


RESEARCH ARTICLE

Open Access



Genome-wide analysis of microsatellite and sex-linked marker identification in *Gleditsia sinensis*

Jianjun Li*  and Chenglin Ye

Abstract

Background: *Gleditsia sinensis* Lam. (Leguminosae), a dioecious perennial arbor, demonstrates important medicinal properties and economic value. These properties can be harnessed depending on the sex of the plant. However, the sex of the plants is difficult to identify accurately through morphological methods before the flowering.

Results: We used bulked segregant analysis to screen sex-specific simple sequence repeat (SSR) markers in *G. sinensis*. Five male and five female plants were pooled to form the male and female bulks, respectively, and subjected to whole-genome sequencing. After high-throughput sequencing, 5,350,359 sequences were obtained, in which 2,065, 210 SSRs were searched. Among them, the number of duplicated SSRs was the highest. The male plants could reach 857,874, which accounted for 60.86% of the total number of male plants. The female plants could reach 1,447,603, which accounted for 56.25% of the total model of the female plants. Among all the nucleotide repeat types, the A/T-rich motif was the most abundant. A total of 309,516 female strain-specific SSRs were selected by clustering. After designing the primers, the male and female gene pools were amplified, and five pairs of primers (i.e., 27, 34, 36, 39, and 41) were found to amplify the differential bands in the male and female gene pools. Using the five pairs of primers, we performed PCR verification on 10 individuals of known sex, which constructed the gene pool. The female plants amplified a single fragment of lengths (i.e., 186, 305, 266, 203, and 260 bp) and no male plant strip, thereby completing the identification of the male and female sexes of the *G. sinensis*.

Conclusions: This study provides accurate sex identification strategies between female and male plants, thus improving the utilization rate of *G. sinensis* resources.

Keywords: SSR, *Gleditsia sinensis* lam., Sex identification

Background

Gleditsia sinensis Lam. is an important native tree species in China [1]. It can be used for environmental protection, landscaping, and as an economic tree species and wood. In addition, the *G. sinensis* has a high medicinal value, including antidiabetic, antihyperglycemic, antioxidant, anti-inflammatory, anticancer, and anticoagulant activities.

Gleditsiae Sinensis Fructus, *Fructus Gleditsiae*, and *Abnormalis Gleditsiae* Spina are recorded in the Chinese Pharmacopoeia as medicinal parts of *G. sinensis*. The flavonoids and steroids of *Gleditsiae* Spina play an important role in tumor treatment [2, 3]. In addition to its anticancer and antiviral effects, saponins can decontaminate and foam and have been extensively used in manufacturing cosmetics and detergents, with high economic value.

With the deepening of research, market demand for *G. sinensis* has increased over the years. *G. sinensis* is a dioecious plant species. The female tree has a strong pod-forming ability and long vegetative period, whereas

* Correspondence: 043081@htu.cn

College of Life Science, Henan Normal University, Green Medicine Biotechnology Henan Engineering Laboratory, Engineering Technology Research Center of Nursing and Utilization of Genuine Chinese Crude Drugs in Henan Province, Xinxiang 453007, China



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

the male plant does not form pods. Therefore, an ideal ratio of male-to-female individuals must be maintained in producing *G. sinensis* to improve economic efficiency. Numerous male trees will be a waste of land and labor given their slow growth rate. However, the *G. sinensis* tree has a long juvenile period and typically takes 6–8 years to blossom and bear fruit. Thus, the sexuality of trees is difficult to identify in the vegetative stage. Although the sexuality based on the differences of their external characteristics, such as flower morphology and physiological metabolites, is easy to identify at the mature stage, the accuracy and reliability are not high. Therefore, developing sex-specific molecular markers is crucial for identifying male trees at the seedling stage before transplanting.

Simple sequence repeat (SSR), also called microsatellite, is a class of repetitive DNA sequences, which are particularly abundant in plant genomes and have important influence on the function and evolution of genomes and chromosomes [4, 5]. It is also a DNA molecular marker technology based on a PCR technology. SSR is a group of nucleotides composed of several nucleotide (typically one to six) repeats, and the sequence on each side of the SSR is generally a relatively conservative single-copy sequence [6]. SSR markers have several advantages, such as abundant quantity, high information, and co-dominance, over other molecular markers. Each locus is determined by the sequence of designed primers. Therefore, a collaboration for developing primers is convenient for different laboratories [7]. With the rapid accumulation of genome sequences in many plant species, numerous SSR markers have been emerging in many plant species, such as rice [8], cotton [9], and oilseed rape [10]. These SSR markers have been extensively used in genetic map construction, gene mapping, fingerprint mapping, and marker-assisted breeding [11]. In a previous study, we have identified SSRs and have analyzed the distribution and frequency of various base repeat sequences in the *G. sinensis* genome [12]. In the present study, we identify the sex-specific SSR markers by combining bulked segregant analysis (BSA) and next-generation sequencing. The identified markers will be useful for early identification of sex and breeding in *G. sinensis*, thereby helping in maintaining an ideal ratio of male-to-female plants in production and in saving land and increasing pod yield.

Results

Number and distribution of SSR loci in *G. sinensis*

In 5,350,359 sequences from Male and female, 2,065,210 (38.6%) sequences contained SSRs, thus suggesting that the SSRs are highly enriched; a total of 461,166 sequences contained more than one SSR, and 582,826 SSRs contained composite forms. Moreover, 3,627,691 scaffolds assembled from the male genome of *G. sinensis*. The total length of the sequence was 1,375,655,435 bp, and the number of sequences containing SSR was 933,173, with a frequency of 25.72%. Furthermore, the total number of identified SSRs was 1,409,677. In terms of distribution, 1 SSR per 976 bp was set. A total of 3,713,022 scaffolds assembled from the female genome, and the total length of the sequence was 1,359,709,142 bp. The number of sequences containing SSR was 1,464,359, with a frequency of 39.43%, and the total number of identified SSRs was 2,573,292. In the female *G. sinensis* genome distribution, 1 SSR loci per 929 bp was set. The specific SSR search results are summarized in Table 1.

Statistics and analysis of the SSR motif types

The different types of SSR motif were counted. The specific results are presented in Table 2. In this table, repeat SSRs from single-nucleotide repeat SSRs to six nucleotides were distributed in the male and female plants. Among them, the number of dinucleotide repeat motifs was the largest, with 857,874 males accounting for 60.86% of the total number of males and 1,447,603 females accounting for 56.25% of the total number of females. The number of mononucleotide and trinucleotide repeat motifs was also abundant, whereas the number of tetranucleotide to hexanucleotide repeat SSRs was small, and that of pentanucleotide repeat motifs was the least. There are 2200 pentanucleotide repeat motifs in female plants, accounting for 0.09% of the total SSR in female plants. The male plants had 1184 pentanucleotide repeat motifs, accounting for 0.08% of the total SSR of the male plants. Comparative analysis showed that genomic sequences of male and female plants are nearly distinct (Fig. 1). We further subcategorized each SSR repeating module in accordance with its sequence composition. The results of mononucleotide, dinucleotide, and trinucleotide repeat motifs are summarized in Table 3. In this table, the dominant repeat motifs of the

Table 1 SSR results statistics

Category	Male plant	Female plant
Total Number of Sequences Examined	3,627,691	3,713,022
Total Size of Examined Sequences (bp)	1,375,655,435	1,359,709,142
Total Number of Identified SSRs	1,409,677	2,573,292
Number of SSR Containing Sequence	933,173	1,464,359
Number of Sequences Containing More Than One SSR	330,488	665,965
Number of SSRs Present In Compound Formation	418,789	1,006,474

Table 2 Percentage of SSR motifs

Type of SSR motif	Male plant		Female plant	
	SSR motif number	Percentage	SSR motif number	Percentage
Mononucleotide Repeat Motifs	377,120	26.75%	359,835	13.98%
Dinucleotide Repeat Motifs	857,874	60.86%	1,447,603	56.25%
Trinucleotide Repeat Motifs	153,104	10.86%	721,598	28.04%
Tetranucleotide Repeat Motifs	18,625	1.32%	38,623	1.50%
Pentanucleotide Repeat Motifs	1184	0.08%	2220	0.09%
Hexanucleotide Repeat Motifs	1770	0.13%	3413	0.13%
Total Repeat Motifs	1,409,677	100.00%	2,573,292	100.00%

mononucleotide repeat motifs in male and female plants are A/T. The dominant repetitive motif of male dinucleotide repeat motifs was AC/GT, female was AG/CT, and trinucleotide repeat motifs were mainly AAG/CTT.

SSR clustering and polymorphism analysis

The clustering results of SSRs are summarized in Table 4. 1,309,516 SSR sequences with a flanking length of more than 20 bp were selected from the sequence that contains SSRs, including 312,951 males and 516,961

females. The sequence of a specific SSR is clustered, and the result of clustering is displayed in Table 5.

SSR polymorphism assessment

In Table 6, a female-specific SSR length polymorphism was mainly concentrated in SSLP = 1, a total of 142,676, which account for 56.46% of the total. The contents of SSLP = 2 and SSLP = 3 were also high, accounting for 21.64 and 10.26%, respectively.

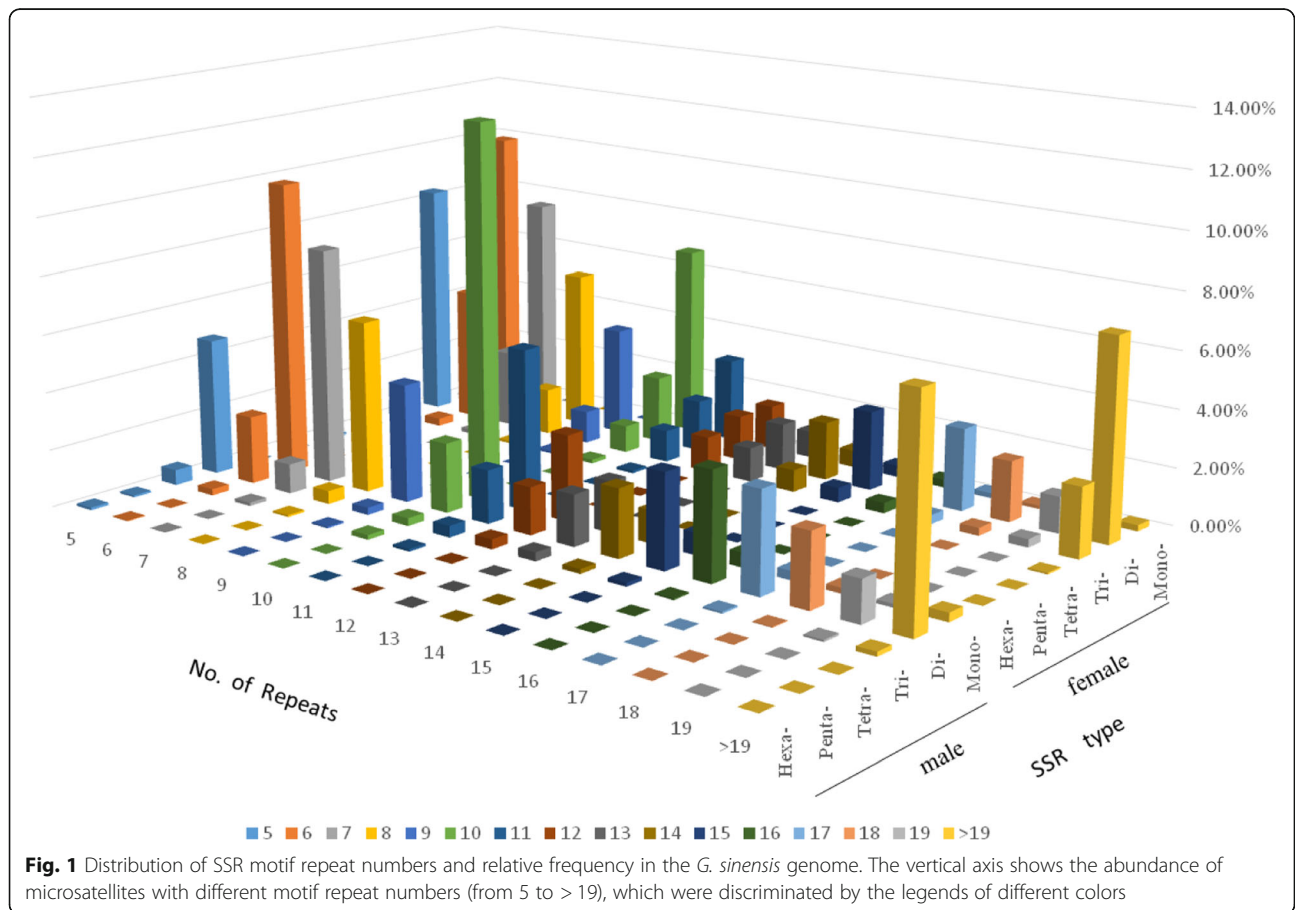


Table 3 Results of Mononucleotide, Dinucleotide, and Trinucleotide Repeat Motifs

Repeat type	Male plant	Female plant
Mononucleotide	377,120	359,835
A/T	370,642	351,803
C/G	6478	8032
Dinucleotide	857,874	1,447,603
AC/GT	318,764	470,223
AG/CT	288,660	688,073
AT/AT	249,210	286,695
CG/CG	1240	2612
Trinucleotide	153,104	721,598
AAC/GTT	26,203	219,657
AAG/CTT	83,164	391,472
AAT/ATT	18,428	34,974
ACC/GGT	1665	6499
ACG/CGT	676	6892
ACT/AGT	606	3551
AGC/CTG	832	6492
AGG/CCT	18,014	36,867
ATC/ATG	3393	14,554
CCG/CGG	123	640

BSA pool PCR amplification results

We selected 50 pairs of random primers from the measured primers and amplified the PCR gene from the male and female gene pools. The amplified products were detected by 1% agarose gel electrophoresis. The results showed that 5 of the 50 pairs of primers (i.e., 27, 34, 36, 39, and 41) amplified different bands in the female gene pool but not in the male gene pool. The detailed information of the primers is presented in Table 7.

PCR verification of known-sex plants

Five pairs of screened primers were used to amplify PCR in 10 plants, which had previously constructed gene pools (Fig. 2). At the same time, five pairs of screened primers were used to amplify PCR in 20 plants from Xinxiang (Fig. 3). This comparison showed 100% accuracy in terms of sex identification. The female plants amplified a single fragment of lengths 186, 305, 266, 150, and 260 bp and no male plant strips.

Table 4 SSR clustering data

Items	Male plant		Female plant	
	Count	Percentage	Count	Percentage
SSR-Containing Sequences	933,173	100.00%	1,464,359	100.00%
SSR-Containing Sequences With Flanking Sequence Length \geq 20 bp	312,951	33.54%	516,969	35.30%

Discussion

With the rapid development of sequencing technology, numerous plant genomes are sequenced and published to the public database, thereby providing an excellent opportunity for searching SSR loci at the genome level and large-scale development of SSR markers. A bulk of previously published report demonstrates SSR markers based on genome sequencing data. For example, an average SSR locus per 1.14 kb is determined in *Arabidopsis thaliana* genome [13], and an average SSR locus per 4 kb is obtained in cabbage genome [14]. In the present study, 2,065,210 SSR loci are found in the genome of *G. sinensis*, with an average of 1 SSR locus per 976 bp for males and 1 SSR locus per 929 bp for females. Considering that only part of the gene sequence of *G. sinensis* is used, the results may not be directly compared with SSR data of other plants mentioned above. However, the results reveal that SSR is extensively distributed in the *G. sinensis* genome.

In the present study, differences in the types of SSR dominant motif are observed in most plants. Most of them are mainly dinucleotide and trinucleotide repeat motifs. Among them, dinucleotide repeat units are dominant in *Brassica napus* [15] watermelon [16] and *Camellia sinensis* [17], whereas trinucleotide repeat units are dominant in the genomes of millet [18] and citrus [19]. At present, the main SSR repeats in the *G. sinensis* genome are dinucleotide repeats, which account for 56.25 and 60.86% of SSRs in male and female plants, correspondingly, followed by mononucleotide and trinucleotide repeat motifs. A previous study on the transcriptome SSR of *G. sinensis* demonstrated an SSR repeat element of the *G. sinensis* transcriptome, this SSR repeat element is mainly based on the dinucleotide repeats [20]. It has provided a new way for identifying the genetic relationship and phylogenetic evolution of the genus *Gleditsia*. Thus, the genomes and transcriptomes of *G. sinensis* are dominated by dinucleotide repeat units. The dominance of motifs in *G. sinensis* may be due to the motifs are rendered into certain amino acids in the process of translation, and the proteins that are present in these amino acids account for a large proportion of the species [16].

In recent years, research on *Gleditsia* has mainly focused on introduction and cultivation, reproductive technology, chemical composition analysis, stress resistance, and provenance test. The lack of molecular

Table 5 Statistics of female SSR clustering results

Items	Female specific SSR	
	Count	Percentage
SSR-Containing Sequence	241,081	100.00%
Cluster	122,999	51.02%

markers has limited the molecular identification and genetic diversity of *Gleditsia* germplasm resources and has affected the formulation of strategies for conserving *Gleditsia* germplasm resources. With the development of molecular biology, molecular markers are extensively investigated in various plant studies. The high polymorphism of SSR molecular markers combined with related biological analysis methods plays an important role in constructing a genetic map, genetic diversity determination, variety identification, seed purity detection, and marker-assisted breeding of medicinal plants. Lin et al. [20] constructed 18 DNA fingerprints of *G. sinensis* germplasm resources by using 8 polymorphic loci amplified by 3 primers, their results provided a theoretical basis for identifying the germplasm of *G. sinensis* and breeding new varieties [21]. However, to date, few reports on using microsatellite DNA sequence markers for early sex identification in plants are available. The application of DNA molecular marker technology in the sex identification of dioecious plants can improve the accuracy and reliability of the identification results, which are unaffected by development time and tissue specificity. The early sex identification of dioecious plants can obtain reliable identification information. Another study used a Microsatellite (GATA)₄ probe to identify 1 band of 5 kb only in male papaya plants [22]. This previous study indicated that microsatellite markers can be used to identify male and female papaya. The accuracy rate of SSR markers for early sex identification of *Populus*

Table 6 SSR length polymorphism assessment

Items	Female specific SSR	
	Count of clusters	Percentage
SSLP = 1	142,676	56.46%
SSLP = 2	54,669	21.64%
SSLP = 3	25,913	10.26%
SSLP = 4	12,456	4.93%
SSLP = 5	5793	2.29%
SSLP = 6	2975	1.18%
SSLP = 7	1719	0.68%
SSLP = 8	1251	0.50%
SSLP = 9	842	0.33%
SSLP ≥ 10	4387	1.74%
Total	252,681	100.00%

SSLP Length polymorphism of SSR in the same class

davidiana is 100% [23]. Our results are consistent with the 100% accuracy of SSR markers associated with the sex chromosomes of *G. sinensis*. At present, by combining expressed sequence tags (EST) and SSR markers, the shortcomings of high cost in developing traditional SSR markers have been solved. The EST has a higher occurrence than the genomic SSR markers among species. In addition, a previous report indicated a primer EST-Eu059 related to the sex of a male *Eucommia ulmoides* plant, which was screened using 140 pairs of EST-SSR primers, thereby indicating that identifying the sex of plants is feasible using EST-SSR molecular markers [24].

Currently, little is known about the sex determination and differentiation mechanisms in the dioecious plant *G. sinensis*. Intensive studies of the regulation mechanisms of sex expression will provide important insights into the genetic regulation of *G. sinensis* and are useful for genetic studies and breeding applications. Particularly, the transcriptome sequencing strategy has been applied to detect differentially expressed genes of different sex types, such as those in *Salix suchowensis* [25], *Asparagus officinalis* [26], *Diospyros kaki* [27], and so on. We are now using this method to detect the sex-biased expression genes, and further analysis will be expected to add new insights of the genetic regulation of *G. sinensis* sex expression.

Conclusions

The early morphological characteristics of *Gleditsia* plants are similar, thus limiting the potential use of these plants. In this study, SSR markers (i.e., 27, 34, 36, 39, and 41) related to the sex of a female *G. sinensis* plant have been screened out. These SSR markers may be used to detect the sex of a female plant in the early stage of development and have the advantages of 100% accuracy and efficiency. Simultaneously, the bands obtained through amplification demonstrate clear results, high repeatability, and robust stability. It provides accurate sex identification strategies between female and male plants, thus improving the utilization rate of *G. sinensis* resources.

Material and methods

Plant material

Complete and insect-free young leaves of *G. sinensis* (five males and five females) from the preserved forest at Xiaojing Town, Boai County, Henan Province, China, were collected to mark the sex and were brought to the laboratory in liquid nitrogen. In addition, 10 samples from Boai (five males and five males) and 20 samples from Xinxiang (10 males and females) were selected to validate the plant material and for further investigation. All land owners allow us to sample on their private land.

Table 7 Primer information and sequence

Primer	Upstream sequence	Downstream sequence	Tm (°C)	Size (bp)
Primer27	CGTCCGAGGACACGTAACCTT	GCCGTAGAAGCAGAGCAGTT	60.17	186
Primer34	TCATCCACTGCGACTTTTCAG	ACGTTTGCGTTGATACGTCA	59.98	305
Primer36	TTACCTAGATGGACCCGTGA	GGTTCTGTCCGTCTGGTAT	58.02	266
Primer39	GCGTTACGCCACGTTATATG	AGGTTGGTCACAGGATACGC	59.13	203
Primer41	CTTGCGAGAGTCCAACATGA	CATGAAGGTTGACTCCGGTT	59.98	260

The materials were identified as *G. sinensis* by Li at Henan Normal University, China. This material has been deposited in the biological specimen museum of Henan Normal University.

DNA extraction

A Plant Genomic DNA Kit (Beijing BIOTEKE Biotechnology Corporation) was used to extract DNA from young leaves of *G. sinensis* according to the kit instructions. Take the leaf of *G. sinensis* in a mortar, add liquid nitrogen to fully grind it into a fine powder, transfer it to a centrifuge tube, add 550 μ i Buffer P1 and 4 μ L RNase A, mix vigorously for 1 min, and leave it at room temperature for 10 min. Add 130 μ L of Buffer P2, mix vigorously for 1 min, and centrifuge at 12000 rpm for 3 min. Pipette the supernatant into separation column A, centrifuge at 12,000 rpm for 1 min, and collect the filtrate. Transfer the filtrate to a new centrifuge tube, add 1.5 volumes of Buffer P3 and mix well. Add the mixture to an adsorption column AC, centrifuge at 12,000 rpm for 1 min, and discard the waste in the collection tube. Add 500 μ L of rinse solution WB, centrifuge at 12,000 rpm for 1 min, and discard the waste solution. Put the adsorption column AC back into the empty collection tube, centrifuge at 13,000 rpm for 3 min, remove the adsorption column AC, put it in a clean centrifuge tube, add 50 μ L of elution buffer in the middle of the adsorption membrane, and leave it at room temperature for 3–5 min, Centrifuge at 12,000 rpm for 1 min to collect DNA. The purified DNA was quantified using a Nanodrop spectrophotometer.

Establishment of female and male DNA pools: According to the principle of BSA [28], an equal amount of DNA from the five female and male plants from Boai was mixed to form the female and male bulks, respectively. The two DNA bulks were fragmented by sonication. 400 bp fragments were recovered for library construction using Illumina Nextera library kit. Selective hybridization method (beads enrichment procedure) is used to enrich the SSR fragments in genomic libraries as described [29]. Eight probes, namely, P (AG)₁₀, P (AC)₁₀, P (AAC)₈, P (ACG)₈, P (AAG)₈, P (AGG)₈, and P (ACATAT)₆, were used for SSR enrichment. The enriched genomic libraries were sequenced with pair-end 250 bp (PE250) using an Illumina

MiSeq platform. Raw reads were filtered in accordance with the requirements.

SSR search and statistics

A microsatellite identification tool (MISA) [30] was used to search for SSR loci from all sequences. The parameters used in the search were mono-10, di-6, tri-5, tetra-5, penta-5, and hexa-5. The maximum allowable interval between two SSRs in the composite sequence is set to 100 bp.

SSR cluster analysis

We used Perl program to mask repetitive sequences and replaced them with the letter R. SSRs with a flanking sequence that is less than 20 bp were removed because short flanking sequences are unsuitable for alignment for similarity comparison. CDhit software was used to cluster the remaining sequences. Cluster analysis was performed with the sequence similarity of 95%, coverage of 70%, and vacancy penalty score of -gap 1-gep-ext 0. Among them, a sequence with two or more SSRs could be used to make separate statistical clustering groups.

Identification of sex-linked SSRs

To identify sex-specific SSRs, the SSR-containing sequences of the male and female plants were independently clustered by CDhit software. The similarity of a nucleotide sequence is set to 80%, the coverage is 70%, and the vacancy penalty score is set to -gap 1-gep-ext 0. The clusters only detected in the female plants were considered female-specific SSRs, whereas the clusters only detected in the male trees were considered male-specific SSRs.

SSR polymorphism assessment

Perl program was used to analyze the length of a repeat sequence in the clusters. If the lengths of repeat sequence in a cluster were the same, then the polymorphism of this cluster was scored as 1. If the lengths of the repeat sequence in a cluster were different, then the polymorphism of this cluster was scored as 2 if two lengths are available, 3 if three lengths are involved, and so on.

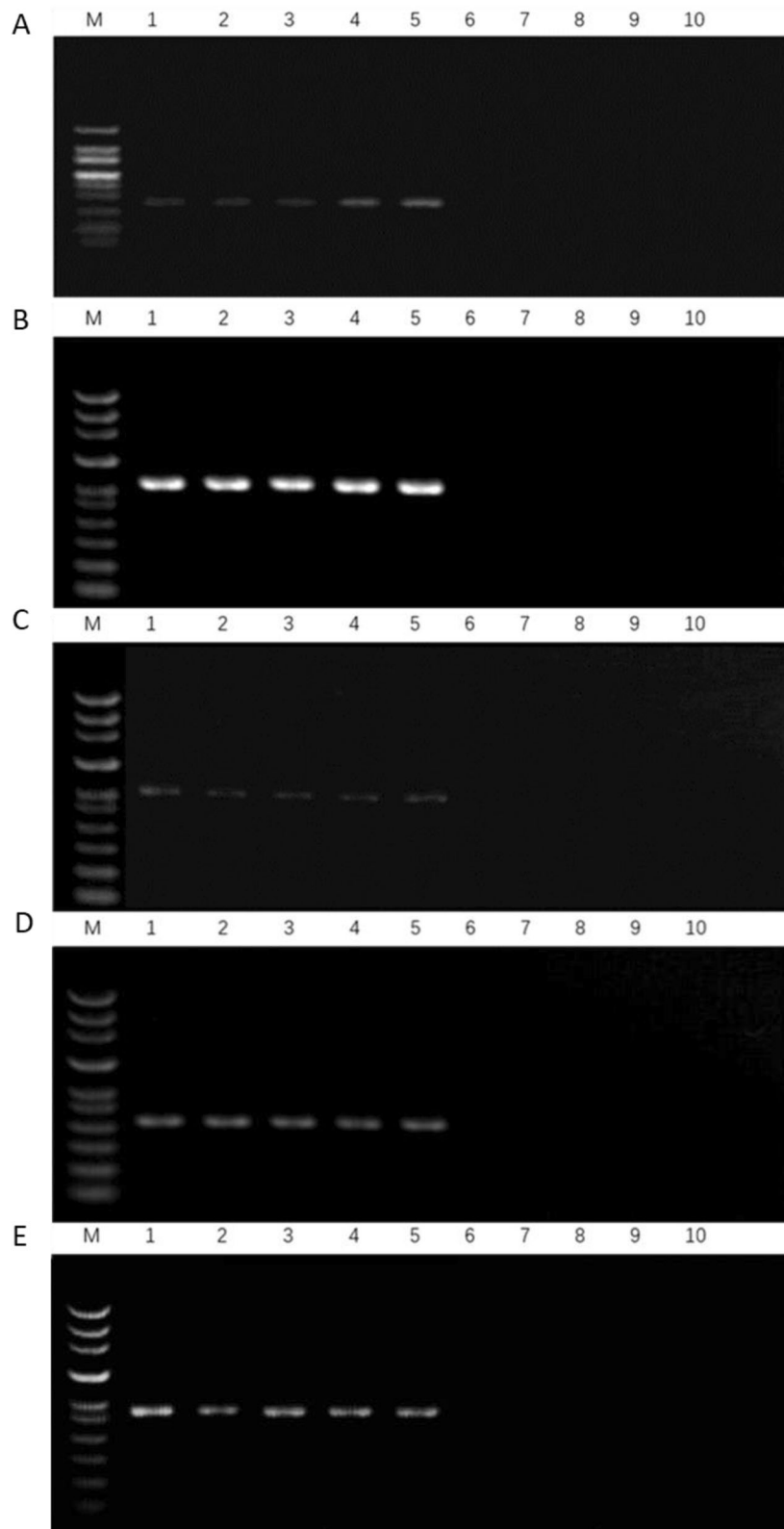


Fig. 2 Amplification results of the five pairs of primers in 10 *G. sinensis*. **a** Primer27; **b** Primer34; **c** Primer36; **d** Primer39; **e** Primer41; (M) makerDL700; 1–5, female plant; 6–10, male plant

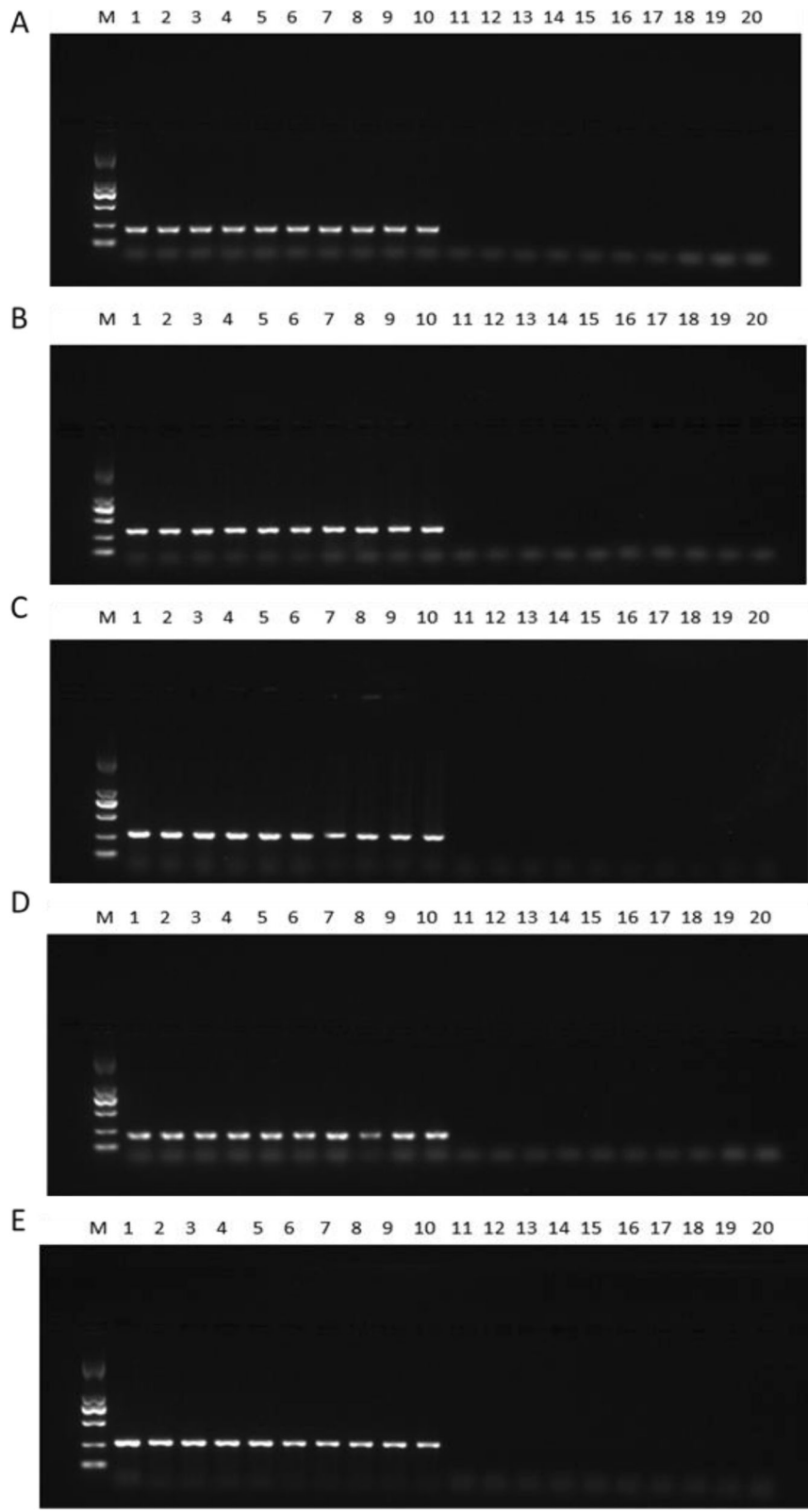


Fig. 3 Amplification results of the five pairs of primers in 20 *G. sinensis*. **a** Primer27; **b** Primer34; **c** Primer36; **d** Primer39; **e** Primer41; (M) makerDL2000; 1–10, female plant; 11–20, male plant

SSR primer design

Primer 3 v2.3.6 was used to design primers [15] for SSR sequences in a cluster with a polymorphism score of ≥ 2 . The PCR product was controlled in 100–400 bp, and the target fragment was amplified from the first base of the repeat sequence to the last five bases of the repeat sequence. Moreover, other parameters were defaults. All the primers were screened with the following criteria: (1) Single-nucleotide and complex repeat motifs are excluded from the primer design. (2) Each SSR marker only contains one SSR type to avoid the influence on polymorphism through the second SSR during amplification. (3) The designed primers must be within a cluster, and the length polymorphism of SSRs in clusters is more than 2. (4) Every designed primer must support two primers in clusters. (5) Identical primer results are removed.

PCR amplification and agarose gel electrophoresis

A Kit (RR002A) from TAKARA was used for PCR amplification, PCR amplification was performed in a volume of 25 μL with 1.0 μL DNA template (40 $\text{ng}\cdot\mu\text{L}^{-1}$), 15.5 μL ddH_2O , 2.0 μL dNTP (0.00625 $\text{mM}/\mu\text{L}$), 1.0 μL forward primers (5 pmol), 1.0 μL reverse primers (5 pmol), 2.5 μL $10\times$ buffer, 1.5 μL MgCl_2 (2.5 mM/mL), and 0.5 μL Taq DNA polymerase (5 U/ μL). The PCR program is described as follows: denaturation at 95 °C for 5 min, then repeat with 30 cycles for 30 s at 95 °C, 52 °C, and 72 °C, with a final extension of 7 min at 72 °C. PCR products were electrophoretically separated on 1% agarose gels. Product sizes were estimated by comparison with a 700 bp DNA ladder.

Abbreviations

SSR: Simple sequence repeat; BSA: Bulk segregant analysis; *G. sinensis*: *Gleditsia sinensis* Lam. (Leguminosae); EST: Expressed sequence tags

Acknowledgments

At the point of finishing this paper, I'd like to express my sincere thanks to all those who have lent me hands in the course of my writing this paper.

Authors' contributions

LJJ designed the research. YCL performed the experiments. All authors analyzed and discussed the data. YCL prepared the figures and LJJ wrote the article. All Authors read and approved the final manuscript.

Funding

This work was funded by the Chinese Medicine Public Health Service Subsidy Special "National Chinese Medicine Resources Census Project" (Caishe [2017] No. 66). There is no role of the funding body in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The datasets has been uploaded to the NCBI SRA database. SRA accession: PRJNA644493. SRA records will be accessible with the following link: <https://www.ncbi.nlm.nih.gov/sra/PRJNA644493>

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 17 April 2019 Accepted: 12 July 2020

Published online: 17 July 2020

References

- Zhang JP, Tian XH, Yang YX, Liu QX, Wang Q, Chen LP, Li HL, Zhang WD. *Gleditsia* species: an ethnomedical, phytochemical and pharmacological review. *J Ethnopharmacol.* 2015;178:155–71.
- Li G, Wang ZP, Xian YX, Zhou HL, Wang X, Liu W, Yu JQ. Flavonoids from *Gleditsia spina* and their antitumor activities. *Chin Tradition Herbal Drugs.* 2016;47(16):2812–6.
- Li J, Jiang K, Wang LJ, Yin G, Wang J, Wang Y, Jin YB, Li Q, Wang TJ. HPLC-MS/MS determination of flavonoids in *Gleditsia spina* for its quality assessment. *J Sep Sci.* 2018;41(8):1752–63.
- Li SF, Su T, Cheng GQ, Wang BX, Li X, Deng CL, Gao WJ. Chromosome evolution in connection with repetitive sequences and epigenetics in plants. *Genes.* 2017;8:290.
- Samoluk SS, Chalup LMI, Chavarro C, Robledo G, Bertoli DJ, Jackson SA, Seijo G. Heterochromatin evolution in *Arachis* investigated through genome-wide analysis of repetitive DNA. *Planta.* 2019;249:1405–15.
- Zhang ZC, Hou XL. Strategies for development of SSR molecular markers. *Hereditas.* 2004;26(5):763–8.
- Xiao Y, Xia W, Ma J, Mason AS, Fan H, Shi P, Lei X, Ma Z, Peng M. Genome-wide identification and transfer ability of microsatellite markers between palmar species. *Front Plant Sci.* 2016;10(7):1578–88.
- McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y, Zhang Q, Kono I, Yano M, Fjellstrom R, DeClerck G, Schneider D, Cartinhour S, Ware D, Stein L. Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Res.* 2002;9(6):199–207.
- Kirungu JN, Deng Y, Cai X, Magwanga RO, Zhou Z, Wang X, Wang Y, Zhang Z, Wang K, Liu F. Simple sequence repeat (SSR) genetic linkage map of D genome diploid cotton derived from an interspecific cross between *Gossypium davidsonii* and *Gossypium klotzschianum*. *Int J Mol Sci.* 2018;19(1):204.
- Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *Plant Cell.* 2007;19(11):3403–17.
- Siew GY, Ng WL, Tan SW, et al. Genetic variation and DNA fingerprinting of durian types in Malaysia using simple sequence repeat (SSR) markers. *Peer J.* 2018;2(6):e4266.
- Han S, Wu Z, Wang X, et al. De novo assembly and characterization of *Gleditsia sinensis* transcriptome and subsequent gene identification and SSR mining. *Genet Mol Res.* 2016;15(1):1–12.
- Lawson MJ, Zhang L. Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. *Genome Biol.* 2006;7(2):14–24.
- Iniguez-Luy FL, Voort AV, Osborn TC. Development of a set of public SSR markers derived from genomic sequence of a rapid cycling *Brassica oleracea* L. genotype. *Theor Appl Genet.* 2008;117(6):977–85.
- Shi J, Huang S, Zhan J, Yu J, Wang X, Hua W, Liu S, Liu G, Wang H. Genome-wide microsatellite characterization and marker development in the sequenced *Brassica* crop species. *DNA Res.* 2013;21(1):53–68.
- Zhu H, Song P, Koo DH, Guo L, Li Y, Sun S, Weng Y, Yang L. Genome wide characterization of simple sequence repeats in watermelon genome and their application in comparative mapping and genetic diversity analysis. *BMC Genomics.* 2016;17(1):557.
- Liu S, et al. Genome-wide identification of simple sequence repeats and development of polymorphic SSR markers for genetic studies in tea plant (*Camellia sinensis*). *Mol Breed.* 2018;38:59.
- Pandey G, Misra G, Kumari K, Gupta S, Parida SK, Chattopadhyay D, Prasad M. Genome-wide development and use of microsatellite markers for large-scale genotyping applications in foxtail millet [*Setaria italica* (L.)]. *DNA Res.* 2013;20(2):197–207.
- Lin FR, Xing JL, Meng YQ, Huang P, Zheng XQ. Development and assessment of EST-SSR for *Gleditsia sinensis* lam. Based on transcriptome sequences. *J Plant Genet Res.* 2017;18(1):148–54.
- Zhang AS, Luo Y, Fan DC, Zhang ZH. Genetic diversity and fingerprints of *Gleditsia sinensis* germplasm resource based on SCoT. *Guihaia.* 2017;37(11):1378–85.

21. Parasnis AS, Ramakrishna W, Chowdari KV, Gupta VS, Ranjekar PK, Ramakrishna W. Microsatellite (GATA)_n reveals sex-specific differences in papaya. *Theor Appl Genet.* 1999;99(6):1047–52.
22. Pakul B, Groppe K, Meyer M, Markussen T, Fladung M. Genetic linkage mapping in aspen (*Populus tremula* L. and *Populus tremuloides* Michx.). *Tree Genet Genomes.* 2009;5(3):505–15.
23. Lin KQ, Li Y, Zhao DG. SSR mining and marker development in *Eucommia ulmoides* Oliver transcriptome. *Mol Plant Breed.* 2016;14(6):1548–58.
24. Meng LZ, Xiang C, Liu HW, Yang L, Mai CY, Yu LQ, Wei YL, Li HJ, Zhang HJ, Zhou Y. The impact of modern plant breeding on dominant Chinese wheat cultivars (*Triticum aestivum* L.) revealed by SSR and functional markers. *Genet Resour Crop Evol.* 2018;65(1):55–65.
25. Liu J, Yin T, Ye N, Chen Y, Yin T, Liu M, Hassani D. Transcriptome analysis of the differentially expressed genes in the male and female shrub willows (*Salix suchowensis*). *PLoS One.* 2013;8:e60181.
26. Li SF, Zhang GJ, Zhang XJ, Yuan JH, Deng CL, Gao WJ. Comparative transcriptome analysis reveals differentially expressed genes associated with sex expression in garden asparagus (*Asparagus officinalis*). *BMC Plant Biol.* 2017;17:143.
27. Li S, Sun P, Du G, Wang L, Li H, Fu J, Suo Y, Han W, Diao S, Mai Y, Li F. Transcriptome sequencing and comparative analysis between male and female floral buds of the persimmon (*Diospyros kaki* Thunb.). *Sci Hortic.* 2019;246:987–97.
28. Xu XY, Sun XL, Zhang Y, Zhang LG, Fang ZY, Friedt W. Identification of AFLP and SSR markers linked with the male fertility restorer gene of CMS 06J45 in heading Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Plant Breed.* 2014;133(5):615–9.
29. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics.* 2011;27(21):2957–63.
30. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 2012;40(15):115.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

