ChinaRxiv 中X

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202207.00025

EST-SSR Primer Development and Validation for *Liquidambar formosana* Based on Transcriptome Sequencing (Postprint)

Authors: Li Hui, Feng Yuanheng, Tang Shengsen, Yang Zhangqi

Date: 2022-07-05T00:00:00+00:00

Abstract

Liquidambar formosana is one of the important native tree species in Guangxi, with high timber, ornamental, and medicinal value. This study developed EST-SSR primers based on transcriptome sequencing results of L. formosana, which can provide effective and reliable molecular marker tools for population genetics research of this species and is of great significance for its conservation, development, and utilization. Based on transcriptome sequencing technology, SSR loci were detected in L. formosana and primers were designed. EST-SSR primers with high polymorphism were screened through PCR amplification and polyacrylamide gel electrophoresis, and the practical application effect of these SSR primers was verified through genetic diversity analysis of one natural population of L. formosana. The results showed that: (1) A total of 23 777 SSR loci were identified in the SSR Unigenes of L. formosana obtained from transcriptome sequencing, with mononucleotide repeat type SSR loci accounting for the highest proportion (46.54%). In terms of repeat numbers, SSR loci with 5-12 repeats accounted for the highest proportion (72.36%). (2) A total of 262 pairs of SSR primers were developed, with an effective amplification rate of 53.1%, and 18 pairs of primers with stable amplification and clear bands were finally selected. (3) Polymorphism detection results showed that all loci were polymorphic. Genetic diversity results of the natural population revealed that the ranges of number of alleles (Na), number of effective alleles (Ne), Shannon's information index (I), and observed heterozygosity (Ho) in this natural population were 2-4, 1.112 8~2.609 6, 0.208 9~1.112 7, and 0.275 9~1.000 0, respectively, with average values of 2.333 3, 1.957 4, 0.708 5, and 0.722 6, respectively. In summary, the dominant SSR locus repeat types and repeat motifs in L. formosana are basically the same as those in other species. The 18 pairs of EST-SSR primers developed in this study can meet the needs for conducting population genetics research of L. formosana and provide abundant marker primers for subsequent studies on genetic diversity of L. formosana populations.

ChinaRxiv 中X

Full Text

Development and Validity Evaluation of *Liquidambar formosana* EST-SSR Primers Based on Transcriptome Sequencing

LI Hui[12], FENG Yuanheng[2], TANG Shengsen[2], YANG Zhangqi[12]*

[1]College of Life Sciences, Guangxi Normal University, Guilin 541006, Guangxi, China

[2]Guangxi Forestry Research Institute, Engineering Research Center of Masson Pine of State Forestry Administration, Engineering Research Center of Masson Pine of Guangxi, Guangxi Key Laboratory of Superior Timber Trees Resource Cultivation, Nanning 530002, China

Abstract: *Liquidambar formosana* is an important native tree species in Guangxi with high timber, ornamental, and medicinal value. This study developed EST-SSR primers based on transcriptome sequencing results, providing an effective and reliable molecular marker tool for population genetics research on this species, which is significant for its conservation, development, and utilization. Using transcriptome sequencing technology, we detected SSR loci in *L. formosana* and designed primers. Through PCR amplification and polyacrylamide gel electrophoresis, we screened EST-SSR primers with high polymorphism and validated their practical application through genetic diversity analysis of a natural *L. formosana* population. The results showed: (1) A total of 23,777 SSR loci were identified from the transcriptome-derived SSR Unigenes, with mononucleotide repeats being the most abundant type (46.54%). In terms of repeat numbers, SSR loci with 5–12 repeats accounted for the highest proportion (72.36%). (2) A total of 262 SSR primer pairs were developed, with an effective amplification rate of 53.1%, and 18 pairs of primers with stable amplification and clear bands were finally selected. (3) Polymorphism detection results indicated that all loci were polymorphic. Genetic diversity analysis of the natural population revealed that the number of alleles (Na), effective number of alleles (Ne), Shannon's information index (I), and observed heterozygosity (Ho) ranged from 2–4, 1.1128–2.6096, 0.2089–1.1127, and 0.2759–1.0000, respectively, with average values of 2.3333, 1.9574, 0.7085, and 0.7226. In conclusion, the dominant SSR repeat types and motifs in *L. formosana* are basically the same as those in other species. The 18 pairs of EST-SSR primers developed in this study can meet the needs of population genetics research on *L. formosana* and provide abundant marker primers for subsequent studies on genetic diversity of *L. formosana* populations.

Keywords: *Liquidambar formosana*, transcriptome, SSR molecular markers, primer development, genetic diversity

---

Liquidambar formosana (Hamamelidaceae) is a colorful deciduous tree species

mainly distributed south of the Qinling-Huaihe line, with presence throughout Guangxi. Its leaves are green in spring and summer, turning deep red in autumn, giving it high ornamental value. As a traditional medicinal plant in China, *L. formosana* bark can be used to promote blood circulation and unblock collaterals, its fruits are effective for indigestion, and its leaves have heat-clearing and detoxifying effects. The leaves are also an indispensable material for making five-colored rice during the Zhuang ethnic minority's "March 3rd" traditional festival in Guangxi. The species also has high timber value for furniture and construction materials, and plays an important role in maintaining soil fertility. With strong adaptability and tolerance to poor soil conditions, *L. formosana* shows strong resistance to toxic gases such as Cl_2 and SO_2, making it an important species for ecological and forest structure adjustment, thus holding significant research value.

In recent years, transcriptome sequencing (RNA-seq) technology has been widely applied in molecular biology research and has become one of the most commonly used techniques, enabling rapid and efficient acquisition of gene sequences and revealing gene function and structure at the whole-genome level during specific developmental stages. Mining and analyzing transcriptome sequencing data can reveal SSR locus distribution characteristics, providing a reference for subsequent SSR primer screening and design. Simple sequence repeats (SSR), also known as microsatellite DNA, refer to repetitive sequences composed of 1–6 nucleotides as repeat units tandemly arranged into dozens of nucleotides. SSRs are present throughout the genome of eukaryotes and are widely used in genetic diversity analysis, genetic linkage map construction, gene mapping, and molecular marker-assisted breeding due to their abundance, high polymorphism, good repeatability, and strong specificity. They are considered one of the most effective markers for detecting population genetic diversity and differentiation. Based on their origin, SSRs can be divided into genomic SSR (G-SSR) and expressed sequence tag (EST)-based SSR. G-SSR markers are based on genomic sequences with complex development processes, high costs, and low efficiency. EST-SSR markers, based on expressed sequence tags, not only share the advantages of G-SSR markers but also exhibit relatively high sequence conservation and good transferability among different species within the same plant genus.

As an important native tree species in China, research on *L. formosana* has mainly focused on physiological property analysis, plantation cultivation, and new variety certification. Few studies have addressed population genetics of *L. formosana*, primarily due to insufficient molecular markers. Previous genetic diversity studies on *L. formosana* mainly used dominant molecular markers. For example, Huang et al. (2021) explored optimal DNA extraction methods from *L. formosana* leaves and optimized ISSR primers and ISSR-PCR amplification systems. Bi et al. (2010) analyzed genetic diversity in natural *L. formosana* populations using ISSR molecular markers. Li (2015) optimized DNA extraction and SRAP-PCR reaction systems. However, ISSR markers have poorer experimental repeatability than SSR markers and cannot detect multi-

allelic loci, distinguish heterozygotes, or perform genotyping as SSR markers can. SRAP markers produce too many amplification bands, making fingerprint identification more difficult than with SSR markers when constructing variety fingerprints using multiple primer combinations. SSR molecular markers can reveal relatively higher genetic information and are significantly superior to other traditional molecular markers, yet current reports on *L. formosana* SSR primer development are limited. Previous studies developed *L. formosana* SSR primers mainly through magnetic bead enrichment methods and by retrieving data from NCBI. These methods suffer from limited locus information, difficult primer development, and small numbers of developed primers, resulting in insufficient SSR primers to support more in-depth population genetics research on *L. formosana*.

This study used seven counties in Hechi City, Guangxi, and Libo County in Guizhou as the research area. Based on transcriptome sequencing technology, we obtained rich and comprehensive SSR locus information for *L. formosana*, screened and designed EST-SSR primers, and selected polymorphic and stably amplifying EST-SSR primers through PCR amplification and polyacrylamide gel electrophoresis. We validated primer application effectiveness through genetic diversity analysis of a wild population. The study aimed to address: (1) the specific distribution and main sequence characteristics of SSR loci in *L. formosana*; (2) the development efficiency of *L. formosana* SSR primers and factors affecting development efficiency; and (3) the polymorphism of developed *L. formosana* SSR primers and their application effectiveness in genetic diversity analysis. This research provides a foundation for molecular marker-assisted breeding and functional gene tagging of *L. formosana*, offers reliable research tools for population genetics and genetic breeding studies, and provides references for genetic diversity evaluation of *L. formosana* germplasm resources and molecular plant-assisted breeding in Guangxi.

1.1 Study Materials

A natural *L. formosana* population was randomly selected in Huanjiang County, Hechi City, Guangxi. Fresh red, yellow, and green leaves were collected in 50 mL centrifuge tubes, frozen in liquid nitrogen, and sent to Beijing Novogene Bioinformatics Technology Co., Ltd. for transcriptome sequencing.

After literature review and field investigation, one natural *L. formosana* population was selected from each of seven counties in Hechi City, Guangxi, and Libo County, Guizhou. One sample (A1–A8) was randomly collected from each population for preliminary primer screening. Additionally, 2–4 *L. formosana* samples were collected from each of the eight natural populations (with >50 m distance between sampled individuals), totaling 24 samples (B1–B24) for primer rescreening and polymorphism detection, ensuring coverage of more natural populations and individuals in the test area. Furthermore, 30 samples (C1–C30) were collected from another natural *L. formosana* population in Huanjiang County, Hechi City, for genetic diversity analysis. Detailed material sources are shown

in Table 1 .

1.2 *Liquidambar formosana* Transcriptome Sequencing

Quality control of the sequencing process was implemented as follows: (1) Sample quality testing included agarose gel electrophoresis to analyze RNA degradation and contamination, Nanodrop to detect RNA purity ($OD_{260}/_{280}$ ratio), Qubit for precise RNA concentration quantification, and Agilent 2100 to accurately detect RNA integrity, ensuring RNA integrity and quantity. (2) Full-length transcriptome libraries were constructed using Oligo(dT) magnetic bead enrichment. During PCR amplification and enrichment of synthesized cDNA, optimal PCR conditions were determined through cycle optimization to ensure library construction quality. After passing quality inspection, sequencing was performed on the PacBio Sequel platform according to the effective concentration of the library and data output requirements, ensuring the final SSR locus information was valid and reliable.

1.3 *Liquidambar formosana* EST-SSR Marker Development

SSR loci detected in the *L. formosana* transcriptome were screened according to the following criteria: sequence length range of 16–28 bp, repeat numbers >9 for dinucleotide units and $ $5 for tri- to hexanucleotide units (mononucleotide units excluded), and a maximum 3 °C difference in annealing temperature between forward and reverse primers. SSR primers were designed and synthesized by Guangzhou Aiji Biotechnology Co., Ltd.

1.4 *Liquidambar formosana* Genomic DNA Extraction

Genomic DNA was extracted from *L. formosana* using a rapid genomic DNA extraction system (Wang et al., 2013). Extracted DNA was assessed for purity and concentration using a spectrophotometer, with OD_{260}/OD_{280} values of 1.7–1.9. DNA quality was verified by 1% agarose electrophoresis. Qualified template DNA was diluted to 50 ng· L^{-1} and stored at -20 °C.

1.5 PCR Amplification

The total PCR reaction volume was 10 L, containing 2 L DNA (50 ng· L^{-1}), 1 L PCR Buffer (10×), 0.2 L dNTPs, 0.25 L each of forward and reverse primers, 0.07 L Taq DNA Polymerase (5 U· L^{-1}), and 6.23 L ddH_2O. Amplification was performed in a thermal cycler with the following program: initial denaturation at 94 °C for 4 min, followed by 30 cycles of denaturation at 94 °C for 15 s, annealing at 55 °C for 15 s, extension at 72 °C for 30 s, and a final extension at 72 °C for 20 min, with storage at 12 °C.

1.6 Primer Effectiveness and Polymorphism Detection

PCR products were separated by 8% polyacrylamide gel electrophoresis. The procedure was as follows: 1.2 L of amplified product with bromophenol blue indicator was loaded onto the polyacrylamide gel using a pipette, with 1 L of 50–500 bp marker loaded at both ends. Electrophoresis was performed at a constant voltage of 240 V, with electrophoresis time adjusted according to the bromophenol blue indicator. After electrophoresis, the gel was washed twice with ddH_2O, shaken in fixative solution for 10 min, then washed 2–3 times in ultrapure water for 2 min each. The gel was silver-stained for 7 min, washed 2–3 times in distilled water for 2 min each, then developed with shaking until clear bands appeared. The gel was washed twice with ddH_2O before band reading and photography.

1.7 Data Statistics

Bands were manually scored. Primer effectiveness was determined by whether effective bands were amplified. For effective primers, different amplification products were numbered A, B, C… in descending order of band length. Popgene 32 software was used to calculate: (1) observed number of alleles (Na); (2) effective number of alleles (Ne) (Hartl et al., 1989); (3) Shannon's information index (I) (Shannon et al., 1949); (4) observed heterozygosity (Ho); and (5) expected heterozygosity (He) (Nei et al., 1973). PIC-CALC software was used to calculate polymorphism information content (PIC) values (Botstein et al., 1980). PIC is an important indicator for judging SSR primer polymorphism: $PIC \leq 0.25$ indicates low polymorphism, $0.25 < PIC < 0.5$ indicates moderate polymorphism, and $PIC \geq 0.5$ indicates high polymorphism. The PIC calculation formula (Yang et al., 2004) is:

$$PIC = 1 - \sum_{i=1}^{n} P_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2P_i^2 P_j^2$$

where PIC is polymorphism information content, P_i and P_j represent the frequencies of the ith and jth alleles, respectively, and n is the number of alleles.

Primer effective amplification rate refers to the ratio of primers that can effectively amplify to the total number of primers. Primer polymorphism ratio refers to the ratio of primers that can effectively amplify and show polymorphism to the total number of primers.

2.1 Distribution Characteristics of SSR Loci in *L. formosana* Transcriptome

A total of 23,777 SSR loci were mined from the *L. formosana* transcriptome, with an average of one SSR locus per 3.13 kb. Statistical analysis revealed that SSRs in the *L. formosana* transcriptome mainly consisted of six simple repeat

unit types from mononucleotide to hexanucleotide, plus a few complex repeat types (Table 2). Mononucleotide repeats accounted for 46.54% of total SSR loci, followed by dinucleotide repeats (33.10%). Tri-, tetra-, penta-, and hexanucleotide repeats accounted for 17.80%, 1.17%, 0.49%, and 0.88%, respectively. The most frequent mononucleotide type was T/A (66.81%), followed by A/T (24.99%). Among dinucleotide types, CT/AG (24.7%) was most frequent, followed by TC/GA (23.72%). For trinucleotide types, CAG/CTG (6.17%) was most common, followed by GAA/TTC (5.08%). The most frequent tetranucleotide was TTTA/TAAA (17.2%). For pentanucleotides, CTTTT/AAAAG (8.62%) showed the highest frequency. The most common hexanucleotide was ACCAGC/GCTGGT (4.72%). In summary, *L. formosana* SSR loci were dominated by mononucleotide and dinucleotide repeats, accounting for 79.64% of total SSR loci, while tri- to hexanucleotide repeat types suitable for primer design accounted for only 19.46%, limiting the number of SSR primers that could be developed.

SSR repeat numbers ranged from 5–78 (Table 3), mainly concentrated between 5–24 repeats. The most abundant repeat number category was 5–8 repeats (8,787 loci, 36.96% of total), followed by 9–12 repeats (8,417 loci, 35.40%). SSR loci with 13–16, 17–20, and 21–24 repeats accounted for 3,934 (16.55%), 1,830 (7.70%), and 598 (2.52%) loci, respectively. Those with >24 repeats were the least abundant, with only 211 loci (0.89%). Among all repeat numbers, 10 repeats showed the largest proportion (3,570 loci, 15.01%), followed by 6 repeats (3,220 loci, 13.54%). SSR polymorphism arises from replication slippage, and higher repeat numbers generally correlate with higher polymorphism. The dominant repeat number of 5–8 in *L. formosana* SSRs basically meets the requirements for SSR primer design and development.

2.2 Validation of *L. formosana* EST-SSR Primer Effectiveness

As shown in Table 4 , 262 primer pairs were initially screened using DNA from eight *L. formosana* samples (A1–A8). A total of 139 primer pairs amplified effectively, yielding an effective amplification rate of 53.1%. The effective amplification rates for different nucleotide repeat types were basically the same, suggesting that the designed and synthesized primers of various nucleotide repeat types have certain applicability in *L. formosana* populations (Yang et al., 2021).

2.3 Polymorphism Analysis of *L. formosana* EST-SSR Primers

Among the 139 effective primer pairs, 41 amplified polymorphic bands, with a polymorphism ratio of 15.64%, indicating relatively strong conservation of *L. formosana* SSR sequences. Eighteen primer pairs with stable amplification, clear bands, and polymorphism (Table 5) were selected for polymorphism detection using DNA from 24 samples (B1–B24) and analyzed using PopGene32 software.

Polymorphism analysis revealed that the observed number of alleles (Na) at

18 loci ranged from 2–5, with an average of 3. PIC values ranged from 0.30–0.72, with 10 loci showing PIC > 0.5 (high polymorphism) and the remaining 8 loci showing PIC > 0.25 (moderate polymorphism). No low-polymorphism loci were detected. The 18 selected *L. formosana* SSR primer pairs exhibited high polymorphism.

2.4 Validation of EST-SSR Genetic Analysis Effectiveness in a Natural *L. formosana* Population

The 18 primer pairs were used for genetic diversity analysis of 30 samples (C1–C30) from a natural *L. formosana* population. Amplification results (Table 6) showed that 42 alleles were amplified across 18 loci. The observed number of alleles (Na) ranged from 2–4, averaging 2.3333. The effective number of alleles (Ne) ranged from 1.1128–2.6096, averaging 1.9574. Shannon's index (I) ranged from 0.2089–1.1127, averaging 0.7085. Observed heterozygosity (Ho) ranged from 0.2759–1.0000, averaging 0.7226. Expected heterozygosity (He) ranged from 0.3706–0.8968, averaging 0.5228. These results demonstrate that the 18 developed SSR primer pairs can be used for genetic diversity evaluation of *L. formosana* populations.

3.1 Distribution Characteristics of SSR Loci in *L. formosana* Transcriptome

SSR markers have the advantage of transferability among plant species, and many studies have transferred existing markers within genera. However, few SSR primers are currently available for *L. formosana*, with only 10 pairs developed by Gu (2016) using magnetic bead enrichment and 11 pairs developed by Sun (2017) from NCBI database information. The magnetic bead enrichment method mainly involves constructing microsatellite-enriched libraries, followed by positive clone screening and sequencing to obtain SSR locus information. Although this method improves SSR primer development efficiency to some extent, the procedure is cumbersome, requires genomic library construction and screening, has high development difficulty, and yields limited SSR locus information. The NCBI database method allows researchers to obtain useful information and saves time and experimental costs, but current *L. formosana* sequence data in databases are limited, resulting in scarce SSR locus information and inability to describe distribution characteristics in detail. With the development of high-throughput sequencing technology and decreasing RNA-seq costs, EST-SSR marker development based on transcriptome data has become an important technical approach for gene discovery and molecular marker development. This study is the first to obtain relatively comprehensive SSR locus information for *L. formosana* through transcriptome sequencing. We found that as a Tertiary relict plant, *L. formosana* has rich SSR motif repeat types, including all six nucleotide repeat types. The high proportion of mononucleotide repeats (46.54%) is an important distribution characteristic of *L. formosana* EST-SSRs, with repeat motif distribution patterns similar to those of *Cornus florida* (Liu et al.,

2020) and *Keteleeria calcarea* (Shi et al., 2021). Overall, the number and types of SSR loci in the *L. formosana* transcriptome are relatively abundant, providing an important foundation for SSR primer development and subsequent research.

3.2 Development Efficiency of *L. formosana* EST-SSR Loci

Among the 262 EST-SSR primer pairs designed in this study, the effective amplification rate was 53.1%. Previous studies by Qi et al. (2004) and Wei et al. (2008) indicated that effective amplification rates for EST-SSR primers should be 60–90%. The relatively low effective amplification rate in this study may be attributed to three aspects: (1) Tri- to hexanucleotide repeat types accounted for a low proportion (19.46%), resulting in a limited total number of SSR loci available for screening; (2) Some primers were poorly designed, as all primers in this study were designed from second-generation sequencing results, which may contain assembly errors, leading to relatively low effective amplification rates; (3) The polymorphism ratio of developed *L. formosana* SSR primers was low (15.64%), mainly because potential SSR polymorphism increases with repeat number (Liu et al., 2020), while the dominant repeat numbers in *L. formosana* were 5–12 with short repeat types, limiting potential polymorphism. The underlying reason is that transcriptome SSR loci originate from relatively conserved gene coding regions, and polymorphic sites are generally located in microsatellite repeat fragments of functional sequences, resulting in lower polymorphism of transcriptome SSRs (Yang et al., 2018).

3.3 Polymorphism of *L. formosana* EST-SSR Primers and Validation of Population Genetic Analysis

Parameters such as polymorphism information content (PIC), observed number of alleles (Na), and effective number of alleles (Ne) can be used to evaluate SSR primer usability and polymorphism richness. Regarding PIC index, Sun (2017) developed 14 SSR primer pairs for *L. formosana*, including 6 highly polymorphic loci, 4 moderately polymorphic loci, and 4 low-polymorphism loci. Compared with previous studies, the 18 SSR primer pairs developed in this experiment included 10 highly polymorphic loci and 8 moderately polymorphic loci, showing higher PIC values. Regarding allele number (Na), Sun (2017) developed 14 SSR primer pairs in a nationwide study on genetic diversity and phylogeography of Chinese *L. formosana*, with average Na of 3.2580 and average Ne of 1.8480. Gu (2016) developed 10 SSR primer pairs in a study on the impact of fragmented habitats on genetic diversity in the Thousand Island Lake region, with mean Na of 4.6 and mean Ne of 2.1300. The 18 SSR primer pairs developed in this study detected average Na of 2.3333 in the Huanjiang population, lower than previous studies because Na largely depends on sample size, and our sampling range and number were smaller than previous studies. The average Ne (1.9574) was similar to previous studies, and Ne better measures increases in population genetic variation than heterozygote and homozygote frequencies, indicating rich

genetic diversity in the Huanjiang population. This discussion demonstrates that the SSR primers developed in this study can effectively reflect genetic variation in *L. formosana* populations and are suitable for genetic diversity research.

The purpose of primer development is to obtain SSR molecular markers with high PIC. Previous studies using conventional SSR marker development technology obtained limited *L. formosana* sequences, resulting in small numbers of developed primers with low PIC. This study is the first to comprehensively explore *L. formosana* SSR locus information through transcriptome sequencing technology, which can efficiently obtain complete transcriptome information of cells at a certain developmental stage and overcome problems of cumbersome procedures and high time and labor costs. We independently designed *L. formosana* EST-SSR primers, validated their effectiveness and polymorphism, and obtained 18 newly developed polymorphic EST-SSR primer pairs with relatively high PIC. Validation results showed that these 18 SSR primer pairs can be used for subsequent genetic diversity analysis and germplasm resource conservation research of *L. formosana*, expanding the number and types of current *L. formosana* SSR primers. Population validation results showed that different EST-SSR molecular markers exhibit certain differences in population expression, indicating that evolutionary rates differ among genes at different loci, with low-polymorphism SSR loci often located in conserved genes (Liu et al., 2020). In addition to genetic diversity analysis, *L. formosana* EST-SSR primers developed through transcriptome sequencing can be used to trace gene fixation or evolution at their loci to explore population genetic differentiation issues. Combined with transcriptome information, functional studies of sequences at these loci can be conducted to construct specific SSR markers associated with superior phenotypes, providing assistance for subsequent research on molecular mechanisms of leaf color change and synthesis of medicinal active components in *L. formosana*. This has important guiding significance for the development, preservation, and utilization of *L. formosana* germplasm resources.

In summary, this study obtained comprehensive SSR locus information for *L. formosana* based on transcriptome sequencing technology, successfully developed 18 EST-SSR primer pairs, provided necessary supplementation to the number of medium- and high-polymorphism SSR primers for *L. formosana*, helped address the current problem of insufficient numbers and types of high-polymorphism SSR primates to support more in-depth genetic diversity research, and provided more and better research tools for population genetics and molecular-assisted breeding of *L. formosana*.

References

Bi ZX, Jin ZX, Li JM, et al., 2010. Genetic diversity in the natural populations of *Liquidambar formosana* revealed by ISSR molecular markers [J]. *Bulletin of Botanical Research*, 30(1): 120-125.

Botstein D, White RL, Skolnick M, et al., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms [J]. *American Journal of Human Genetics*, 32(3): 314-331.

Fan SB, Su JY, Fang T, et al., 2021. Pathogen of *Liquidambar formosana* leaf spot [J]. *Mycosystema*. https://kns.cnki.net/kcms/detail/11.5180.Q.20211228.1418.002.html

Gu XY, 2016. Impacts of habitat fragmentation on genetic diversity and differentiation of *Liquidambar formosana* populations [D]. Shanghai: East China Normal University: 6-7.

Hartl DL, Clark AG, 1989. *Principles of Population Genetics* [M]. 2nd ed. Sunderland: Sinauer Associates, Inc Publishers, 1990(1): 128-129.

Huang M, Huang XP, Li BQ, et al., 2021. Optimization of *Liquidambar formosana* ISSR-PCR reaction system and primer [J]. *Molecular Plant Breeding*. https://kns.cnki.net/kcms/detail/46.1068.s.20210429.1428.010.html

Lai JX, 2020. Comparative analysis reveals different mechanisms of leaf color changes in *Liquidambar formosana* Hance during different seasons and among different varieties [D]. Beijing: Chinese Academy of Forestry: 12-13.

Li FF, 2015. Analysis on the genetic diversity of *Liquidambar formosana* of Zhejiang province [D]. Zhengzhou: Henan Agricultural University: 18-21.

Liu FF, Xu GB, Li TT, et al., 2021. Characterization and development of SSR markers from transcriptome sequence of *Bretschneidera sinensis* Hemsl [J]. *Plant Science Journal*, 39(3): 288-296.

Liu JQ, Fang Y, Chen YX, 2020. Analysis of SSR distribution characteristics and primer development of *Cornus florida* transcriptome [J]. *Molecular Plant Breeding*, 18(20): 6769-6775.

Liu Z, Zhao Y, Yang PD, et al., 2014. Comparison of parent identification for tea variety based on SSR, SRAP and ISSR markers [J]. *Journal of Tea Science*, 34(6): 617-624.

Nei M, 1973. Analysis of gene diversity in subdivided populations [J]. *Proceedings of the National Academy of Sciences USA*, 70(12): 3321-3323.

Shannon CE, Weaver RW, Wiener RN, 1949. The mathematical theory of communication [J]. *Urbana University of Illinois Press*, 85(2): 24-30.

Sun RX, 2017. Genetic diversity and phylogeography of *Liquidambar formosana* Hance in China [D]. Beijing: Chinese Academy of Forestry: 26-27.

Wang DD, Yang DX, 2017. Development and transferability of EST-SSR primers in *Actinidia arguta* [J]. *Journal of Northwest Forestry University*, 32(4): 147-152.

Wang JJ, Zhang B, Zhang WS, 2015. A new variety, *Liquidambar formosana* 'Jinjue' [J]. *Scientia Silvae Sinicae*, 51(10): 154.

Wang X, Liao HH, Huang HM, et al., 2013. Comparative test for DNA extraction methods of Citrus Huanglongbing pathogen [J]. *Journal of Southern Agriculture*, 44(2): 225-229.

Wei LD, Liang LH, Tang SS, et al., 2016. Plantation growth process of *Liquidambar formosana* [J]. *Guangxi Forestry Science*, 45(4): 409-413.

Wei LB, Zhang HY, Zheng YZ, et al., 2008. Development and utilization of EST-derived microsatellites in sesame (*Sesamum indicum* L.) [J]. *Acta Agronomica Sinica*, 34(12): 2077-2084.

Xin Y, Cui HR, Zhang ML, et al., 2005. Development of EST (expressed sequence tags) marker in Chinese cabbage and its transferability to rapeseed [J]. *Hereditas*, 27(3): 410-416.

Xu ZJ, Zhao S, Xu L, et al., 2020. Discovery of microsatellite markers from RNA-seq data in cultivated peanut (*Arachis hypogaea*) [J]. *Scientia Agricultura Sinica*, 53(4): 695-706.

Yang GZ, Zhang JB, Zhao YM, et al., 2004. Study on polymorphism of 8 microsatellite markers in grassland red cattle crossed population [J]. *Grass-Feeding Livestock*, 124(3): 23-27.

Yang SP, Sun XM, Wang LH, et al., 2018. Development and identification of SSR markers based on transcriptome of *Helianthus tuberosus* L. [J]. *Molecular Plant Breeding*, 16(2): 484-492.

Yang X, Yang N, Yuan QH, et al., 2021. Development and application of EST-SSR molecular markers in *Pinus bungeana* [J]. *Journal of Beijing Forestry University*, 43(7): 1-11.

Ye P, Li XH, Tang JR, et al., 2019. Distribution and characteristics of SSR in transcriptome of *Camellia fascicularis* [J]. *Journal of Central South University of Forestry and Technology*, 39(9): 86-91.

Yi LB, Li F, Cui ZX, 2017. Application of high-throughput sequencing technology in the research of ramie (*Boehmeria nivea* L. Gaud) biology [J]. *Plant Fiber Sciences in China*, 39(2): 88-93.

Yin GP, Liu XS, Jiang Y, et al., 2021. The changes of tissue structure, photosynthetic characteristics and pigment content of *Liquidambar formosana* leaves in the process of turning red [J/OL]. *Guihaia*. https://kns.cnki.net/kcms/detail/45.1134.Q.20210722.1715.004.htm

Zhang LD, Tang KX, 2010. Development of plant EST-SSR markers and its application [J]. *Genomics and Applied Biology*, 29(3): 534-541.

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

Source: ChinaXiv — Machine translation. Verify with original.