Transcriptome Characterization and SSR Marker Development in Choerospondias axillaris (Post-print)

Authors: Xu Mengyang, Cai Yuyu, Li Ting, Wu Nansheng, Sun Rongxi

Date: 2025-03-19T00:00:00+00:00

Abstract

Through analysis of male and female transcriptome characteristics and development of SSR markers in Choerospondias axillaris, this study provides theoretical support and scientific basis for its genetic evaluation and molecular-assisted breeding for sex. This paper mainly analyzes differential expression between male and female transcriptomes of Choerospondias axillaris, distribution and sequence characteristics of SSR loci, conducts SSR loci mining, and performs development and validation of SSR primers. The results showed: (1) Transcriptome sequencing of Choerospondias axillaris yielded a total of 40,341 unigenes, with total length, N50, average length, and GC content of 52,806,369 bp, 2,409 bp, 1,309 bp, and 38.75%, respectively. A total of 1,949 differentially expressed genes between male and female were screened, among which 1,052 genes were significantly up-regulated and 897 genes were down-regulated in male plants compared to female plants. (2) A total of 5,251 SSR loci were detected across all unigenes, with 619 unigene sequences containing two or more loci, the SSR occurrence frequency was 11.18%, and the average distribution distance was 10.06 kb. Among all SSR loci, dinucleotide repeats accounted for the highest proportion (46.95%), followed by trinucleotide repeats (34.27%). (3) Through screening and validation, a total of 20 pairs of SSR polymorphic primers were obtained, which detected 80 alleles across 85 germplasm resources, with an average polymorphic information content (PIC) of 0.56. In conclusion, the leaf transcriptome sequencing of Choerospondias axillaris achieved high quality with good assembly results. The 20 pairs of polymorphic primers provide a reference for future genetic evaluation, sex-assisted breeding, and fingerprint map construction.

Full Text

Transcriptome Characteristic Analysis and SSR Marker Development of *Choerospondias axillaris*

Authors: XU Mengyang[1,2], CAI Yuyu[1,2], LI Ting[1,2], WU Nansheng[1,2], SUN Rongxi[1,2*]

Affiliations:

1. Institute of *Choerospondias axillaris* Research, College of Forestry, Jiangxi Agricultural University, Nanchang 330045, China
2. Jiangxi Provincial Key Laboratory of Subtropical Forest Resources Cultivation, College of Forestry, Jiangxi Agricultural University, Nanchang 330045, China

Abstract: This study analyzed the transcriptome characteristics and developed SSR markers for *Choerospondias axillaris* to provide theoretical support and a scientific basis for genetic evaluation and sex-assisted molecular breeding. We analyzed differential expression between male and female transcriptomes, examined the distribution and sequence characteristics of SSR loci, and conducted SSR mining, primer development, and validation. The results showed: (1) Transcriptome sequencing yielded 40,341 Unigenes with a total length of 52,806,369 bp, N50 length of 2,409 bp, average length of 1,309 bp, and GC content of 38.75%. A total of 1,949 differentially expressed genes were identified between sexes, with 1,052 genes significantly upregulated and 897 downregulated in males compared to females. (2) Among all Unigenes, 5,251 SSR loci were detected, with 619 Unigenes containing two or more loci. The SSR occurrence frequency was 11.18%, with an average distribution distance of 10.06 kb. Dinucleotide repeats were most abundant (46.95%), followed by trinucleotide repeats (34.27%). (3) After screening and validation, 20 polymorphic SSR primer pairs were obtained, detecting 80 alleles across 85 samples with an average polymorphism information content (PIC) of 0.56. In summary, the leaf transcriptome sequencing of *C. axillaris* demonstrated high quality and good assembly performance. These 20 polymorphic primers provide valuable resources for future genetic evaluation, sex-assisted breeding, and fingerprinting studies.

Keywords: *Choerospondias axillaris*, primer development, transcriptome, SSR, locus characteristics

---

Choerospondias axillaris (Roxb.) B.L. Burtt & A.W. Hill belongs to the family Anacardiaceae and is a traditional medicinal plant in Mongolian medicine. Its bark is used for tannin extraction, while its fruits are edible and medicinal, and the wood has high economic value. Widely distributed across southern China, *C. axillaris* is an important fast-growing timber species with significant economic, ecological, and social benefits. As a dioecious species endemic to East Asia, it constitutes an important component of subtropical and tropical deciduous broad-leaved forests. Paleogeographic studies suggest that the genus

Choerospondias likely originated in early Eocene Europe and dispersed along the warm and humid Tethys Ocean to Asia, with the current East Asian distribution pattern possibly established before the late Miocene. Fossil records from the Oligocene onward indicate high adaptability, with Miocene fossils from Fujian showing greater morphological diversity than modern descendants. The genus shares similar distribution patterns with *Rhus* in Anacardiaceae, which originated in early Eocene western North America and became disjunctly distributed across Asia and North America during the middle Eocene. The unique topography and climate of eastern Asia may have served as refugia for both genera, with *C. axillaris* persisting in southern China long after its disappearance from Europe.

Despite its strong adaptability to varying temperature and precipitation regimes, *C. axillaris* exhibits substantial phenotypic variation in growth characteristics, nutritional components, stress resistance, and phenology across different regions. As a dioecious species with long-term outcrossing, it maintains highly heterozygous genotypes. However, the lack of reference genomes from congeneric species and limited molecular markers have constrained genetic evaluation and molecular-assisted breeding efforts, hindering analysis of its evolutionary potential and future development prospects.

Genetic diversity is a crucial indicator of a species' evolutionary potential and adaptive capacity, providing a theoretical foundation for utilization and development. Molecular marker-based genetic diversity analysis not only informs germplasm collection, preservation, evaluation, and utilization but also establishes a basis for advanced genetic research. Previous studies have optimized ISSR-PCR systems, conducted preliminary SSR marker exploration, and performed transcriptome sequencing of leaves and fruits, while the complete chloroplast genome has also been sequenced. Nevertheless, sequence data remain unavailable in public databases, and molecular marker information has not been explicitly reported, limiting molecular-level genetic diversity studies.

Microsatellite markers (SSR) are among the most widely used molecular markers due to their high mutation rates, polymorphism, stability, reproducibility, and broad distribution across genomes. They are extensively applied in genetic evaluation, cultivar identification, genetic map construction, QTL mapping, and conservation of endangered species. SSR loci typically provide higher polymorphism than other markers, better reflecting genetic differentiation within and among populations and offering more accurate evolutionary information. For non-model species without reference genomes, transcriptome-based marker development is considered the most effective approach. This study utilized transcriptome data from male and female leaves of *C. axillaris* to analyze sex-biased expression, characterize SSR distribution patterns, and develop validated SSR markers to establish a foundation for genetic diversity assessment, sex marker development, and fingerprinting.

1.1 Experimental Materials

Transcriptome sequencing materials were collected from the *C. axillaris* germplasm resource nursery at Jiangxi Agricultural University. Fresh leaves were sampled from six 10-year-old clones (three females: Female-1, Female-2, Female-3; three males: Male-1, Male-2, Male-3), immediately frozen in liquid nitrogen, and stored on dry ice for transcriptome sequencing. For primer screening, one sample each was collected from eight geographically distant populations: Chongyi, Jiangxi (JXCY); Jiangkou, Guizhou (GZJK); Sangzhi, Hunan (HNSZ); Mangshi, Yunnan (YNMS); Pingnan, Guangxi (GXPN); Chibi, Hubei (HBCB); Qimen, Anhui (AHQM); and Shapingba, Chongqing (CQSPB). For polymorphism validation, 85 samples were obtained from nine populations across three provinces (Table 1). Healthy, disease-free young leaves from all samples were rapidly dried with silica gel and stored for DNA extraction.

1.2 Experimental Methods

1.2.1 Transcriptome Assembly and Analysis Raw data from high-throughput sequencing were filtered to obtain high-quality clean data. Trinity software (Grabherr et al., 2011) was used for de novo transcriptome assembly, generating transcript sequences and Unigene information. Clean reads from each sample were mapped to the assembled Unigenes to obtain read counts, which were then converted to FPKM (Fragments Per Kilobase Million) values to quantify gene expression levels.

1.2.2 SSR Locus Mining and Primer Design MISA software (http://pgrc.ipk-gatersleben.de/misa/misa.html) was used to search for SSR loci in the assembled Unigenes, with repeat parameters set to 10, 6, 5, 5, 5, and 5 for mono- to hexanucleotide repeats, respectively (Pan et al., 2019). SSR distribution characteristics and repeat types were statistically analyzed. Primer3.0 (Rozen & Skaletsky, 2000) was employed for batch primer design with the following parameters: annealing temperature 50–60°C, GC content 40–60%, primer length 18–27 bp, and product length 100–300 bp. One hundred primer pairs targeting differentially expressed genes between sexes were synthesized by Wuhan Tsingke Biotechnology Co., Ltd.

1.2.3 Genomic DNA Extraction and Quality Assessment Genomic DNA was extracted using the CTAB method. DNA concentration and quality were assessed using a UV spectrophotometer and 1% agarose gel electrophoresis. Samples with clear bands, high purity, and good integrity were diluted to 50 $ng \cdot \mu L^{-1}$ and stored at -20°C for subsequent experiments.

1.2.4 Primer Screening and PCR Amplification Initial screening was performed using 1.5% agarose gel electrophoresis with two randomly selected samples. Primers producing clear, stable bands within the expected size range were considered qualified. The PCR reaction system contained 12.5 L 2×Taq

PCR Master Mix, 1 L each of forward and reverse primers (10 μmol · L^{-1}), 1 L DNA template, and 9.5 L double-distilled water. Touchdown PCR was performed as follows: 94°C for 5 min; 10 cycles of 94°C for 30 s, 63°C for 30 s, and 72°C for 45 s; 20 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 45 s; final extension at 72°C for 7 min; hold at 4°C (Sun, 2017). For secondary screening, eight samples from distant populations were analyzed via capillary electrophoresis to select primers with strong peak signals and no background peaks.

For polymorphism validation using 85 samples from nine populations, a three-primer system was employed for capillary electrophoresis: a forward primer with an M13 tail (5 -TGTAAAACGACGGCCAGT-3), a reverse primer, and a fluorescently labeled M13 tail primer (HEX, MRA, FAM, or ROX). Two-step PCR was conducted: Step 1 used a 10 μL reaction containing 5 L 2×Taq PCR Master Mix, 0.1 μL each of M13-tailed forward and reverse primers (10 μmol · L^{-1}), 1 L DNA template, and 3.8 L double-distilled water, with cycling conditions of 95°C for 5 min; 20 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 30 s; final extension at 72°C for 10 min. Step 2 used a 20 μL reaction containing 10 L 2×Taq PCR Master Mix, 0.15 μL fluorescently labeled M13 primer (FAM, TAMRA, or ROX), 0.15 μL reverse primer (10 μmol · L^{-1}), 2 L of the Step 1 PCR product, and 7.7 L double-distilled water, with cycling conditions of 95°C for 5 min; 35 cycles of 95°C for 30 s, 52°C for 30 s, and 72°C for 30 s; final extension at 72°C for 10 min; hold at 4°C (Sun, 2017).

1.2.5 Capillary Electrophoresis Sequencing A mixture of formamide and GeneScan-500LIZ internal size standard (100:1 v/v) was prepared, and 10 μL was added to each well followed by 0.3 μL of PCR product. Samples were denatured at 95°C for 5 min, cooled at 4°C, centrifuged, and then analyzed using an ABI 3730xL DNA Analyzer (Applied Biosystems, Foster City, CA, USA) with two- or three-color multiplexing.

1.3 Data Analysis

SSR distribution characteristics were analyzed with occurrence frequency calculated as the ratio of total SSR loci to total Unigene sequences (Huang et al., 2019). GeneMaker 2.2.0 software (Lukashin & Borodovsky, 1998) was used for genotyping from raw capillary electrophoresis peak data. GenAlEx 6.5 (Peakall & Smouse, 2012) calculated genetic parameters including expected heterozygosity (He), observed heterozygosity (Ho), effective allele number (Ne), and allele number (Na). PowerMarker V3.25 (Liu & Muse, 2005) computed polymorphism information content (PIC).

2.1 Transcriptome Data Assembly and Analysis

All six samples yielded over 5.4 Gb of clean data, totaling 35.4 Gb with Q20 and Q30 values exceeding 96.64% and 90.97%, respectively, and GC content around

42%, indicating high sequencing quality. All raw transcriptome data have been submitted to NCBI (PRJNA1142239). Female leaf samples (Female-1, Female-2, Female-3) generated 44,329,038; 41,676,510; and 40,128,104 clean reads, respectively, while male samples (Male-1, Male-2, Male-3) produced 36,771,300; 36,540,830; and 37,214,014 clean reads, totaling 236,659,796 clean reads.

Trinity assembly yielded 40,341 Unigenes with a total length of 52,806,369 bp, average length of 1,309 bp, and GC content of 38.75%. The N50 statistic comprised 7,123 Unigenes with a length of 2,409 bp, while maximum and minimum lengths were 16,889 bp and 201 bp, respectively. Length distribution analysis revealed that 15,482 Unigenes (38.38%) were 201–500 bp, 8,371 (20.75%) were 501–1,000 bp, 7,109 (17.62%) were 1,001–2,000 bp, and 9,379 (23.25%) exceeded 2,000 bp.

Mapping statistics showed that 40,341 genes were detected across all six samples, with female samples mapping to over 72% of total genes and male samples mapping to approximately 70% (Table 3).

2.2 Differential Gene Expression Analysis

Clustering analysis based on $\log_{10}$(FPKM+1) values was performed to quantify expression levels. Using a false discovery rate P-value < 0.05 and $|\log_2\text{FoldChange}| > 1$ as screening criteria, 1,949 differentially expressed genes were identified between female and male leaf groups, including 1,052 upregulated and 897 downregulated genes in males relative to females (Figure 1 [Figure 1: see original paper]).

2.3 SSR Quantity and Distribution in the Transcriptome

MISA analysis of 40,341 Unigenes detected 5,251 SSR loci distributed across 4,511 Unigenes, with an SSR occurrence frequency of 13.02% and a frequency of 11.18%. Among these, 3,892 Unigenes contained a single SSR locus, 619 contained two or more loci, 386 contained compound SSRs, and 18 had overlapping compound SSRs (Table 4).

Excluding mononucleotide repeats, dinucleotide repeats were most abundant (42.93% of total SSR loci) with an average distribution distance of 23.43 kb, followed by trinucleotide (31.06%, 32.38 kb) and tetranucleotide repeats (11.52%, 87.28 kb). Pentanucleotide and hexanucleotide repeats were rare (3.52% and 3.28%, respectively), with average distances of 285.44 kb and 307.01 kb. The frequency of dinucleotide repeats was highest at 5.59%, followed by trinucleotide (4.04%), while tetra-, penta-, and hexanucleotide repeats showed lower frequencies of 1.50%, 0.46%, and 0.43%, respectively. Compound and overlapping compound repeats accounted for 7.35% and 0.34% of loci, with frequencies of 0.96% and 0.04%, respectively.

Total SSR sequence length was 117,134.87 bp, with dinucleotide repeats comprising the longest region (41,248.20 bp), followed by trinucleotide (29,472.17

bp), tetranucleotide (10,545.15 bp), pentanucleotide (3,975.65 bp), and hexanucleotide repeats (4,434.16 bp). The average SSR length was 22.31 bp, with one SSR locus per 10.06 kb on average (Table 5).

2.4 SSR Motif Length and Repeat Characteristics

SSR loci with motif lengths < 20 bp were most common (3,601 loci, 68.58%), followed by 21-30 bp loci (1,112, 21.18%). Loci > 30 bp represented only 10.24% of the total. Among 5,251 SSR loci, 412 repeat motif types were identified. Dinucleotide repeats comprised 10 types, with AG/CT being dominant (1,309 loci, 24.93%), followed by AT/AT (730, 13.90%). Trinucleotide repeats included 60 types, with AAG/CTT as the major type (556, 10.59%) and AAT/ATT second (277, 5.28%). Tetranucleotide repeats comprised 95 types, led by AAAT/ATTT (312, 5.94%) and AAAG/CTTT (103, 1.96%). Pentanucleotide repeats included 99 types, with AAAAT/ATTTT predominant (53, 1.01%). Hexanucleotide repeats comprised 148 types, while compound and overlapping compound motifs included 404 and 18 types, respectively, all with frequencies below 0.5% (Figure 2 [Figure 2: see original paper]).

2.5 SSR Repeat Numbers in the Transcriptome

Excluding compound loci, SSR repeat numbers ranged from 4 to 29 across 2-6 nucleotide motifs. The most frequent repeat number was 6 tandem repeats (1,100 loci, 22.69%), followed by 5 repeats (1,057, 21.81%) and 4 repeats (727, 15.00%). Dinucleotide repeats predominantly showed 6 tandem repeats, trinucleotide repeats showed 5 repeats, and tetra-, penta-, and hexanucleotide repeats primarily showed 4 repeats (Table 6). SSR abundance decreased with increasing repeat number.

2.6.1 Primer Primary and Secondary Screening

Among 100 SSR primer pairs, 55 successfully amplified clear target bands in initial screening (55% success rate; 24 examples shown in Figure 3 [Figure 3: see original paper]). Secondary screening via capillary electrophoresis identified 20 highly polymorphic primer pairs (36.36% of effective primers), with representative results for primer NSZ-9141 shown in Figure 4 [Figure 4: see original paper].

2.6.2 SSR Locus Population Genetic Analysis Validation

Validation using 85 samples from nine populations across Hubei, Hunan, and Guangxi detected 80 alleles (Na) across the 20 primer pairs, averaging 3.82 alleles per locus. Mean effective allele number (Ne) was 2.61 (range: 1.14 for NSZ-22900 to 7.16 for NSZ-131). Shannon' s information index (I) ranged from 0.21 (NSZ-22900) to 2.06 (NSZ-131), averaging 0.98. Mean observed (Ho) and expected (He) heterozygosity were both 0.53. PIC values ranged from 0.12

(NSZ-22900) to 0.92 (NSZ-131), averaging 0.56, with 15 loci showing high polymorphism (PIC > 0.5) and four showing moderate polymorphism (0.25 < PIC ≤ 0.5) (Table 7). These results confirm that the 20 SSR primer pairs are suitable for subsequent molecular genetic diversity evaluation in *C. axillaris*.

Discussion

For non-model species lacking reference genomes, transcriptome-based marker development is the most effective approach. This study generated 236,659,796 clean reads from male and female leaf transcriptomes of *C. axillaris*, with an N50 length of 2,409 bp and average length of 1,309 bp, which exceeds values reported for mango (Mangifera indica) (N50 = 2,160 bp; average = 1,219 bp) and previous *C. axillaris* transcriptome studies (N50 = 1,841 bp; average = 1,287 bp), as well as other broadleaf species such as *Castanopsis carlesii* (N50 = 1,427 bp; average = 1,034 bp) and *Melia azedarach* (N50 = 1,955 bp; average = 1,431.82 bp). These metrics indicate high sequencing quality and effective assembly, providing a robust foundation for molecular marker development.

We identified 5,251 SSR loci in the *C. axillaris* leaf transcriptome with an average distribution distance of 10.06 kb and an occurrence frequency of 11.18%. This frequency is higher than that reported for loquat (Eriobotrya japonica) (6.77%) and mango (7.57%), though lower than a previous *C. axillaris* study (25.52%). Such interspecific variation may arise from differences in tissue type, gene expression levels, SSR search criteria, and Unigene quantity and length. Most plants, including *Myrica rubra*, mango, and *Trifolium repens*, predominantly feature di- and trinucleotide repeats. Similarly, our study found dinucleotide repeats most frequent (46.95%), followed by trinucleotides (34.27%), together accounting for 81.22% of all SSRs. Although some species show predominance of mononucleotide repeats (e.g., *Lycium chinense*, *Platanus acerifolia*), these were excluded from our analysis due to their inherent limitations.

Polymorphism is a key indicator of marker utility, with SSR length being a critical factor. Highly polymorphic SSRs typically exceed 20 bp, moderately polymorphic loci range 12–20 bp, and low-polymorphism loci are <12 bp. In *C. axillaris*, 1,112 SSR loci (21.18%) fell within the 21–30 bp range, indicating high polymorphism potential. Our 20 validated primer pairs detected 80 alleles across nine populations, with mean Ho and He of 0.53 and average PIC of 0.56. Nineteen loci showed moderate to high polymorphism, confirming their suitability for downstream applications. Repeat motif length correlates with selective pressure, with shorter motifs evolving more rapidly and showing higher polymorphism. Lower-order repeats (mono-, di-, and trinucleotides) exhibit longer evolutionary histories and higher mutation rates. The predominance of dinucleotide repeats in *C. axillaris* suggests relatively high evolutionary rates, consistent with its ancient lineage and continuous presence in southern China since the Miocene. This long evolutionary history, combined with high adaptability, may underlie the species' high polymorphism levels.

In conclusion, the *C. axillaris* transcriptome yielded an SSR occurrence frequency of 11.18% with dinucleotide and trinucleotide repeats predominating. The 20 polymorphic primer pairs detected 80 alleles with an average PIC of 0.56, demonstrating their effectiveness for future studies on genetic diversity evaluation, sex-assisted molecular breeding, fingerprint construction, and core germplasm development in *C. axillaris*.

References

CAI JF, YANG XM, YU WW, et al., 2021. Development of SSR molecular markers based on transcriptome sequencing of *Melia azedarach* [J]. *Scientia Silvae Sinicae*, 57(6): 85-92.

CHEN L, YU FL, YAO MZ, et al., 2008. Preparation of the UPOV guidelines for the conduct of tests for distinctness, uniformity and stability—tea plant [J]. *Scientia Agricultura Sinica*, 41(8): 2400-2406.

DING YG, HOU Q, ZHOU L, et al., 2024. Transcriptome analysis of the effects of foliar water application on the nap time of mango leaves [J]. *Acta Botanica Boreali-Occidentalia Sinica*, 44(7): 1094-1104.

Editorial Committee of Flora of China, Chinese Academy of Sciences, 1980. *Flora Reipublicae Popularis Sinicae*: Vol. 45, Part one [M]. Beijing: Science Press: 86.

GAO Y, BA YF, LI HY, et al., 2021. Changes of color and texture of *Choerospondias axillaris* fruit at different storage temperatures [J]. *Non-wood Forest Research*, 39(4): 239-245.

GRABHERR MG, HAAS BJ, YASSOUR M, et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome [J]. *Nature Biotechnology*, 29(7): 644.

GU ZJ, YANG CX, DING W, et al., 2019. Characterization analysis of SSR sequences in the transcriptome of *Choerospondias axillaris* and its molecular marker development [J]. *South China Forestry Science*, 47(3): 12-15.

GUO JH, LIU SJ, WU YH, et al., 2019. Genetic structure of *Toxicodendron vernicifluum* in southern Shanxi Province based on SSR marker [J]. *Molecular Plant Breeding*, 17(9): 2950-2955.

HUANG JM, LI F, LIU YJ, et al., 2019. Bioinformatic analysis of SSR information in *Tirpitzia sinensis* transcriptome [J]. *Jiangsu Agricultural Sciences*, 47(2): 54-58.

KURODA Y, KAGA A, TOMOOKA N, et al., 2006. Population genetic structure of Japanese wild soybean (*Glycine soja*) based on microsatellite variation [J]. *Molecular Ecology*, 15(4): 959-974.

LI D, WU NS, GUO CL, et al., 2020. Investigation on the pest species of Lapsi *Choerospondias axillaris* ( ): Wood-boring pests and shoot pests [J]. *Biological*

Disaster Science, 43(3): 239-243.

LI JE, DONG JJ, XIE MZ, et al., 2024. Antioxidant, hypoglycemic and hypolipidemic effects of *Choerospondias axillaris* polysaccharide in vitro [J]. *Food Research and Development*, 45(3): 59-66.

LI RX, ZHANG N, LUO XD, et al., 2023. Effects of different lactic acid bacteria fermentation on physicochemical indexes and functional components of *Choerospondias axillaris* beverage [J]. *China Brewing*, 42(6): 161-165.

LIANG X, WU Z, GUO L, et al., 2022. Late Miocene leaves and endocarps of *Choerospondias* (Anacardiaceae) from Zhejiang, eastern China: implications for paleogeography and paleoclimate [J]. *Biology*, 11(10): 1399.

LIU K, MUSE SV, 2005. Power Marker: an integrated analysis environment for genetic marker analysis [J]. *Bioinformatics*, 21(9): 2128-2129.

LUKASHIN AV, BORODOVSKY M, 1998. GeneMark.hmm: new solutions for gene finding [J]. *Nucleic Acids Research*, 26(4): 1107-1115.

LUO C, WU HX, YAO QS, et al., 2015. Data mining and primer screening of simple sequence repeats in mango transcriptome [J]. *Chinese Journal of Tropical Crops*, 36(7): 1261-1266.

MAI KL, ZHU CS, FENG LX, et al., 2022. Early performances of *Toona sinensis* and *Choerospondias axillaris* among different provenances from northwestern Guangxi, China [J]. *Guangxi Forestry Science*, 51(2): 236-240.

MAGURRAN AE, 1988. *Ecological Diversity and Its Measurement* [M]. Dordrecht: Springer: 179.

PAN LQ, LI JY, LI SC, et al., 2019. Development of SSR markers based on transcriptome of *Camellia japonica* and analysis of genetic relationship [J]. *Journal of Beijing Forestry University*, 41(7): 111-120.

PEAKALL R, SMOUSE PE, 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update [J]. *Bioinformatics*, 28(19): 2537-2539.

REN Z, BAI Q, SU SC, et al., 2022. Genetic diversity analysis of *Pistacia chinensis* bunge based on SSR markers [J]. *Acta Botanica Boreali-Occidentalia Sinica*, 42(9): 1530-1539.

ROZEN SG, SKALETSKY H, 2000. Primer3 on the WWW for general users and for biologist programmers [J]. *Methods in Molecular Biology*, 132: 365-386.

SU YJ, WEI WY, XIE A, et al., 2023. Quality standard of *Choerospondias* cortex [J]. *China Pharmaceuticals*, 32(08): 73-76.

SUN RX, 2017. Genetic diversity and phylogeography of *Liquidambar formosana* Hance in China [D]. Beijing: Chinese Academy of Forestry.

TEMNYKH S, PARK WD, AYRES N, et al., 2000. Mapping and genome organization of microsatellite sequence in rice (*Oryza sativa* L.) [J]. *Theoretical and Applied Genetics*, 100(5): 697-712.

UNAMBA C, AKSHAY N, SHARMA RK, 2015. Next generation sequencing technologies: the doorway to the unexplored genomics of non-model plants [J]. *Frontiers in Plant Science*, 6: 1074.

WANG JW, DUAN LY, DAI GL, et al., 2022. Exploiting and primer design of SSR loci based on the transcriptome of Lycium Berry [J/OL]. *Molecular Plant Breeding*, 1-19[2024-12-13]. http://kns.cnki.net/kcms/detail/46.1068.S.20221208.1526.003.html.

WANG ZX, HERRERA F, SHU JW, et al., 2020. A new *Choerospondias* (Anacardiaceae) endocarp from the middle Miocene of southeast China and its paleoecological implications [J]. *Review of Palaeobotany and Palynology*, 283: 104312.

WU SQ, ZHANG JX, HONG XG, et al., 2001. Advance in molecular marker' s technology and its application [J]. *Chinese High Technology Letters*(4): 99-103.

WU ZL, 2023. Taxonomic identification and paleogeographic migration of the late Miocene Anacardiaceae from eastern Zhejiang and comparative study of monsoon intensity [D]. Chang' an: Chang' an University.

YANG CX, LI KQ, DING W, et al., 2018. Transcriptome analysis of *Choerospondias axillaris* based on high-throughput sequencing [J]. *Molecular Plant Breeding*, 16(6): 1798-1806.

YANG GL, ZHONG CL, ZHANG Q, et al., 2022. Optimization of extraction of total flavonoids from *Choerospondias axillaris* leaves by response surface methodology and analysis of antioxidant activity [J]. *Cereals & Oils*, 35(9): 133-137.

YAN MM, DAI XG, LI SX, et al., 2011. Sequence analysis and comparison of EST-SSRs in pine, poplar and eucalyptus [J]. *Genomics and Applied Biology*, 30(1): 103-109.

YAO YP, ZHAO L, HE WC, et al., 2021. A study on geographic variation of *Choerospondias axillaris* fruit phenotype characters and germination [J]. *Acta Agriculturae Universitatis Jiangxiensis*, 43(5): 996-1006.

YE JS, LI WH, GONG B, et al., 2015. Establishment and optimization of ISSR-PCR system for *Choerospondias axillaris* [J]. *South China Forestry Science*, 43(2): 6-9.

YE XM, CHEN FS, SUN RX, et al., 2019. Prediction of potential suitable distribution areas for *Choerospondias axillaris* based on MaxEnt model [J]. *Acta Agriculturae Universitatis Jiangxiensis*, 41(3): 440-446.

ZHANG K, ZHANG X, WANG Q, et al., 2021. The complete chloroplast genome of *Choerospondias axillaris* (Roxb.) B. L. Burtt et A. W. Hill, an ancient and versatile plant [J]. *Mitochondrial DNA Part B: Resources*, 6(8): 2400-2401.

ZHANG PZ, DREISIGACKER S, MELCHINGER AE, et al., 2005. Quantifying novel sequence variation and selective advantage in synthetic hexaploid wheats and their backcross-derived lines using SSR markers [J]. *Molecular Breeding*, 15(1): 1-10.

ZHANG Q, LU LX, ZHENG YF, et al., 2022. Isolation, purification, and antioxidant activities of polysaccharides from *Choerospondias axillaris* leaves [J]. *Molecules*, 27(24): 8881.

ZHANG SS, YAO J, YIN Q, et al., 2022. Development and identification of SSR molecular markers based on transcriptome sequencing of *Platanus acerifolia* [J]. *Molecular Plant Breeding*, 20(3): 919-925.

ZHANG SW, LIANG SM, ZHENG XL, et al., 2019. Development of genomic SSR and application in Chinese bayberry [J]. *Acta Horticulturae Sinica*, 46(1): 149-156.

ZHANG TT, ZHANG HS, SONG KJ, et al., 2023. Analysis of SSR site characteristics of *Trifolium repens* transcriptome and primer development [J]. *Pratacultural Science*, 40(9): 2266-2275.

ZHENG TT, WEI WL, YANG X, et al., 2015. Development of SSR molecular markers based on transcriptome sequencing of *Eriobotrya japonica* [J]. *Subtropical Plant Science*, 44(4): 274-278.

ZHONG XR, XU MY, LI T, et al., 2023. Development of EST-SSRs based on the transcriptome of *Castanopsis carlesii* and cross-species transfer ability in other *Castanopsis* species [J]. *PLoS ONE*, 18(7): e288999.

ZALAPA JE, CUEVAS H, ZHU HY, et al., 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences [J]. *American Journal of Botany*, 99(2): 193-208.

Note: Figure translations are in progress. See original paper for figures.