

Article

Comparison of Parentage Verification and Parentage Assignment Methods for Walnut (*Juglans regia* L.) Controlled-Cross Offspring Using SSR Markers

Shengqun Chen ^{1,2,†}, Tao Zhou ^{3,†}, Yajun Zeng ^{1,2,4,*} and Na Hou ^{1,2,*}

¹ Guizhou Institute of Walnut, Guizhou Academy of Forestry, Guiyang 550005, China; squnchen@163.com

² Guizhou Academy of Forestry, Guiyang 550005, China

³ Honghe Tropical Agriculture Institute of Yunnan, Honghe 661100, China; zhoutao960712@163.com

⁴ Guizhou Key Laboratory of Forest Cultivation in Plateau Mountain, Guiyang 550005, China

* Correspondence: yjz410@163.com (Y.Z.); houna1018@163.com (N.H.)

† These authors contributed equally to this work.

Abstract

To evaluate the effectiveness of microsatellite (SSR) markers for verifying the authenticity of walnut controlled-cross offspring, seven controlled-cross combinations were established using improved and widely cultivated walnut varieties from Guizhou Province, China, generating a total of 108 candidate offspring. Among 60 previously reported SSR primer pairs, 15 loci exhibiting stable amplification and clear genotyping profiles were selected for analysis. Parentage identification was conducted using the genotype exclusion and maximum likelihood methods, and the applicability of each method was assessed through genetic diversity and cluster analyses. The 15 SSR loci detected a total of 68 alleles, with an average of 4.53 alleles per locus. The polymorphism information content (PIC) ranged from 0.017 to 0.608, with a mean value of 0.396. The genotype exclusion method confirmed 76 offspring as genetically consistent with their recorded parental combination, corresponding to an overall identification rate of 70.37%. In the maximum likelihood analysis, 27 offspring were assigned to their most likely parent pairs when both parents were unknown, with an identification rate of 25.00%. When the maternal parent was known and only the paternal parent was unknown, 47 offspring were successfully assigned, increasing the identification rate to 43.52%. Cluster analysis showed that most offspring clustered near their corresponding parents, although some individuals deviated from their original cross combinations, suggesting allele sharing or complex genetic origins among the parental materials. The shared maternal parent structure, closely related paternal backgrounds, high-frequency alleles, and null alleles may jointly influence the efficiency of SSR-based parentage assignment. Compared with the maximum likelihood method, the genotype exclusion method showed greater practical stability for verifying recorded cross combinations in a structured walnut breeding population characterized by a shared maternal parent, multiple closely related paternal parents, and moderate SSR polymorphism. These findings indicate that likelihood-based parentage assignment should be interpreted cautiously when candidate parents are genetically similar and marker informativeness is limited. This study provides a technical reference for authenticity verification of walnut controlled-cross offspring, pedigree management, and marker-assisted breeding.

Keywords: *Juglans regia* L.; SSR markers; parentage identification; genotype exclusion method; maximum likelihood method



Academic Editor: Evangelia V. Avramidou

Received: 8 June 2026

Revised: 16 July 2026

Accepted: 20 July 2026

Published: 21 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Walnut (*Juglans regia* L.), the most widely cultivated tree nut species worldwide, has substantial value for both nut production and timber utilization and occupies an important position in the forestry industry of temperate regions and China [1,2]. Hybrid breeding is a major approach for creating new germplasm and developing improved varieties in the genetic improvement of *J. regia* [1,3]. In hybrid breeding, the authenticity of offspring, accuracy of parental records, and traceability of pedigrees directly affect the reliability of cross-combination evaluation, the efficiency of superior individual selection, and the long-term utilization of breeding materials [4,5]. However, artificial pollination errors, pollen contamination, label management omissions, and material mixing make it difficult to accurately verify offspring origin using only field pollination records, morphological traits, or breeding experience [6,7]. Previous studies have provided a technical foundation for early authenticity identification and molecular fingerprinting of walnut and its controlled-cross offspring [4,6]. Nevertheless, in practical breeding populations, stable and operational methods for verifying controlled-cross offspring authenticity and confirming parentage relationships remain a key issue in breeding implementation.

SSR markers are characterized by codominant inheritance, high polymorphism, good reproducibility, rich locus-specific information, and relatively controllable detection costs. They have been widely used in plant parentage identification, population genetic analysis, and molecular fingerprinting studies [8,9]. In fruit trees and other woody plants, SSR markers have shown good resolution and operability in germplasm identification, cultivar fingerprint construction, hybrid authenticity verification, and genetic relationship analysis [2,6,10,11]. In walnut, SSR markers have been successfully applied to major cultivar identification, germplasm resource fingerprinting, genetic diversity evaluation, and molecular identification of specific hybrid materials or F₁ offspring, confirming their potential for parentage analysis in this species [2,6,12]. However, most existing walnut studies have focused on cultivar discrimination, germplasm resource evaluation, and authenticity verification of single cross combinations. Systematic investigations remain limited regarding the applicability and robustness of different parentage assignment strategies in breeding systems characterized by a shared maternal parent and multiple paternal parents. In practical breeding, shared maternal parent structure, the number of candidate paternal parents, and marker informativeness may all influence parentage assignment outcomes, yet these effects have not been fully validated [6,13,14]. Although SSR markers can effectively identify walnut controlled-cross offspring, their identification efficiency and result robustness may vary markedly under different candidate parent structures, locus combinations, and analytical methods [6,15,16]. Under conditions involving a shared maternal parent, concurrent multiple crosses, and a relatively restricted set of candidate parents, the genotype exclusion method and likelihood-based parentage assignment methods may differ in identification efficiency, robustness, and dependence on marker informativeness [6,15,16]. In addition, uneven locus polymorphism, skewed allele frequency distributions, and null alleles can further affect the reliability of parentage assignment results for controlled-cross offspring [17]. Therefore, it is necessary to systematically evaluate the performance of different SSR-based parentage identification strategies in well-characterized controlled pollination populations of walnut, thereby clarifying their applicability and limitations in practical breeding materials [6,17]. Unlike previous studies on walnuts, which primarily focused on the validation of single-combination pedigrees or variety identification, this study—characterised by the use of a common maternal line and multiple paternal lines—systematically compared the genotype exclusion method with likelihood-based methods and assessed the robustness of both approaches in practical breeding.

In this study, improved and widely cultivated walnut varieties from Guizhou Province were used to establish seven controlled cross combinations through controlled pollination, generating 108 candidate offspring. From previously reported walnut SSR markers, 15 loci with stable amplification, clear genotyping profiles, and moderate polymorphism were selected. Parentage identification was conducted using both the genotype exclusion and maximum likelihood methods, and the applicability of the two methods was evaluated through genetic diversity and cluster analyses. This study aimed to clarify the effectiveness of SSR markers for authenticity verification of controlled-cross offspring in walnut, compare the stability of different parentage assignment strategies in a breeding system characterized by a shared maternal parent and multiple paternal parents, and provide methodological references for pedigree verification, material management, and subsequent breeding efforts in walnut hybrid breeding.

2. Materials and Methods

2.1. Plant Materials

This study analyzed 108 candidate offspring from seven controlled cross combinations. The parental materials comprised improved and widely cultivated walnut varieties from Guizhou Province. Across all combinations, ‘Xue Ninghong’ (XNH) served as the maternal parent. The paternal parents included ‘Long Yuanhong’ (LYH), ‘Long Wuren’ (LWR), ‘Long Bofeng’ (LBF), ‘Qian Guiren’ (QGR), ‘Qian Panhe No. 4’ (QPH4), and ‘Qian Panhe No. 3’ (QPH3), as well as ‘Xue Ninghong’ (XNH) itself in the self-pollination combination. The seven hybrid combinations were: XNH × LYH (12 offspring), XNH × LWR (13 offspring), XNH × LBF (12 offspring), XNH × QGR (6 offspring), XNH × QPH4 (13 offspring), XNH × QPH3 (13 offspring), and XNH × XNH (39 offspring) (Table 1).

Table 1. Crossing combinations and numbers of putative controlled-cross offspring.

No.	Female ♀	Male ♂	No. of Offspring
1	XNH	LYH	12
2	XNH	LWR	13
3	XNH	LBF	12
4	XNH	QGR	6
5	XNH	QPH4	13
6	XNH	QPH3	13
7	XNH	XNH	39

2.2. DNA Extraction and SSR Genotyping

2.2.1. DNA Extraction

Genomic DNA was extracted from fresh walnut leaves using the cetyltrimethylammonium bromide (CTAB) method, following a previously published protocol with minor modifications [18]. DNA integrity was evaluated by electrophoresis on a 1% agarose gel, while DNA concentration and purity were measured using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). The extracted DNA was diluted to a working concentration of 20–50 ng·μL^{−1} for subsequent SSR amplification and genotyping [4,18].

2.2.2. SSR Primer Screening and PCR Amplification

A preliminary screening was conducted using 60 previously reported walnut SSR primer pairs [4]. Seven parental materials and eight representative controlled-cross offspring were selected as pre-screening materials to evaluate amplification performance and polymorphism. The screening criteria were as follows: (1) clear, stable, and reproducible

amplification bands, with fragment sizes generally consistent with previously reported ranges and without nonspecific amplification; (2) complete and clear amplification peak profiles suitable for capillary electrophoresis genotyping; (3) high polymorphism among different materials, with priority given to loci showing stable amplification, clear genotyping profiles, and high polymorphism, including some loci with a PIC greater than 0.5; this threshold was adopted to ensure that the selected markers provided sufficient discriminatory power for parentage analysis in a structured breeding population, where allele sharing among related parental materials may reduce marker informativeness; and (4) preferential selection of loci distributed across different genomic regions without significant linkage, to enhance genetic informativeness and the resolution of subsequent analyses [4,13]. In addition, the primers were required to show good discriminatory power across different hybrid combinations, enabling effective detection of genotypic differences between parents and offspring [4,19]. Based on these criteria, 15 SSR primer pairs with stable amplification, strong discriminatory power, and relatively high polymorphism were selected for subsequent parentage identification and genetic analysis (Table 2).

Table 2. SSR loci and primer sequences used in this study.

Locus	Forward Primer (5'-3')	Reverse Primer (5'-3')
JC8125	AGCAACCAGAGCAGAGCATT	AACCTCAACACCAACTATGCT
JH89978	ACCTTCCCTGCTCCTCTCTT	GAGCCTTGTGGAAGCAAACG
JC7329	TGCAGCGCATCAGTGAGTTA	ACGCTCGAGTGTAGTAGCAAG
JC5411	AAGCTGTTTGTGCCAAAAGC	TTCTAGCGAGAATTCCGGCC
JR4964	CTCGATCTGAACTCGGCTCC	TCTACTCTCTCCGCACCACA
JR4616	AGCCCTTTTGCATCGGCTAT	AGCTGACCGATCGATCAACA
JH42753	CAGTTTTGGCCAGCTGCAAT	TGTGCCCATGCTAAGACTGG
JH91908	GAAAAGCATGGTCCTGCTGC	ATTGAGCGACGAAAAGGGGT
JR3773	GGTGGTTTGACCCTTAATTCTGT	ACCCTGCCACAATGACCAAA
JH84548	TCTGAGGAAGCTGCATGGAA	AACTCTGGACACATGCCGC
JM68820	TCCTTCTGTGTGAGTGCGTG	GGTCAGGTGAGTGGAGCAAA
JR6439	TCGATGCGATCATCTCCGTG	CGGCACCAAAACAGAACTCG
JR3434	CCGCCCAGCAGATTGTCATA	CGTCCCCTCAAGTTCTTGCT
JH2096	AAGCTATGTTGGCTGCTGGT	ATTGTTTACGCGGTTGCCCTA
JH6044	CCTCGTCTCCTCCCTAACA	GTAGGATAGTGTGGCGTCGG

PCR amplification was performed in a total reaction volume of 20 μ L, containing 2 $\times$ Taq PCR Master Mix (TransGen Biotech, Beijing, China), 0.5 μ L each of forward and reverse primers (10 μ mol·L⁻¹), 2 μ L of template DNA (20–50 ng· μ L⁻¹), and nuclease-free water added to a final volume of 20 μ L. The amplification program was as follows: initial denaturation at 95 °C for 5 min; 35 cycles of denaturation at 95 °C for 30 s, annealing at 55–60 °C for 30 s, and extension at 72 °C for 45 s; followed by a final extension at 72 °C for 10 min and holding at 4 °C. The amplification products were detected using an ABI 3730xl automated capillary electrophoresis platform (Applied Biosystems, Foster City, CA, USA), with GeneScan 500 LIZ size standard (Applied Biosystems, Foster City, CA, USA) used as the internal size standard. Allele calling and genotyping were performed using GeneMapper v4.0 software [20]. To ensure the reliability of SSR genotyping, all PCR amplifications were performed with at least two independent technical replicates. Samples showing weak amplification, ambiguous peak profiles, or inconsistent allele calls were re-amplified and re-scored. A no-template negative control was included in each PCR batch to monitor potential contamination. Allele calls were independently checked by two researchers, and only reproducible genotypes were retained for subsequent parentage identification and genetic diversity analyses.

2.3. Parentage Identification

Parentage analysis was conducted using both genotype exclusion and maximum likelihood approaches. The genotype exclusion method was used primarily to verify whether offspring genotypes were compatible with the recorded parental combinations, whereas the likelihood-based method statistically assigned the most probable parent or parent pair among candidate parents [13]. In the genotype exclusion analysis, the multilocus SSR genotype of each offspring was compared with those of the recorded parental pair. This procedure constitutes a compatibility check, determining whether Mendelian incompatibilities exist between the offspring and the assumed parents, as opposed to serving as strict parentage identification. At each locus, the offspring was required to carry at least one allele inherited from the maternal parent and one allele inherited from the paternal parent. If the alleles of an offspring at a given locus could not be explained by the putative parental combination, a mismatch was recorded [13].

In the maximum likelihood analysis, parentage assignment was conducted using CERVUS 3.0.7 software based on multilocus genotype data and allele frequencies [21]. In contrast, the likelihood-based method performs statistical assignment of the most probable parent pair among candidate parents. The simulation analysis included 10,000 iterations. Following previous SSR-based parentage studies in perennial woody species, the number of candidate parents was conservatively set to five times the actual number of participating parents to account for potential pollen contamination and undocumented pollen donors. The candidate parent sampling proportion was set to 100% because all known parental materials involved in the controlled pollinations were sampled. A genotyping error rate of 1% was adopted because occasional allele dropout, stutter peaks, or scoring errors may occur during SSR genotyping, and this value has been widely used in previous CERVUS analyses [21].

2.4. Genetic Diversity and Population Structure Analyses

Based on the multilocus genotype data of 108 offspring individuals and seven parental lines, genetic diversity parameters were calculated for the selected SSR loci. These parameters included the number of alleles per locus (N_a), observed heterozygosity (H_o), expected heterozygosity (H_e), PIC, and null allele frequency [$F(\text{Null})$] [22]. In addition to standard genetic diversity parameters, private alleles were counted for each parental genotype and crossing combination to evaluate the genetic uniqueness of the candidate paternal materials. Private alleles were defined as alleles detected in only one parental genotype or one crossing combination within the analyzed dataset. This analysis was used to assess whether insufficient paternal-specific allelic variation contributed to the limited resolution of likelihood-based parentage assignment [22].

2.5. Genetic Distance and Clustering Analysis

Genetic distances among the 108 offspring individuals and their parents were calculated using GenAlEx v6.51 [23]. Based on the resulting genetic distance matrix, a clustering tree was constructed using MEGA11 to evaluate the overall genetic relationships between the offspring and parental materials [23]. All 15 SSR loci, including JM68820, were retained in all analyses to reflect practical marker usage conditions in walnut breeding programs.

3. Results

3.1. Genetic Diversity Analysis of SSR Loci

Fifteen SSR primer pairs were used to amplify DNA samples from 108 controlled-cross offspring and seven parental materials. A total of 68 alleles were detected, with an average of 4.53 alleles per locus. The PIC values ranged from 0.017 to 0.608, indicating

considerable variation in polymorphism among loci. Loci such as JH91908, JR3773, and JR6439 showed high PIC values and strong potential discriminatory power. Together with the subsequent exclusion results, these loci may serve as important auxiliary markers for parentage identification. In contrast, JC5411 and JC7329 showed low polymorphism and contributed little to offspring discrimination. Null allele frequency [F(Null)] varied considerably among loci, with JM68820 exhibiting a particularly high value, suggesting possible amplification bias or the presence of null alleles (Table 3). These results indicate that the informativeness of SSR loci depends not only on the number of alleles but also on the distribution of allele frequencies.

Table 3. Genetic diversity parameters of 15 SSR loci used in this study.

Locus	Na	Ho	He	PIC	F (Null)
JH2096	4	0.145	0.137	0.125	−0.027
JH6044	2	0.782	0.478	0.368	−0.231
JR3434	3	0.473	0.376	0.323	−0.081
JC5411	2	0.589	0.018	0.017	0.318
JC7329	2	0.487	0.230	0.136	0.125
JC8125	6	0.609	0.568	0.495	−0.030
JH89978	3	0.545	0.462	0.389	−0.102
JH42753	4	0.445	0.529	0.424	0.124
JH91908	7	0.664	0.667	0.608	−0.017
JR4616	5	0.755	0.584	0.547	−0.157
JR4964	4	0.518	0.489	0.395	−0.052
JH84548	2	0.582	0.500	0.374	−0.093
JM68820	11	0.145	0.614	0.553	0.644
JR3773	7	0.718	0.641	0.592	−0.066
JR6439	6	0.745	0.656	0.593	−0.078

3.2. Parentage Identification Analysis Using the Genotype Exclusion Method

The genotype exclusion method was used to verify the authenticity of 108 candidate offspring from seven controlled cross combinations. In total, 76 individuals were successfully assigned to their corresponding parental combinations, with an overall identification rate of 70.37% (Table 4). The identification rate varied markedly among combinations, ranging from 53.85% to 76.92%. Among them, the XNH × QPH4 and XNH × QPH3 combinations showed the highest identification rates, whereas the XNH × LWR combination showed the lowest rate. SSR loci such as JR3434, JH6044, JC8125, and JH91908 showed strong exclusion power across multiple combinations, indicating that loci with higher heterozygosity and more balanced allele frequencies are more favorable for improving identification efficiency.

Table 4. Verification results based on the genotype exclusion method.

Hybrid Combination	No. of Offspring	No. of Identified Offspring	Identification Rate (%)	Core Identification Loci
XNH × LYH	12	9	75.00	JH6044
XNH × LWR	13	7	53.85	JR3434
XNH × LBF	12	8	66.67	JH6044
XNH × QGR	6	4	66.67	JC8125
XNH × QPH4	13	10	76.92	JR3434
XNH × QPH3	13	10	76.92	JR3434
XNH × XNH	39	28	71.79	JH6044
Total	108	76	70.37	—

The variation in identification rates among different combinations may be associated with the shared maternal parent, XNH, across all combinations. This shared maternal structure increased allele sharing among different offspring populations, thereby reducing the discriminatory power of some SSR loci. In addition, close genetic relationships among some paternal parents may have further increased allele overlap, consequently limiting the overall identification efficiency of the genotype exclusion method (Table 4).

3.3. Parentage Assignment Analysis Using the Maximum Likelihood Method

Maximum likelihood parentage assignment analysis was performed on 108 controlled-cross offspring using CERVUS 3.0.7. When the genotypes of both parents were known, the combined exclusion probability for the parent pair (CE-PP) of the 15 SSR loci reached 0.9991, indicating that the selected loci collectively had high parentage identification capability. At the 95% confidence level, when the genotypes of both parents were unknown, only 27 offspring were successfully assigned to their most likely parent pairs, corresponding to an overall identification rate of 25.00%. When the maternal parent was known and only the paternal parent required assignment, 47 offspring were successfully matched to their corresponding paternal parents, increasing the identification rate to 43.52%. Among these offspring, some obtained high LOD scores in the CERVUS analysis and were assigned to their respective parents or paternal parents at the 95% or 80% confidence level (Table 5).

Table 5. Representative parentage assignment results based on CERVUS analysis.

Offspring ID	Female	LOD Score	Confidence Level	Male	LOD Score	Confidence Level
LYH-1	XNH	2.402	*	LYH	3.077	*
LYH-4		2.969	*	LYH		
LYH-8		3.077	*	LYH		
LYH-9		1.041	+	LYH		
LYH-12		2.408	*	LYH		
QGR-3		1.983	+	QGR	1.983	*
QPH4-1		2.390	*	QPH4		
QPH4-4		2.795	*	QPH4	6.001	*
QPH4-7		4.013	*	QPH4		
QPH4-8		2.100	+	QPH4		
QPH4-10		5.997	*	QPH4		
QPH4-11		6.316	*	QPH4		
QPH4-12		6.001	*	QPH4		
QPH4-13		3.966	*	QPH4		
QPH3-1		1.148	+	QPH3	4.106	*
QPH3-2		1.286	+	QPH3		
QPH3-3		4.502	*	QPH3		
QPH3-4		5.183	*	QPH3		
QPH3-5		1.560	+	QPH3		
QPH3-7		4.531	*	QPH3		
QPH3-8		3.169	*	QPH3		
QPH3-10		4.106	*	QPH3		
QPH3-12		2.431	*	QPH3		

Note: * denotes a 95% confidence level; + denotes an 80% confidence level.

Allele frequency analysis revealed that several SSR loci contained one or a few pre-dominant alleles, particularly at loci such as JH2096, JR3434, JH89978, JR4964, JR4616, JH84548, and JR3773 (Figure 1A,B). Such skewed allele-frequency distributions may reduce the discriminatory power of these loci and partially explain the limited resolution of the likelihood-based parentage assignment in closely related walnut breeding populations. Furthermore, the shared maternal parent, XNH, across all crosses markedly increased the

similarity of allele composition among different offspring populations. These observations indicate that, in closely related walnut populations, shared maternal structure, allele sharing, and skewed high-frequency alleles are major factors affecting the efficiency of SSR-based maximum likelihood parentage assignment (Figure 1). Assignment success was generally higher under the 95% confidence threshold than under the 80% threshold, indicating that assignment reliability was primarily constrained by population genetic structure rather than statistical threshold selection.

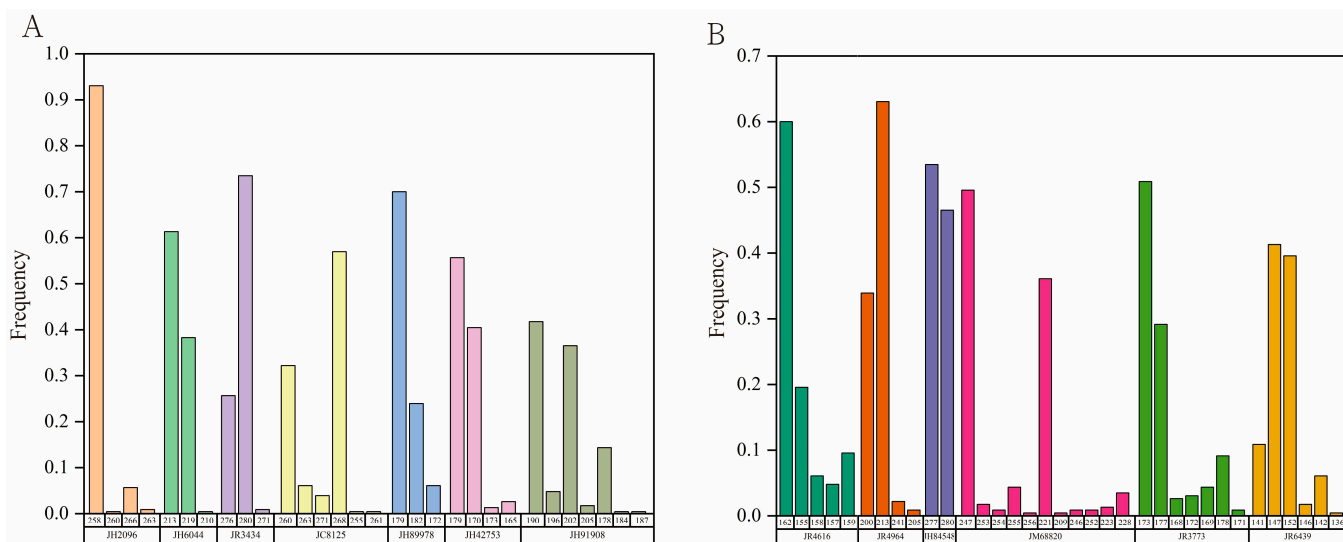


Figure 1. Allele frequency distribution of the 15 SSR loci used for walnut parentage assignment. (A) Allele frequency distributions of JH2096, JH6044, JR3434, JC8125, JH89978, JH42753, and JH91908; (B) Allele frequency distributions of JR4616, JR4964, JH84548, JM68820, JR3773, and JR6439.

3.4. Clustering Analysis of Controlled-Cross Offspring and Parents

Based on the SSR multilocus genotype data, 115 materials, including 108 offspring and seven parents, were divided into three main genetic clusters (Figure 2). Most offspring clustered near their corresponding parents, indicating that the selected SSR loci effectively reflected the genetic relationships among different hybrid populations. It should be noted that cluster analysis primarily reflects overall genetic similarity among individuals and does not provide direct evidence for pedigree confirmation. Among them, the XNH $\times$ XNH offspring formed a relatively compact cluster, suggesting high genetic similarity within this group. This clustering pattern may reflect increased genetic similarity among self-pollinated progeny, possibly resulting from the limited segregation of parental alleles during selfing or the relatively narrow genetic background of the parental genotype. However, this interpretation should be considered as a possible explanation rather than a direct measurement of heterozygosity or inbreeding level.

Some offspring, including LYH-1, QGR-1, QPH3-1, QPH3-2, and QPH4-14, did not cluster with the corresponding branches of their original cross combinations. This pattern may be associated with pollen contamination or shared genetic backgrounds among parental lines. Certain combinations derived from closely related paternal parents showed pronounced mixed clustering, indicating that multiple walnut materials had relatively similar genetic backgrounds and a high degree of allele sharing among different populations. This closely related genetic structure may further explain the relatively low accuracy of parentage assignment using the maximum likelihood method (Figure 2).

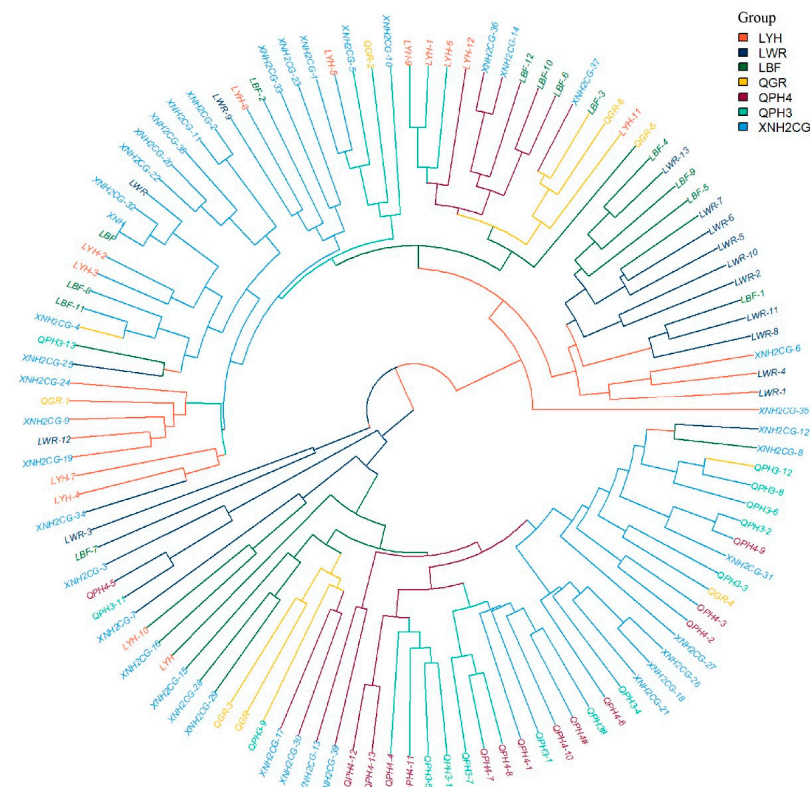


Figure 2. Cluster analysis diagram of 115 individuals.

4. Discussion

SSR markers have been widely established for parentage identification in fruit trees and woody plants; however, their resolution remains strongly influenced by population genetic structure in closely related walnut breeding populations [10,17]. In this study, 15 SSR loci were systematically analyzed across seven controlled walnut cross combinations. The genotype exclusion method showed higher consistency for verifying recorded cross combinations, whereas the likelihood-based assignment was more sensitive to allele frequency structure and candidate parent relatedness. It should be emphasized that these two methods address different analytical objectives and therefore their efficiency values are not directly comparable in a strict statistical sense. The results showed that the selected loci exhibited moderately high genetic diversity overall. The genotype exclusion method consistently enabled reliable verification of hybrid authenticity, whereas the maximum likelihood method showed markedly lower identification efficiency under a shared maternal-parent structure.

The mean PIC value of the SSR loci in this study was 0.396, which is comparable to the mean PIC value of 0.41 reported by Nurzhuma et al. [12] in Central Asian walnut populations. Wani et al. [17] reported PIC values mainly ranging from 0.30 to 0.50 in Himalayan walnut germplasm, and Zhang et al. [13] observed similar polymorphism levels in the identification of walnut F_1 offspring. These findings suggest that most cultivated walnut populations show moderate SSR genetic diversity overall, with limited genetic differentiation. Although JH91908, JR3773, and JR6439 did not have the highest allele numbers in this study, they showed greater stability in parentage exclusion analysis, possibly because of their more balanced allele frequency distributions. In contrast, some loci with higher allele numbers had limited discriminatory power due to an excessively high proportion of high-frequency alleles. This observation is consistent with Zhou et al. [16]

and Zheng et al. [6], who suggested that the informativeness of SSR loci depends not only on allele number but also on the structure of allele frequency distributions.

Compared with the genotype exclusion method, the maximum likelihood method showed substantially lower overall identification efficiency in this study. However, this result should be interpreted within the context of a highly structured breeding population with a shared maternal parent and closely related paternal lines, which likely increased allele sharing and reduced LOD differentiation among candidate parents. In particular, when both parents were unknown, only 25.00% of the offspring could be reliably assigned to their corresponding parent pairs. This finding differs from the high assignment efficiencies reported in some fruit tree studies [7]. Our results suggest that the shared maternal parent XNH may be an important factor contributing to insufficient LOD score differentiation. Because all crosses shared the same maternal parent, extensive maternal allele overlap occurred among different offspring populations, thereby compressing LOD differences among multiple candidate paternal parents. In addition, most paternal parents used in this study originated from the local walnut breeding population in Guizhou Province and had relatively close genetic backgrounds, which further increased allele sharing. Another possible explanation is that some SSR loci exhibited cryptic amplification bias, reducing the ability of the likelihood-based method to correctly identify the true paternal parent. Pollegioni et al. [15] and Harrison et al. [8] both noted that, in closely related populations or half-sib structures, a limited number of SSR loci often fail to maintain stable parentage assignment accuracy. The results of the present study are generally consistent with these observations.

JM68820 showed a typical pattern of high polymorphism coupled with a high null allele frequency. Therefore, the reduced assignment efficiency observed in the likelihood-based analysis should be interpreted as the combined effect of null alleles, marker heterogeneity, and the structured breeding population rather than a single methodological limitation. This locus had a PIC value of 0.553 but an F(Null) value as high as 0.644, which substantially exceeded the levels commonly reported in recent walnut SSR studies [14]. This pattern suggests substantial amplification bias or null allele interference at this locus. One possible explanation is that insertions, deletions, or local recombination events in the flanking sequences of this locus among different walnut genotypes may lead to marked differences in primer binding efficiency, thereby generating apparent homozygosity. Alternatively, extensive repeat sequence expansion in this region may compromise stable amplification [24]. Because JM68820 was retained in the analyses, its elevated null allele frequency may have contributed to reduced assignment confidence in CERVUS. Therefore, the relatively low assignment success observed in this study should be interpreted as the combined consequence of marker limitations and the highly structured breeding population rather than solely as a limitation of likelihood-based approaches. From an applied perspective, such loci may still have auxiliary value in the genotype exclusion method but should be used with caution in likelihood-based software such as CERVUS, because a high null allele frequency can lead to erroneous exclusion of true parents [24,25]. In addition to marker-specific effects, the relatively low assignment success observed in this study should also be interpreted in the context of multiple contributing factors. These include skewed allele frequency distributions across several SSR loci, the presence of null alleles (particularly at JM68820), close genetic relatedness among paternal parents derived from the same breeding population, and the possibility of unintended pollen contamination or undocumented pollination events during controlled crosses. Together, these factors likely reduce the differentiation of likelihood scores among candidate parents in structured breeding populations, thereby limiting the performance of likelihood-based assignment methods under such conditions. In contrast, loci such as JH91908 and JR3773 had both

high PIC values and low null allele frequencies, maintaining stable exclusion capability across multiple combinations [17]. These findings indicate that the combination of high PIC and low null allele frequency may be more important than the sole pursuit of high polymorphism in walnut parentage identification [6].

Cluster analysis further reflected the prevalence of closely related genetic backgrounds among walnut breeding materials, consistent with the low-to-moderate population differentiation and extensive genetic admixture observed in recent walnut population genetic studies [17,26]. Most offspring clustered near their corresponding parents, indicating that the selected SSR loci effectively reflected genetic relationships among populations, in agreement with the results of Manthos et al. [20] and Zhang et al. [6] in cluster analyses of walnut germplasm resources. However, some individuals, such as LYH-1, QGR-1, and QPH3-1, did not stably cluster within their original branches, suggesting relatively complex genetic origins. This pattern may be related to exogenous pollen input or long-distance pollen dispersal. Zhou et al. [16] found that more than 60% of offspring could not be assigned to nearby candidate paternal parents in SSR-based parentage analysis of walnut, indicating that pollen flow and unexpected pollination are common in walnut populations. Another possible explanation is the long-term sharing of ancestral gene pools among different landraces. Whole-genome resequencing studies have shown that most cultivated walnuts and landraces exhibit mixed ancestral components and shared genetic origins [5]. The XNH $\times$ XNH offspring formed a relatively tight cluster, suggesting that self-pollinated offspring had more homogeneous genetic backgrounds. This pattern may reflect reduced genetic differentiation among self-pollinated progeny; however, further population genetic analyses would be required to quantitatively evaluate the genetic variation within this group. The results also indicate that, even when cross combinations are clearly defined, offspring may still form mixed clusters because of extensive maternal allele sharing under a common maternal parent background. This genetic structure may reduce LOD differences among candidate paternal parents, thereby affecting the efficiency of parentage assignment using the maximum likelihood method [16].

5. Conclusions

In this study, 15 SSR loci were used for parentage identification and genetic relationship analysis of 108 candidate offspring from seven controlled walnut cross combinations. The results showed that the selected SSR loci had moderate polymorphism and parentage identification capability, with 68 alleles detected in total and a mean PIC value of 0.396. The genotype exclusion method successfully identified 76 offspring, achieving an overall identification rate of 70.37%, which was substantially higher than those obtained using the maximum likelihood method when both parents were unknown (25.00%) or when the maternal parent was known (43.52%). Cluster analysis showed that most offspring maintained relatively close genetic relationships with their corresponding parents, although some individuals showed clustering deviations, suggesting a relatively high degree of allele sharing among walnut breeding materials. Shared maternal structure, closely related paternal backgrounds, high-frequency alleles, and elevated null allele frequencies were identified as major factors affecting the efficiency of SSR-based parentage assignment. Under the breeding conditions examined in this study, the genotype exclusion method appeared to be more suitable for controlled walnut cross populations characterized by a shared maternal parent and multiple paternal parents.

Author Contributions: S.C. and T.Z. conceived and designed the study. S.C. and N.H. performed data curation and formal analysis. S.C. wrote the original draft. T.Z. and Y.Z. reviewed and revised the manuscript. Y.Z. and N.H. supervised the project and contributed to funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (31860215, 32560369); Innovation and Application of Guizhou Walnut Germplasm (Qianlin ke [2022] 17); the Guizhou Provincial Key Laboratory of Forest Cultivation in Plateau Mountain Areas; the Breeding of New Walnut and *Zanthoxylum* L. Varieties Project (Guilinkehe [2025] No. 9); the Research, Development, and Demonstration of an Efficient Cultivation Technology System for ‘Xueninghong’ Walnut (Qianlinkehe [2026] Support No. 002); and the Junior Researcher Growth Fund of Honghe Tropical Agriculture Institute of Yunnan (HHS2026006).

Institutional Review Board Statement: All experimental procedures complied with institutional, national, and regional guidelines for plant research in China.

Data Availability Statement: The datasets generated and analyzed during the current study, including SSR genotype data and parentage assignment results, are available from the corresponding author upon reasonable request.

Acknowledgments: We thank Lianwen Shen and Zhenshan Liu for their valuable assistance in sample collection and data analysis during this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhou, B.-B.; Ma, Q.-G.; Wang, H.; Wang, W.; Zhang, M.-Y.; Pei, D. Establishment of a parental identification method for walnut. *Sci. Agric. Sin.* **2011**, *44*, 4258–4264. (In Chinese)
2. Chen, L.; Ma, Q.; Chen, Y.; Wang, B.; Pei, D. Identification of major walnut cultivars grown in China based on nut phenotypes and SSR markers. *Sci. Hortic.* **2014**, *168*, 240–248. [\[CrossRef\]](#)
3. Bernard, A.; Lheureux, F.; Dirlewanger, E. Walnut: Past and future of genetic improvement. *Tree Genet. Genomes* **2017**, *14*, 1. [\[CrossRef\]](#)
4. Zhang, S.; Li, Y.; Li, Y.; Zhang, Y.; Hao, Y.; Hou, Z.; Qi, J. Development of SSR Markers for and Finger-printing of Walnut Genetic Resources. *Forests* **2024**, *15*, 405. [\[CrossRef\]](#)
5. Liu, Y.; Zhang, P.; Wang, Z.; Fu, J.; Han, J.; Gao, Y.; Feng, Z. Genetic analysis of wild walnuts in Xinjiang based on whole-genome resequencing. *Front. Plant Sci.* **2025**, *16*, 1645319. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Zheng, C.; Wang, W.; Bai, B.; Li, P.; Wu, N.; Xing, W.; Wang, J.; Yang, J. Parentage assignment of walnut hybrid offspring by microsatellite markers. *J. Sichuan For. Sci. Technol.* **2024**, *45*, 103–108. (In Chinese)
7. Alexander, L.; Wu, X.; Gouker, F. Production and verification of novel *Osmanthus* hybrids. *Front. Hortic.* **2024**, *3*, 1382450. [\[CrossRef\]](#)
8. Harrison, H.B.; Saenz-Agudelo, P.; Planes, S.; Jones, G.P.; Berumen, M.L. Relative accuracy of three common methods of parentage analysis in natural populations. *Mol. Ecol.* **2013**, *22*, 1158–1170. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Dangl, G.S.; Woeste, K.; Aradhya, M.K.; Koehmstedt, A.; Simon, C.; Potter, D.; Leslie, C.A.; McGranahan, G. Characterization of 14 Microsatellite Markers for Genetic Analysis and Cultivar Identification of Walnut. *J. Am. Soc. Hortic. Sci.* **2005**, *130*, 348–354. [\[CrossRef\]](#)
10. Testolin, R.; Messina, R.; Cipriani, G.; De Mori, G. SSR-based DNA fingerprinting of fruit crops. *Crop Sci.* **2023**, *63*, 390–459. [\[CrossRef\]](#)
11. Dangl, G.S.; Yang, J.; Golino, D.A.; Gradziel, T. A practical method for almond cultivar identification and parental analysis using simple sequence repeat markers. *Euphytica* **2009**, *168*, 41–48. [\[CrossRef\]](#)
12. Nurzhuma, M.; Kokhmetova, A.; Kumarbayeva, M.; Keishilov, Z.; Bakhytuly, K.; Bolatbekova, A.; Kokhmetova, A.; Mukhamet-zhanov, K.; Akan, K. Assessment of Genetic Diversity in Walnut (*Juglans regia* L.) Genotypes from Southern and Southeastern Kazakhstan Using Microsatellite Markers. *Horticulturae* **2025**, *11*, 810. [\[CrossRef\]](#)
13. Zhang, S.; Li, Y.; Li, Y.; Zhang, Y.; Hao, Y.; Chen, Y.; Hou, Z.; Qi, J. Identification and genetic diversity analysis of specific walnut F1 progeny based on SSR molecular markers: Taking heart-shaped walnuts and Jinghong 1 as examples. *Sci. Rep.* **2024**, *14*, 27869. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Itoo, H.; Shah, R.A.; Qurat, S.; Jeelani, A.; Khursheed, S.; Bhat, Z.A.; Mir, M.A.; Rather, G.H.; Zargar, S.M.; Shah, M.D.; et al. Genome-wide characterization and development of SSR markers for genetic diversity analysis in northwestern Himalayas Walnut (*Juglans regia* L.). *3 Biotech* **2023**, *13*, 136. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Pollegioni, P.; Woeste, K.; Mugnozza, G.S.; Malvolti, M.E. Retrospective identification of hybridogenic walnut plants by SSR fingerprinting and parentage analysis. *Mol. Breed.* **2009**, *24*, 321–335. [\[CrossRef\]](#)

16. Zhou, H.; Zhao, P.; Woeste, K.; Zhang, S. Gene flow among wild and cultivated common walnut (*Juglans regia*) trees in the Qinling Mountains revealed by microsatellite markers. *J. For. Res.* **2021**, *32*, 2189–2201.
17. Wani, A.W.; Hassan, G.I.; Bhat, K.M.; Ahmad, M.; Siddiqui, M.H.; Kumar, S.; Naresh, R.K.; Gupta, R.K. Utilizing SSR markers to examine the population structure and molecular genetic diversity of walnut (*Juglans regia* L.) genotypes in the northwestern Himalayan region of Jammu and Kashmir. *BioResources* **2024**, *19*, 4213–4237. [[CrossRef](#)]
18. Çilesiz, Y. Genetic diversity assessment of walnut (*Juglans regia* L.) genotypes from inner Anatolia region Türkiye using ISSR and RAPD markers. *Sci. Rep.* **2025**, *15*, 30850. [[CrossRef](#)] [[PubMed](#)]
19. Guney, M.; Kafkas, S.; Keles, H.; Zarifikhosroshahi, M.; Gundesli, M.A.; Ercisli, S.; Necas, T.; Bujdoso, G. Genetic Diversity among Some Walnut (*Juglans regia* L.) Genotypes by SSR Markers. *Sustainability* **2021**, *13*, 6830. [[CrossRef](#)]
20. Manthos, I.; Sotiropoulos, T.; Karapetsi, L.; Ganopoulos, I.; Pratsinakis, E.D.; Maloupa, E.; Madesis, P. Molecular Characterization of Local Walnut (*Juglans regia*) Genotypes in the North-East Parnon Mountain Region of Greece. *Int. J. Mol. Sci.* **2023**, *24*, 17230. [[CrossRef](#)] [[PubMed](#)]
21. Azka, N.; Taryono; Wulandari, R. Assessment of tea plant (*Camellia sinensis* L.) accessions for pollen sources in natural crossing by using microsatellites. *SABRAO J. Breed. Genet.* **2021**, *53*, 673–684. [[CrossRef](#)]
22. Yin, J.; Zhao, H.; Wu, X.; Ma, Y.; Zhang, J.; Li, Y.; Shao, G.; Chen, H.; Han, R.; Xu, Z. SSR marker based analysis for identification and of genetic diversity of non-heading Chinese cabbage varieties. *Front. Plant Sci.* **2023**, *14*, 1112748–1112761. [[CrossRef](#)] [[PubMed](#)]
23. Wang, C.-y.; Huang, J.-c.; Yin, M.-y.; Hu, H.-j.; Yang, Y.-p.; Guo, J.-j.; Zeng, J. Geographic isolation and environmental heterogeneity shape population genetic differentiation of a medicinal plant, amla (*Phyllanthus emblica* L.) in river valleys of Yunnan, China. *Front. Plant Sci.* **2025**, *16*, 1648822. [[CrossRef](#)] [[PubMed](#)]
24. Jahnke, G.; Smidla, J.; Deák, T.; Oláh, R.; Szóke, B.Á.; Nyitrai Sárdy, D.Á. The SSR Null Allele Problem, and Its Consequences in Pedigree Reconstruction and Population Genetic Studies in Viticulture. *Horticulturae* **2022**, *8*, 658. [[CrossRef](#)]
25. Wen, Y.F.; Uchiyama, K.; Han, W.J.; Ueno, S.; Xie, W.D. Null alleles in microsatellite markers. *Biodivers. Sci.* **2013**, *21*, 117–126. [[CrossRef](#)]
26. Ye, L.; Shavvon, R.S.; Qi, H.; Wu, H.; Fan, P.; Shalizi, M.N.; Khurram, S.; Davletbek, M.; Turuspekoy, Y.; Liu, J. Population genetic insights into the conservation of common walnut (*Juglans regia*) in Central Asia. *Plant Divers.* **2024**, *46*, 600–610. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.