Research Article

Molecular Sex Identification in the Hardy Rubber Tree (Eucommia ulmoides Oliver) via ddRAD Markers

Wencai Wang,1,2 Guoqian Yang,3 Xin Deng,4 Fengqing Shao,1,2 Yongquan Li,4 Wei Guo,4 Hong Liang,4 and Xianzhi Zhang4

1Science and Technology Innovation Center, Guangzhou University of Chinese Medicine, Guangzhou 510405, China
2Institute of Clinical Pharmacology, Guangzhou University of Chinese Medicine, Guangzhou 510405, China
3School of Agriculture and Biology, Shanghai Jiao Tong University, China
4College of Horticulture and Landscape Architecture, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China

Correspondence should be addressed to Xianzhi Zhang; zhangxianzhi@zhku.edu.cn

Received 21 January 2020; Accepted 3 April 2020; Published 16 May 2020

Academic Editor: Antonio Ferrante

Copyright © 2020 Wencai Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Eucommia ulmoides, also known as the industrially and medicinally important hardy rubber tree, is the sole species of Eucommiaceae. Nevertheless, its dioecious property hinders sex recognition by traditional morphological observation at very early developmental stages, thus inhibiting breeding and economic cropping. In this study, double-digest restriction site-associated DNA sequencing (ddRAD-seq) was applied to screen sex-linked molecular markers for sex identification and investigation of the sex determination system in 20 male and female E. ulmoides individual plants, respectively. In consequence, five candidate male-specific loci but no female-specific loci were predicted among the 183,752 male and 147,122 female catalogue loci by bioinformatics analysis. Subsequent PCR (polymerase chain reaction) amplification and Sanger sequencing examinations were performed on another 24 individuals, 12 for each sex, from a separate population. One ideal sex-linked locus, MSL4, was identified among the five putative male-specific loci that were found using ddRAD data. MSL4 is 479 bp in length and highly conserved in all the male individuals, suggesting its feature of being stable and repeatable. Our results also indicated that the sex of E. ulmoides is likely determined genetically. In short, this study provides a consistent and reproducible ddRAD marker (MSL4) that is able to discriminate male from female seedlings in E. ulmoides, which will be valuable for rapid breeding practice and better commercial production of this economically important tree.

1. Introduction

Eucommia ulmoides Oliver, belonging to the monotypic family Eucommiaceae [1], is an economically important dioecious tree endemic to southern and central China [2]. It is a well-known Tertiary relict tree species and widely cultivated as a medicinal plant in China for more than 2,000 years [3]. In recent years, this species has been called as the hardy rubber tree due to the high productivity of gutta, i.e., TPI (trans-1,4-polyisoprene) in the whole plant [4–6]. Because of its wide adaptability in the subtropical and temperate regions, E. ulmoides is therefore to potentially supersede the commonly known para rubber tree (Hevea brasiliensis in Euphorbiaceae) that is restricted in the tropical zones [7, 8]. However, sex identification of the young seedlings of E. ulmoides is rather difficult for its dioecious sexual system and long life cycle (up to 8-10 years), which limits the efficiency of breeding for improving the gutta (Eu-rubber) yield or other agronomic traits [9].

Dioecy represents the sex separation in distinct individuals, which resembles the majority of mammals that include individuals of males and females [10]. Only about 6% of the flowering plants, i.e., 15,600 species in 987 genera of 175 families, however, are dioecious [11]. In general, there are three kinds of sex determination system via sex chromosomes that have been revealed in dioecious plants: XY, ZW, and XO [12]. XY sex chromosome system has heterogametic sex in males (XY) while females are homogametic XX, which is
determined in Actinidia chinensis [13], Asparagus officinalis [14], Carica papaya [15], and Silene latifolia [16]. In contrast, ZW sex chromosome system features female heterogamety (ZW) and male homogamety (ZZ), such as in Populus trichocarpa [17]. Some other species, e.g., Rumex acetosa, evolve an XO sex determination system, in which the sex of one individual is determined by the X-to-autosome ratio [18]. Until now, heteromorphic sex chromosomes are only verified in 39 plants by reliable cytogenetic and/or molecular evidence [19]. The dioecious hardy rubber tree E. ulmoides has a diploid genotype with a basic composition of 2n = 34 chromosomes and circa 1.2 Gb of genome size [20]. Nevertheless, it remains uncertain whether E. ulmoides has a genetic sex determination or heteromorphic sex chromosome [21, 22], leading to the difficulty and unreliability to identify the sex of E. ulmoides during the vegetative growth phase via karyotype [23].

Moreover, with the advent of genomics, sex-determining genes are emerging in several plants [24]. For instance, in kiwifruit (Actinidia spp.), Y-encoded genes SyG1 and FrBy act independently as the suppressor of feminization (SuF) and the maintenance of male (M) [25]. In persimmon (Diospyros spp.), a Y-encoded psuedogene OGI is sufficient for the expression of androecium and repression of gynoecium through encoding a small RNA and targeting its autosomal counterpart gene MeGI [26]. Two Y-encoded factors, named SOFF and aspTDF, have been discovered in garden asparagus (A. officinalis) for the expression of maleness and repression of female development [27]. Nonrecombinant sex-specific regions including many genes have also been characterized in C. papaya [28], P. trichocarpa [29], and S. latifolia [30]. Recently, one putative MADS box APETALA3-like gene, with male-biased expression, was identified to be probably involved in the sex determination in E. ulmoides based on transcriptomic analysis [31]. Nevertheless, to our knowledge, the sex determination mechanism in E. ulmoides is still in the dark to date.

Sexual dimorphisms for instance morphological and physiological differences between the genders have been reported in dioecious plants [32, 33]. The gutta yields display a remarkable sexual dimorphism between the leaves of male and female trees of E. ulmoides [21]. The fruit accumulates the highest content of gutta among the leaves, barks, roots, fruits, and seeds of E. ulmoides [7]. An optimal proportion of 6-8% of males is considered to be sufficient for providing sperm in the Eucommia orchard [23]. It is thus essential to develop a simple, accurate, and effective method for sex identification of E. ulmoides long before phenomorphological features appear to satisfy the optimal sex ratio for gutta production.

Ideal sex-linked molecular markers can assist to select off the low-value male seedlings and to keep pistillate trees at the juvenile stage, consequently to enlarge economic benefits of the orchard, as that applied in the kiwifruit [34] and sea buckthorn (Hippophae rhamnoides) [35]. Restriction site-associated DNA sequencing (RAD-seq) and its modified version double-digest RAD-seq (ddRAD-seq), high-throughput sequencing methods developed on the basis of next-generation sequencing, enable the discovery of tens of thousands of molecular markers for nonmodel species in ecological and evolutionary studies [36, 37]. Considering that RAD-seq/ddRAD-seq has been successfully used to identify sex-linked DNA sequences in dioecious plants, e.g., in Excoecaria agallocha [38], kiwifruit [39], and S. latifolia [40], in this study, we employed a slightly modified ddRAD (MiddRAD) technique [37] to develop sex-specific DNA markers for gender identification of E. ulmoides during the vegetative growth phase.

2. Materials and Methods

2.1. Plant Materials. A total of 20 female and 20 male trees from a population of E. ulmoides cultivated on the campus of Northwest A&F University, Yangling, Shaanxi, China (34°16′ 46″ N, 108° 04′ 27″ E) were sampled for ddRAD marker development. Their sexes were identified by flower traits during the flowering season (April in 2017), when discrimination of the staminate and pistillate genotypes was decided (Figure 1). For each individual, fresh healthy leaves were collected before being immersed into liquid nitrogen immediately and then stored at -80°C until DNA isolation. In addition, we sampled 12 male and 12 female individuals from Henan Normal University in Xinxiang, Henan, China (34°41′ 33″ N, 113° 40′ 16″ E) in April 2018 as a separate population to validate the sex-specific markers. The distance between these two sampling points is about 650 km.

2.2. DNA Isolation, ddRAD Library Construction, and Illumina Sequencing. Genomic DNA was isolated from the leaf samples using the CTAB method [41]. The quality of DNA was monitored on 1.0% (m/v) agarose gel and NanoPhotometer® spectrophotometer (Implen, CA, USA). We constructed a ddRADseq library for the 20 male individuals and 20 female individuals following the MiddRAD protocol [37]. In brief, Avall+MspI enzyme pair was used to digest the purified genomic DNA and ligated the barcoded adaptors. Then, the DNA samples were pooled and fragments with 500-600 bp in length were selected for PCR amplification. The enriched DNA was purified and measured by gel and Qubit 2.0 Fluorometer (Life Technologies, CA, USA). Finally, the library was sequenced on the Illumina HiSeq 2500 platform at Novogene, Beijing, China, to generate paired-end (PE) reads with 150 nucleotides in length.

2.3. ddRAD-seq Data Analysis. Raw reads were demultiplexed using the process_radtags algorithm in Stacks software version 1.24 [42, 43]. Average sequence quality per read was estimated by FastQC version 0.11.3 [44], and adapter reads were determined using Cutadapt 1.9.1 [45]. Reads of low quality (below a Phred score of Q10), with ambiguous barcodes (the presence of mismatch), or missing the restriction site, and the adaptor reads were removed to produce clean data. All the clean reads were truncated to a final length of 135 base pairs (bp), excluding the barcode and low-quality nucleotides at the 3′ end, for subsequent analyses.

The usstacks module in Stacks was used firstly to merge short-read sequences into tags/loci at the set of -m 10, -M 3,
in which \( m \) is the minimum number of identical reads required to create a stack and \( M \) is the maximum distance (in nucleotides) allowed between stacks [38]. Then, a catalog was built by the \( cstacks \) program \((-n 3\) for all these individuals, in which \( n \) is the maximum number of mismatches allowed between the loci in the catalog. Finally, the output of \( sstacks \) (matches.tsv) from each sample was used to identify putative sex-specific loci by the R script [46]. The ddRAD locus that was present in \( \geq N-1 \) individuals of one sex (\( N \) is the total number of individuals of one gender) while absent in all individuals of the other sex was defined as putative sex-specific markers [38].

Since each member of the PE reads was treated as an independent locus in the above analysis, we further investigated if there were cases where two of the predicted sex-specific ddRAD loci represented two members of a PE read pair. In addition, clean reads of the opposite sex were mapped to each of the chosen loci by the Linux grep command to rule out the possibility of any exact matches in sequence to the selected locus. These two steps can narrow down the number of candidate markers [38].

### 2.4. PCR Validation of Putative Sex-Specific ddRAD Loci

We subsequently conducted PCR to validate these putative sex-specific ddRAD loci based on the additional 24 individuals (12 for each sex) collected from one separate population in Henan. The forward and reverse primers were designed with Primer3 [47] and anchored the sequences of the putative sex-linked ddRAD loci (Table S1). We used the volume of 25 \( \mu L \) for PCR amplification reaction: 12.5 \( \mu L \) PCR mix (Tiangen Biotech, Beijing, China), 2.0 \( \mu L \) random decamer primers (25 \( \mu \text{mol} \cdot \mu L^{-1} \)), 0.5 \( \mu L \) genomic DNA (100 ng \cdot \mu L^{-1}), and 10.0 \( \mu L \) ddH\(_2\)O. The following PCR profile was used: an initial denaturation at 94°C for 3 min followed by 30 cycles of denaturation (30 s at 94°C), annealing (30 s at 59°C) and extension (1 min at 72°C), and a final extension 72°C for 5 min. The PCR amplification products were finally examined on a 1.0% agarose gel. Each PCR experiment was repeated twice to validate both the reliability and stability.

### 2.5. Sanger Sequencing of the Verified ddRAD Marker

In addition, for the verified ddRAD marker that was only amplified via PCR in males but not in females (male-specific locus MSL4), we sequenced the band by Sanger sequencing on ABI 3730X at Sangon Biotech, Shanghai, China. We used the MSL4-F primer (Table S1) as the sequenced primer, and all the 12 male individuals used in PCR validation were sequenced. The obtained 12 sequences were aligned to produce a consensus sequence in Geneious v11.1 [48], which was searched against the \textit{in silico} predicted sex-specific ddRAD loci by blastn with \( e \) value \(< 10^{-5} \) [49] to confirm their origin of same genomic position. The identified ddRAD marker was also blast to the NCBI nucleotide database (https://www.ncbi.nlm.nih.gov/) to investigate their potential functional information. Furthermore, we run the local blastn by setting the marker MSL4 as query and 

### 3. Results

#### 3.1. ddRAD Sequencing and Sex-Linked Loci Development

A total of 40 individuals of \( E.\ ulmoides \), with each sex having 20 individuals, were sequenced according to the ddRAD-seq protocol. About 2-3 Gb data were produced for each individual, reaching a total of 180,645,228 paired-end (PE) reads (read length = 150 bp, 100 Gb). After demultiplexing and quality filtering, we got a total of 178,740,749 reads (read length = 135 bp, 54 Gb). The number of reads in the 40 samples varied from 2,186,096 to 6,742,116 (Table 1). The \( ustacks \) analysis recovered 20,807 to 95,993 ddRAD tags in these samples. Using \( cstacks \), we obtained 183,752 loci and 147,122 loci in males and females, respectively.

Through screening the \( sstacks \) output (matches.tsv), we preliminarily detected 11 candidate male-linked ddRAD loci.
that were present in 19/20 males and none in the 20 females. Based on further checking, if these loci represented a pair of PE reads and had no exact match of female clean reads, we finally identified five loci as putative male-specific ddRAD markers, which were denoted as MSL1, MSL2, MSL3, MSL4, and MSL5 (Table 2). Given that the DNA fragments of ddRAD-seq are about 500 bp in length and should be tightly linked, we considered each marker as one independent genomic locus. Interestingly, there were no candidate female-linked ddRAD loci that were found in our analysis. The ddRAD-seq data have been deposited in the Sequence Read Archive (SRA) of the NCBI database, with the accession number PRJNA607161.

3.2. Male-Specific ddRAD Marker Examination. A total of 12 male individuals and 12 female individuals collected from the independent Henan population were used to examine the identified candidate male-specific ddRAD loci (Table 2) by PCR amplification and Sanger sequencing. We found that none of the PCR gel banding of MSL1, MSL2, MSL3 and MSL5 were male-specific (Figure 2), even appearing no difference between the male and female ones (Figures 2(a), 2(b), and 2(d)). In contrast, the agarose gel electrophoresis of PCR products from the locus MSL4 displayed a distinct male-specific pattern, showing one single bright band with the expected size, about 500 bp, in each of the 12 male individuals whereas no banding in any of the 12 female individuals (Figure 3). Additionally, the results were consistent in two experimental replications.

We further sequenced the male-specific band (MSL4) from 12 male individuals by Sanger sequencing. Sequences of MSL4 were highly conserved, with no sequence variations, in the 12 staminate genotypes (Figure 4(a)). The consensus sequence of this marker was determined as 479 bp in length (Figure 4(b), as expected (circa 500 bp, Figure 3). Moreover, the produced 479 nucleotides matched perfectly with ddRAD sequences of MSL4 (a pair of 135 bp reads, Table 2), indicating the same genomic origin. The blast analysis found no hits for this male-specific ddRAD marker in the NCBI nucleotide database (https://www.ncbi.nlm.nih.gov/). Luckily, we detected a high similarity between the positions of 25,047-25,525 in GWHAAAL00024598#OriSeqID=scaffold571_obj#Len=331914 (Eucommia genome scaffolds, https://bigd.big.ac.cn/gwh/Assembly/13/show) and MSL4. The detailed results are as follows: Score = 835 bits (452), Expect = 0, Identities = 470/479 (98%), Gaps = 0/479 (0%), and Strand=Plus/Plus. We furthermore ascertained the annotation information of this scaffold (#OriSeqID=scaffold571_obj, Accession=GWHAAAL00024598) and found that MSL4 is probably non-protein-coding sequences located between CDS17 (24,544-24,670) and CDS18 (25,716-25,793) of a putative gene with unknown function.

4. Discussion

The evolution of dioecy has long been an intriguing issue since Darwin’s time [50]. However, many dioecious plants have homomorphic sex chromosomes, e.g., wild strawberry (Fragaria virginiana) and garden asparagus (A. officinalis), which are difficult to be cytologically detectable for the lack of visible crossovers or of pairing [12, 51]. Until now, reports about the sex chromosomes of Eucommia ulmoides are missing and gender discrimination by karyotype is undependable [21, 22]. Moreover, young and probably small sex-linked

<table>
<thead>
<tr>
<th>Male</th>
<th>No. of reads</th>
<th>No. of ddRAD tags</th>
<th>Female</th>
<th>No. of reads</th>
<th>No. of ddRAD tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>3,824,332</td>
<td>41,724</td>
<td>F1</td>
<td>3,621,226</td>
<td>47,001</td>
</tr>
<tr>
<td>M2</td>
<td>4,129,114</td>
<td>64,385</td>
<td>F2</td>
<td>2,253,982</td>
<td>23,616</td>
</tr>
<tr>
<td>M3</td>
<td>4,295,004</td>
<td>54,496</td>
<td>F3</td>
<td>5,256,178</td>
<td>79,809</td>
</tr>
<tr>
<td>M4</td>
<td>5,432,212</td>
<td>79,735</td>
<td>F4</td>
<td>4,352,006</td>
<td>64,904</td>
</tr>
<tr>
<td>M5</td>
<td>3,535,126</td>
<td>42,335</td>
<td>F5</td>
<td>6,369,882</td>
<td>92,210</td>
</tr>
<tr>
<td>M6</td>
<td>4,908,108</td>
<td>69,511</td>
<td>F6</td>
<td>4,485,857</td>
<td>56,084</td>
</tr>
<tr>
<td>M7</td>
<td>3,614,255</td>
<td>40,623</td>
<td>F7</td>
<td>2,816,328</td>
<td>30,228</td>
</tr>
<tr>
<td>M8</td>
<td>6,709,238</td>
<td>82,083</td>
<td>F8</td>
<td>4,107,792</td>
<td>56,497</td>
</tr>
<tr>
<td>M9</td>
<td>2,664,860</td>
<td>32,317</td>
<td>F9</td>
<td>3,809,033</td>
<td>46,428</td>
</tr>
<tr>
<td>M10</td>
<td>2,355,478</td>
<td>20,807</td>
<td>F10</td>
<td>2,727,482</td>
<td>24,611</td>
</tr>
<tr>
<td>M11</td>
<td>3,684,222</td>
<td>40,937</td>
<td>F11</td>
<td>5,165,036</td>
<td>82,508</td>
</tr>
<tr>
<td>M12</td>
<td>2,706,853</td>
<td>30,620</td>
<td>F12</td>
<td>5,593,874</td>
<td>76,306</td>
</tr>
<tr>
<td>M13</td>
<td>5,690,638</td>
<td>72,379</td>
<td>F13</td>
<td>6,251,392</td>
<td>95,993</td>
</tr>
<tr>
<td>M14</td>
<td>6,417,624</td>
<td>94,810</td>
<td>F14</td>
<td>3,956,216</td>
<td>38,078</td>
</tr>
<tr>
<td>M15</td>
<td>3,890,158</td>
<td>42,191</td>
<td>F15</td>
<td>6,330,162</td>
<td>90,762</td>
</tr>
<tr>
<td>M16</td>
<td>6,742,116</td>
<td>93,592</td>
<td>F16</td>
<td>4,312,874</td>
<td>63,217</td>
</tr>
<tr>
<td>M17</td>
<td>5,850,922</td>
<td>84,222</td>
<td>F17</td>
<td>5,295,320</td>
<td>78,552</td>
</tr>
<tr>
<td>M18</td>
<td>5,774,106</td>
<td>84,796</td>
<td>F18</td>
<td>3,544,489</td>
<td>35,002</td>
</tr>
<tr>
<td>M19</td>
<td>3,890,228</td>
<td>46,438</td>
<td>F19</td>
<td>4,686,812</td>
<td>65,568</td>
</tr>
<tr>
<td>M20</td>
<td>2,186,096</td>
<td>30,450</td>
<td>F20</td>
<td>5,504,118</td>
<td>84,643</td>
</tr>
</tbody>
</table>
regions are intractable to be uncovered because they have stopped recombination during meiosis [19, 24]. Up to now, only a few of sex determination genes are dissected in Actini-
dia, Asparagus, Diospyros, and so on [10, 52]. Genetic sex determination with either male or female heterogamety in *E. ulmoides* is controversial [23]. Sex-specific molecular
Markers have been proven to be successful in identifying gender of seedlings, inferring sex determination system, and seeking sex determination genes [24, 51]. In this circumstance, we report a reproducible and consistent ddRAD-PCR method for sex identification of *E. ulmoides* based on the high-throughput sequencing technology.

Molecular markers such as SSR (simple sequence repeat), AFLP (amplified fragment length polymorphism), SCARs (sequence-characterized amplified regions) and RAPD (random amplification of polymorphic DNA) have been commonly used in exploring sex-specific markers of dioecious plants [34, 35]. However, a large amount of primers or primer pairs are required to screen sex-linked markers in these conventional methods, which are extremely labor-intensive and time consuming. For example, a total of 1000 decamer primers were screened in order to identify a female-linked SCAR marker in *Humulus lupulus* [53]; in *A. officinalis*, 760 primers were used for developing a SCAR marker associated with the M (male-determining) locus [54]. Especially for tree species like *E. ulmoides*, it is even more difficult to screen sex-linked markers via conventional methods (e.g., SSR, AFLP, and RAPD) since experimental crosses and linkage map constructions are often impractical. For instance, in *P. vera* [55], only one 950 bp marker was found to be tightly linked to the female phenotype by screening 400 primers; a total of 45 decamer primers were examined for detecting a female-linked RAPD-SCAR detected in *H. rhamnoides* [56]. A previous study on *E. ulmoides* explored a pair of primers (EST-Eu059) out of 140 pairs of EST-SSR primers as sex-linked marker that displayed an extra band in male plants [57]. In contrast, the ddRAD-seq method used here generated a great number of genome-wide markers for *E. ulmoides* but with far less labor and cost. We obtained a high density of ddRAD tags for 20 males (183,752) and 20 females (147,122) of *E. ulmoides* (Table 1) across the genome, based on which five putative sex-specific markers were found (Table 2).

**Figure 3**: Agarose gel electrophoresis of PCR products of the male-specific ddRAD marker MSL4 in *Eucommia ulmoides*. Left lanes: male 1-12, right lanes: female 1-12. The result is robust in experimental replicates 1 and 2.

**Figure 4**: Sanger sequencing of the male-specific ddRAD marker MSL4 in *Eucommia ulmoides*. (a) Pairwise sequence alignments derived from MUSCLE program showing highly conserved MSL4 in males. The sequence identity is indicated on the top, and label EM1-12 represents the 12 males of *E. ulmoides* used for marker examination. (b) Consensus nucleotide sequence of the male-specific amplicon MSL4.
segregation [39, 40, 58]. By bioinformatics analysis of ddRAD-seq data, we detected five candidate male-specific markers in *E. ulmoides* (Table 2). Based on further examination of these putative markers by PCR amplification in another independent population (Figures 2 and 3), we successfully identified a reliable sex-specific marker (MSL4) for PCR-based gender identification. The highly specialized banding pattern of MSL4 between males and females is very stable in different experimental replications (Figure 3). It is thus an ideal marker for sex identification of the seedling of *E. ulmoides* with simple PCR and agarose gel electrophoresis, similar to the SCAR markers developed for other dioecious species [35, 56]. Given that it is crucial to identify the gender of young trees for the agricultural industry in perennial woody plants with important economic values, just as the hardy rubber tree (*E. ulmoides*), this study provides a convenient and cost-effective method to screen sex-specific markers for gender discrimination. In addition to present study, the ddRAD-PCR technique has been proven solid in identifying sex-specific markers in the dioecious mangrove tree, *E. agallocha* [38] as well.

Gender identification of the young seedlings of *E. ulmoides* is substantially important in the breeding practice [59, 60]. As a tree species, *E. ulmoides* has to experience 8-10 years of juvenility period before flowering and fruiting [23]; therefore, early ascertaining the sex of seedlings can accelerate the artificial selection in breeding programs. The stable male-specific marker (MSL4) developed in this study is able to be used for rapid sex identification of *E. ulmoides* long before maturity when morphological characteristics of flowers are invisible. Additionally, this marker identified in the Shaanxi population of *E. ulmoides* also worked in the Henan population, which is suggesting that MSL4 probably is a universal marker for all populations of *E. ulmoides* across their national or global distributions. *E. ulmoides* has drawn much attention for hardy rubber (gutta) production during recent years [7, 61]. Previous studies have revealed sexual dimorphism in the leaf gutta content between males and females [21] and the highest gutta accumulation in fruit [23]. Therefore, it is essential to discern the sexes of seedlings of *E. ulmoides* with the molecular marker MSL4 to assure optimal sex ratio in *E. ulmoides* orchard, which will largely enhance their economic value.

Exploring sex-determining regions in nonmodel plant species may help us to understand the evolution of recombination suppression within young sex chromosomes [10, 62]. We sequenced the MSL4 fragment of 12 staminate genotypes from Henan population and found that there were no sequence variations at this locus (Figure 4(a)). Alignment of the 479 bp nucleotides (Figure 4(b)) and corresponding sequences from ddRAD-seq analysis (Table 2) further guaranteed a more convincing conclusion that they shared the same origin of the genome. In spite of null similarity hits of MSL4 in the NCBI nucleotide database (https://www.ncbi.nlm.nih.gov/), we luckily found a homologous sequence of MSL4 in the *Eucomma* genome database (GWHAAAL00024598, OriSeqID=scaffold571_obj), Len=331,914, https://bigd.big.ac.cn/gwh/Assembly/13/show) with nucleotide identities being 98% (positions: 25,047-25,525). It is noteworthy that the individual *E. ulmoides* plant used for genome sequencing was also a male tree [20], which enables us to identify the homologous sequences of MSL4, supporting the conclusion we draw here that the MSL4 is a male-specific marker. Three putative genes have been annotated with unknown functions in this scaffold (#OriSeqID=scaffold571_obj, Accession=GWHAAAL00024598, GWHAAAL00000000.gff), among which the MSL4 is situated in the intron between CDS17 (24,544-24,670) and CDS18 (25,716-25,793) of the first gene (2,411-25,793). Functional studies of this genomics locus will facilitate to unveil the sex determination mechanism of *E. ulmoides* in the future.

Sex-specific markers have also been used to infer sex determination mechanisms in several species, e.g., kiwifruit [13, 39] and papaya [63]. In this study, we identified five candidate male-specific markers and no female-specific loci using the ddRAD-seq technique (Table 2). Further PCR and Sanger sequencing examination demonstrated male-specific sequences at one of the five loci (Figures 3 and 4). The development of a male-specific fragment in a vast portion of the population suggests that the sex of *E. ulmoides* is probably determined genetically. Noteworthy, we only detected male-specific ddRAD markers but no female-specific loci, similar to a recent study that only found a male-specific SSR marker (EST-Eu059) in *E. ulmoides* [57]. The 479 bp male-specific locus may imply its location on sex chromosome, which suggests that *E. ulmoides* likely has an XX/XY sex determination system. But, we should keep it in mind that additional sex-linked markers need to be explored to dissect the genetic architecture of sex determination in *E. ulmoides* in the future.

5. Conclusions

In this study, we performed ddRAD-seq to screen putative sex-linked markers in the hardy rubber tree, *E. ulmoides*, based on which we then identified five candidate male-specific loci while no female-specific loci. One of the five male-specific loci (MSL4) was further validated by PCR amplification and Sanger sequencing in a geographically distant population. As a result, the male-specific ddRAD marker is a reproducible, quick, and precise tool for discriminating males from female seedlings in *E. ulmoides*. The results also indicated that *E. ulmoides* probably has a genetic sex determination system. Collectively, reliable gender identification of *E. ulmoides* at the juvenile stage with the male-specific marker is not only valuable for marker-assisted breeding program but also helpful to ascertain optimal sex ratio in *E. ulmoides* orchard to maximum increase gutta (Eu-rubber) yield. The combination method of ddRAD and PCR also shed light on identifying sex-specific markers in other dioecious tree species.

Data Availability

The ddRAD-seq data generated and analyzed in this study have been deposited in the Sequence Read Archive (SRA) of the NCBI database, with the accession number PRJNA607161.
Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Authors’ Contributions

W-CW and X-ZZ conceived and designed the experiments. W-CW, G-QY, XD, and F-QS performed the experiments. W-CW, G-QY, and X-ZZ analyzed the data. Y-QL, WG, HL, and X-ZZ contributed reagents/materials/analysis tools. W-CW and X-ZZ wrote the paper. All authors reviewed and approved the final manuscript.

Acknowledgments

The authors are grateful to Dr. Chan Zhang (Henan Normal University) and Dr. Qi-Sheng Han (Farmland Irrigation Research Institute) for their kind help in sample collection. The authors acknowledge Dr. Qi-Sheng Han (Farmland Irrigation Research Institute) for their kind help in sample collection. The authors are grateful to Dr. Chan Zhang (Henan Normal University) and Dr. Qi-Sheng Han (Farmland Irrigation Research Institute) for their kind help in sample collection. The authors are grateful to Dr. Chan Zhang (Henan Normal University) and Dr. Qi-Sheng Han (Farmland Irrigation Research Institute) for their kind help in sample collection.

Supplementary Materials

Table S1: PCR primers used to amplify putative sex-linked dRAD loci in Eucommia ulmoides. (Supplementary Materials)

References


