained one accession and group III contained 4 accessions. The PCA showed that 5 *Cleistogenes* accessions, including Acc 02 and Acc 20, which were *C. songorica*, did not cluster into any group, whereas the remaining 18 clustered together into one group. Most of these 18 were *C. songorica*, and one

was *C. caespitosa* (**Figure 4**). Although *C. hackelii* and *C. hancei* clustered in the same group, their genetic distance was large.

The comparative electropherogram analysis of the four miRNA-SSR loci revealed that some of the sequenced alleles from the different miRNA-SSR primers were similar to the original locus from which the marker was developed, whereas others were not. Primer pair Cs-miRNA-SSR-27, with trinucleotide repeats (TTG)₅, and Cs-miRNA-SSR-82, with dinucleotide repeats (TC)₁₁, were matched to the locus used for marker development, but primers Cs-miRNA-SSR-108 and Cs-miRNA-SSR-67, which contained (A)₉ and (T)₄, respectively, were not. In previous studies, including those in *Melilotus* (Yan et al. 2017) and *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 of the 82 polymorphic Cs-miRNA-SSR markers, clustered the 23 accessions into 3 clusters. There was no significant relationship between the pattern of clustering of members of cluster 1 and 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 (Verma and Rana 2011; Wang et al. 2013; Singh et al. 2014; Zhou et al. 2014).

It can be concluded that miRNA-SSR markers of *C. songorica* have great potential for studies of genetic diversity and for transferability to other Gramineae and non-Gramineae species. We hope that 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.

Acknowledgements The author gratefully acknowledge the financial support from National Natural Science Foundation of China (31572453), Special Fund for Agro-scientific, Research in

Public Interest (20120304205), Program for Changjiang Scholars and Innovative Research Team in University (IRT_17R50), Fundamental Research Funds for the Central Universities (Izujbky-2016-10), and the 111 Project (B12002).

Literature Cited

- Ahmad, F., A. Khan, F. Awan, B. Sadia, H. Sadaqat *et al.*, 2010 Genetic diversity of chickpea (*Cicer arietinum* L.) germplasm in Pakistan as revealed by RAPD analysis. *Genetics and Molecular Research* 9 (3):1414-1420.
- Bandelj, D., J. Jakše, and B. Javornik, 2004 Assessment of genetic variability of olive varieties by microsatellite and AFLP markers. *Euphytica* 136 (1):93-102.
- Baraket, G., K. Chatti, O. Saddoud, A.B. Abdelkarim, M. Mars *et al.*, 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 Molecular Biology Reporter* 29 (1):171-184.
- Bartel, B., and D.P. Bartel, 2003 MicroRNAs: at the root of plant development? *Plant physiology* 132 (2):709-717.
- Cavagnaro, P.F., D.A. Senalik, L. Yang, P.W. Simon, T.T. Harkins *et al.*, 2010 Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC genomics* 11 (1):569.
- Chen, M., Z. Tan, J. Jiang, M. Li, H. Chen *et al.*, 2009 Similar distribution of simple sequence repeats in diverse completed Human Immunodeficiency Virus Type 1 genomes. *FEBS letters* 583 (17):2959-2963.
- Cloutier, S., R. Ragupathy, Z. Niu, and S. Duguid, 2011 SSR-based linkage map of flax (*Linum usitatissimum* L.) and mapping of QTLs underlying fatty acid composition traits. *Molecular Breeding* 28 (4):437-451.
- Duan, Z., D. Zhang, J. Zhang, H. Di, F. Wu *et al.*, 2015 Co-transforming bar and CsALDH genes enhanced resistance to herbicide and drought and salt stress in transgenic alfalfa (*Medicago sativa* L.). *Frontiers in plant science* 6:1115.
- Fahlgren, N., S. Jogdeo, K.D. Kasschau, C.M. Sullivan, E.J. Chapman *et al.*, 2010 MicroRNA gene evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. *The Plant Cell* 22 (4):1074-1089.
- Ganie, S.A., and T.K. Mondal, 2015 Genome-wide development of novel miRNA-based microsatellite markers of rice (*Oryza sativa*) for genotyping applications. *Molecular Breeding* 35 (1):51.
- Haasl, R.J., and B.A. Payseur, 2012 Microsatellites as targets of natural selection. *Molecular biology and evolution* 30 (2):285-298.
- Jaiswal, P., D. Ware, J. Ni, K. Chang, W. Zhao *et al.*, 2002 Gramene: development and integration of trait and gene ontologies for rice. *International Journal of Genomics* 3 (2):132-136.
- Jena, S.N., A. Srivastava, K.M. Rai, A. Ranjan, S.K. Singh *et al.*, 2012 Development and characterization of genomic and expressed SSRs for levant cotton (*Gossypium herbaceum* L.). *Theoretical and applied genetics* 124 (3):565-576.
- Jiang, G.-L., 2015 Molecular marker-assisted breeding: a plant breeder's review, pp. 431-472 in *Advances in plant breeding strategies: breeding, biotechnology and molecular tools*. Springer.
- Jin, Q., D. Waters, G.M. Cordeiro, R.J. Henry, and R.F. Reinke, 2003 A single nucleotide polymorphism (SNP) marker linked to the fragrance gene in rice (*Oryza sativa* L.). *Plant Science* 165 (2):359-364.

- Kessuwan, K., S. Kubota, Q. Liu, M. Sano, N. Okamoto *et al.*, 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*). *Marine Biotechnology* 18 (1):57-84.
- Khraiweh, B., J.-K. Zhu, and J. Zhu, 2012 Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 1819 (2):137-148.
- Li, Y.-C., A.B. Korol, T. Fahima, and E. Nevo, 2004 Microsatellites within genes: structure, function, and evolution. *Molecular biology and evolution* 21 (6):991-1007.
- Li, Y.C., A.B. Korol, T. Fahima, A. Beiles, and E. Nevo, 2002 Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular ecology* 11 (12):2453-2465.
- Lin, S.-L., D.T. Wu, and S.-Y. Ying, 2008 Recent application of intronic microRNA agents in cosmetics, pp. 51-72 in *Current Perspectives in microRNAs (miRNA)*. Springer.
- Liu, M., Y. Xu, J. He, S. Zhang, Y. Wang *et al.*, 2016 Genetic diversity and population structure of broomcorn millet (*Panicum miliaceum* L.) cultivars and landraces in China based on microsatellite markers. *International journal of molecular sciences* 17 (3):370.
- Liu, Q., H. Wang, L. Zhu, H. Hu, and Y. Sun, 2013 Genome-wide identification and analysis of miRNA-related single nucleotide polymorphisms (SNPs) in rice. *Rice* 6 (1):10.
- Liu, W.-X., W.-H. Liu, W. Jun, A.-N. Gao, and L.-H. Li, 2010 Analysis of genetic diversity in natural populations of *Psathyrostachys huashanica* Keng using microsatellite (SSR) markers. *Agricultural Sciences in China* 9 (4):463-471.
- Liu, W., X. Jia, Z. Liu, Z. Zhang, Y. Wang *et al.*, 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 (5):8759-8771.
- Min, X., Z. Zhang, Y. Liu, X. Wei, Z. Liu *et al.*, 2017 Genome-wide development of microRNA-based SSR markers in *Medicago truncatula* with their transferability analysis and utilization in related legume species. *International journal of molecular sciences* 18 (11):2440.
- Mondal, T.K., S.A. Ganie, and A.B. Debnath, 2015 Identification of Novel and Conserved miRNAs from Extreme Halophyte, *Oryza coarctata*, a Wild Relative of Rice. *Plos One* 10 (10):e0140675.
- Neilson, J.R., and P.A. Sharp, 2008 Small RNA regulators of gene expression. *Cell* 134 (6):899-902.
- Ni, J., P.M. Colowit, and D.J. Mackill, 2002 Evaluation of genetic diversity in rice subspecies using microsatellite markers. *Crop science* 42 (2):601-607.
- Noli, E., M. Teriaca, M. Sanguineti, and S. Conti, 2008 Utilization of SSR and AFLP markers for the assessment of distinctness in durum wheat. *Molecular Breeding* 22 (2):301-313.
- Nozawa, M., S. Miura, and M. Nei, 2012 Origins and evolution of microRNA genes in plant species. *Genome biology and evolution* 4 (3):230-239.
- Parveen, S., A. Shahzad, and V. Yadav, 2016 Molecular markers and their application in plant biotechnology, pp. 389-413 in *Biotechnological strategies for the conservation of medicinal and ornamental climbers*. Springer.
- Rakoczy-Trojanowska, M., and H. Bolibok, 2004 Characteristics and a comparison of three classes of microsatellite-based markers and their application in plants. *Cellular and Molecular Biology Letters*. 9 (2):221-238.
- Saha, M.C., J.D. Cooper, M.R. Mian, K. Chekhovskiy, and G.D. May, 2006 Tall fescue genomic SSR markers: development and transferability across multiple grass species. *Theoretical and applied genetics* 113 (8):1449-1458.
- Saini, H.K., A.J. Enright, and S. Griffiths-Jones, 2008 Annotation of mammalian primary microRNAs. *BMC genomics* 9 (1):564.
- Schlötterer, C., 2000 Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109 (6):365-371.

- Shan, Z., H. Wu, C. Li, H. Chen, and Q. Wu, 2011 Improved SDS method for general plant genomic DNA extraction. *Guangdong Agricultural Sciences* 8:113-115.
- Singh, I., S. Smita, D.C. Mishra, S. Kumar, B.K. Singh *et al.*, 2017 Abiotic stress responsive miRNA-target network and related markers (SNP, SSR) in Brassica juncea. *Frontiers in plant science* 8:1943.
- Singh, R., D. Narzary, J. Bhardwaj, A.K. Singh, S. Kumar *et al.*, 2014 Molecular diversity and SSR transferability studies in Vetiver grass (*Vetiveria zizanioides* L. Nash). *Industrial Crops and Products* 53:187-198.
- Smulders, M., and G. De Klerk, 2011 Epigenetics in plant tissue culture. *Plant growth regulation* 63 (2):137-146.
- Sujatha, M., 2013 Genetic diversity, molecular markers and marker assisted breeding in Jatropha, pp. 395-422 in *Jatropha, Challenges for a New Energy Crop*. Springer.
- Varshney, R.K., C. Shi, M. Thudi, C. Mariac, J. Wallace *et al.*, 2017 Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nature biotechnology* 35 (10):969.
- Verma, S., and T. Rana, 2011 Genetic diversity within and among the wild populations of *Murraya koenigii* (L.) Spreng., as revealed by ISSR analysis. *Biochemical Systematics and Ecology* 39 (2):139-144.
- Wang, G.-l., Y.-m. Yuan, and J.-l. Li, 2007 SSR analysis of genetic diversity and phylogenetic relationships among different populations of *Hyriopsis cumingii* from the five lakes of China. *Journal of fisheries of China* 2.
- Wang, Z., H. Yan, X. Fu, X. Li, and H. Gao, 2013 Development of simple sequence repeat markers and diversity analysis in alfalfa (*Medicago sativa* L.). *Molecular biology reports* 40 (4):3291-3298.
- Wassom, J.J., V. Mikkilineni, M.O. Bohn, and T.R. Rocheford, 2008 QTL for fatty acid composition of maize kernel oil in Illinois High Oil \times B73 backcross-derived lines. *Crop science* 48 (1):69-78.
- Wiesner, I., D. Wiesnerova, and E. Tejklova, 2001 Effect of anchor and core sequence in microsatellite primers on flax fingerprinting patterns. *The Journal of Agricultural Science* 137 (1):37-44.
- Yan, Z., F. Wu, K. Luo, Y. Zhao, Q. Yan *et al.*, 2017 Cross-species transferability of EST-SSR markers developed from the transcriptome of *Melilotus* and their application to population genetics research. *Scientific Reports* 7 (1):17959.
- Yang, J., G. Zhu, and G. Gao, 2001 Effects of grazing systems on the reproductive feature of key plant population in *Stipa breviflora* steppe. *J Arid Land Res Environ*.
- Zhang, J., Z. Duan, Z. Jahufer, S. An, and Y. Wang, 2014 Stress-inducible expression of a *Cleistogenes songorica* ALDH gene enhanced drought tolerance in transgenic *Arabidopsis thaliana*. *Plant Omics* 7 (6):438.
- Zhang, J., Z. Duan, D. Zhang, J. Zhang, H. Di *et al.*, 2016 Co-transforming bar and CsLEA enhanced tolerance to drought and salt stress in transgenic alfalfa (*Medicago sativa* L.). *Biochemical and biophysical research communications* 472 (1):75-82.
- Zhang, J., U.P. John, Y. Wang, X. Li, D. Gunawardana *et al.*, 2011 Targeted mining of drought stress-responsive genes from EST resources in *Cleistogenes songorica*. *Journal of plant physiology* 168 (15):1844-1851.
- Zhang, Q., B. Ma, H. Li, Y. Chang, Y. Han *et al.*, 2012 Identification, characterization, and utilization of genome-wide simple sequence repeats to identify a QTL for acidity in apple. *BMC genomics* 13 (1):537.
- Zhang, Z., X. Min, Z. Wang, Y. Wang, Z. Liu *et al.*, 2017 Genome-wide development and utilization of novel intron-length polymorphic (ILP) markers in *Medicago sativa*. *Molecular Breeding* 37 (7):87.
- Zhou, Q., T. Chen, Y. Wang, and Z. Liu, 2014 The development of 204 novel EST-SSRs and their use for genetic diversity analyses in cultivated alfalfa. *Biochemical Systematics and Ecology* 57:227-230.