A New Nuclear DNA Marker Revealing Both Microsatellite Variations and Single Nucleotide Polymorphic Loci: A Case Study on Classification of Cultivars in Lagerstroemia indica L

Zhi-li Suo1*, Wen-ying Li2, Xiao-bai Jin3 and Hui-jin Zhang3

1State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, 20 Nanxincun, Haidian District, Beijing, China
2Institute of Forestry New Technologies, Chinese Academy of Forestry, Yiheyuan Hou, Beijing, China
3Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, 20 Nanxincun, Haidian District, Beijing, China

Abstract

Plant cultivars are important germplasm resources for socio-economic development. However, it is difficult to conduct accurate genetic evaluation on plant diversity solely on the basis of morphological traits which are easily affected by environmental conditions and may change during developmental stages of the plant. Developing DNA markers with high resolution and sensitivity, especially at cultivar level, is a global challenge. We report a good methodology for rapid and efficient detection of plant genetic diversity at cultivar level. A unique nucleotide molecular formula (NMF) was constructed for each crape myrtle using polymorphic nucleotide sites. Our results showed that the DNA sequence from chromatin remodeling gene region of the ubiquitin-proteasome system is useful for molecularly characterizing crape myrtle cultivars. This DNA marker technique will be of significant value for plant breeding, cultivar identification, protection of Plant Breeders’ rights and evaluation and protection of plant germplasm resources.

Keywords: SBM-SNP marker; Cultivar; Genetic diversity; Assessment; Lagerstroemia indica L.

Introduction

Identification of cultivars is important for breeding, protection of Plant Breeders’ Rights, commercial production, trade and inspection of plants. However, plant genetic diversity is difficult to evaluate objectively and accurately if based solely on morphological traits which are easily changeable under different environmental conditions and in different developmental stages. Accurate and rapid identification of plant cultivars is still a big challenge. Development of DNA markers with high sensitivity is highly desirable [1-6].

There are three major classes of genetic variations in biological genomes, which are simple sequence repeat variations (SSRs), or microsatellite polymorphisms, single nucleotide polymorphisms (SNPs), and copy number variations (CNVs) [2].

Various DNA markers were designed with different strategies aiming at detecting the genetic variations at genomic DNA level. Along with the development of molecular biology, dozens of molecular marker techniques have been reported [1,2,7-12].

The available DNA fingerprinting techniques can be roughly divided into two categories: (i) Microsatellite information has been considered to be contained in the amplified DNA fragments when the DNA marker was designed, e.g., microsatellite analyses of simple sequence repeats (SSR) or inter simple sequence repeats (ISSRs) [13-15], variable number of tandem repeats, (VNTR) [1], Sequence tagged microsatellite markers (STMS) [16], selectively amplified microsatellite polymorphic loci (SAMPL) (modification of AFLP marker) [17], microsatellite-amplified fragment length polymorphism (M-AFLP) (modification of AFLP marker) [17], and retrotransposon-microsatellite amplified polymorphism (REMAP) [13]. (ii) Microsatellite information was not considered to be included necessarily in the amplified DNA fragments when the DNA marker was designed, e.g., restricted fragment length polymorphism (RFLP) [1], DNA random amplified polymorphic DNA (RAPD) [1,13], amplified fragment length polymorphism (AFLP) [1], single nucleotide polymorphisms (SNPs) [1,2]), the cleaved amplified polymorphic sequences (CAPS) [18], inter-retrotransposon amplified polymorphism (IRAP) [19], start codon targeted (SCoT) polymorphism [20], retro transposon based insertion polymorphism (RBIP) markers [21], SSCPs (single stranded conformation polymorphism) [1], sequence-specific amplified polymorphism (SSAP) [22], CAPS-SNP marker [20] and IRAP-SCAR marker [1]. The molecular marker systems are commonly used either singly or in combination for DNA fingerprinting and cultivar identification.

However, due to lack of complete genome sequence information in early time, the previously developed DNA fingerprinting methods for detecting DNA fragment length polymorphism were designed using various strategies to meet the need of plant genetic diversity detection with no prior knowledge of the genome sequence information as primers are designed. The electrophoresis images are to be converted into data matrix of 0/1 according to the existence or absence of the DNA fragments of certain length.

Compared to the direct use of nucleotide differences in the amplified DNA sequence by sequencing, DNA fingerprinting techniques are indirect detection methods of genetic variations. In practice, some disadvantages exist inevitably. For examples, in certain cases the microsatellite (especially dinucleotide) repeats based allelic variation results in shadow bands or stutter bands during electrophoresis thus

*Corresponding author: Zhi-li Suo, State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, 20 Nanxincun, Haidian District, Beijing, China.

Received May 24, 2016; Accepted June 19, 2016; Published June 29, 2016


Copyright: © 2016 Suo Zl, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
leading to genotyping errors [23,24]. Imperfect repeats and allelic dropout can lead to an overestimation of observable alleles, a decrease in observed heterozygosity, and an increase in the apparent level of inbreeding [2].

In recent years, the epoch-making booster of the global complete genome sequencing projects enabled us to use directly DNA sequences for plant. It can effectively overcome the disadvantages mentioned above in aspects of experimental stability and detection accuracy of the fragment length based DNA fingerprinting techniques. The currently recommended four DNA fragments/markers (rbcl, matK, trnH-psbA and ITS) as the sources of data can only meet the need for plant identification at/above species level with limited or no resolution among closely related species and/or cultivars [6,25,26], due to their lower evolutionary rates. Sufficient invest of time and money is necessary for exploration of more rapidly evolving DNA regions in the genome.

Recently, hyper-variable region of chloroplast genome has been used to reveal genetic differences of tree peony lineages/cultivars [5]. More rapidly evolving E3 ubiquitin ligase gene sequence has been found useful in species/variety/cultivar classification in *Juglans* L. [6].

Ubiquitin ligase genes are part of the genes related with the ubiquitin-proteasome system, which plays an important role in degradation of proteins, and is imperative for maintaining the cellular homeostasis in all eukaryotic cells [27-30]. However, there are no reports on the development of DNA markers from the genomic DNA regions in relation with the ubiquitin-proteasome system for crape myrtle cultivar identification so far.

Crape myrtles (*Lagerstroemia indica* L.) are famous ornamental trees/shrubs for blooms in summer and for city greening with a garden application history of more than 1700 years in China. There are over 500 cultivars in the world. In production and horticultural application, crape myrtle cultivars are divided into Rubra Group with purple, pinkish red, reddish purple, red, or similar-colored flowers, Alba Group with white or similar-colored flowers; Amabilis Group with purplish blue, bluish purple or similar-colored flowers; Sajin Group with multi-colored flowers[31-35]. However, it is difficult to conduct accurate genetic assessment on crape myrtle cultivar diversity solely on the basis of morphological traits which are easily affected by environmental conditions and developmental stages. The fact that more and more new cultivars are bred leads to an increasing need for accurate and rapid identification of cultivars. Crape myrtle flower industries are inhibited due to the shortage of morphological and DNA markers.

In this study, for the first time, we developed a DNA marker which detects simultaneously both microsatellite variations and single nucleotide mutations from the DNA region in relation with ubiquitin-proteasome system. It allows closely related crape myrtle cultivars to be discriminated unambiguously.

Our objectives were: (i) to investigate the possibility of developing new and high-performance DNA markers with advantages from both microsatellite marker and SNP marker; (ii) to evaluate the resolution ability and application value of such DNA marker; and (iii) to estimate the usability of the ubiquitin-proteasome system related DNA region for DNA marker development for crape myrtle cultivar identification. The results of this effort show that the SBMP-SNP (sequencing-based microsatellite polymorphism-single nucleotide polymorphism) marker derived from the ubiquitin-proteasome system related DNA region is sensitive for characterizing genetic diversity at cultivar level in *Lagerstroemia* L.

### Materials and Methods

#### Plant material

The crape myrtle cultivars used in this study were grown in Lianhe Crape Myrtle Resource Nursery (N 27°23′, E 111°47′) of the Hunan Crape Myrtle Institute, Hunan Ziwei Group Co., Ltd., Shaoyang City, Hunan Province, China. Fresh leaves of each cultivar were collected in spring and dried immediately using silica gel for DNA extraction. This is a joint project with the Hunan Crape Myrtle Institute, Hunan Ziwei Group Co. Ltd., Shaoyang City, Hunan Province that allowed us to investigate and collect materials.

#### DNA extraction, PCR amplification and sequencing

Total genomic DNA was extracted using the Plant Genomic DNA Kit (DP305) from Tiangen Biotech (Beijing) Co. Ltd., China. A DNA fragment was amplified using one of the newly developed primer pair CM_99F (5’–GTCCCCGTGAATGTTTGA–3’) and CM_878R (5’–GGTGCTTTGCAGCTAG–3’). Taq DNA polymerase and PCR buffer (TaKaRa Code: DR100B) were from TaKaRa Biotechnology Co. Ltd (Dalian, China). PCR amplification was conducted following the protocol of TaKaRa Code: DR100B. PCR conditions were as follows: preheating at 94°C for 4 min, 34 cycles at 94°C for 45 s, annealing at 57°C for 42 s and elongation at 72°C for 1.2 min., followed by a final extension at 72°C for 10 min. PCR amplification of the DNA regions of interest was performed in an Applied Biosystems VeritiTM 96-Well Thermal Cycler (Model#: 9902, made in Singapore). PCR products were sequenced with the same primer mentioned above from both forward and reverse directions using a 3730xl DNA analyzer (Applied Biosystems, Foster City, CA, USA). Sequencing reactions were 96°C for 1 min, 25 cycles at 96°C for 10 s, annealing at 50°C for 5 s and elongation at 60°C for 4 min, and then lowering at 4°C for storage, according to the protocol of BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA). Primer CM_99F_Nc_153F (5’–GTTGGAAGGAGGAAAACCC–3’) and primer CM_99F_Nc_410R (5’–TCTTTCTGGCCACTCAG–3’) were designed according to the sequence of the amplified DNA region obtained by primer pair CM_99F and CM_878R for further sequencing the amplified DNA fragment in order to obtain the high quality sequences. At least three independent individual plants of each cultivar in *Lagerstroemia* L. have been sequenced, including leaf samples from at least three different branches within an individual plant, each sample was sequenced for at least three times, identical results were obtained in the preliminary experiments [5-6]. Therefore, only one sample data was presented to represent each taxon in this paper.

#### Data analysis

The DNA sequences were aligned with ClustalX [36] and then were manually confirmed using Sequencher (+4.6) software and Mega 6.0 [37]. The sequences of the accessions were deposited in GenBank (GenBank accession numbers: KR162133-KR161230).

#### Results

##### Identification of the amplified DNA sequence

The 618 bp amplified DNA sequences (Figure S1) were found to be matched partially with the chromatin remodeling (*CHR*) involved gene (or synonym: *SNF2* domain-containing protein *CLASSY* 1-like gene) whose functions are related with the ubiquitin-proteasome system, as shown by BLAST to compare the sequence to that in the GenBank database. Additional comparisons were made using the related DNA
sequences of *Panica graminata* (c13592 transcribed RNA sequence, GenBank accession no. GBGR01013568.1), *Arabidopsis thaliana* (chromatin remodeling 38 mRNA, complete cds, NM_114134.2), *Eucalyptus grandis* (SNF2 domain-containing protein CLASSY 1-like (LOC104421037), transcript variant X4, mRNA, XM_010032860.1), *Brassica rapa* (SNF2 domain-containing protein CLASSY 1 (LOC103834007), transcript variant X4, mRNA, XM_009110066.1), and *Camellia sativa* (SNF2 domain-containing protein CLASSY 1-like (LOC104773713), mRNA, XM_010498358.1).

The BLAST search results suggested that the SBMP-SNP marker is localized within the genomic region with high homology to SNF2-like genes of the ubiquitin-proteasome system.

Molecular identification of the eight crape myrtle cultivars and the molecular taxonomic key

The full length of the amplified DNA fragments including primer sequences were 771–780 bp with slight variations among the cultivars investigated. By cutting off the sequences at the two ends of the amplified DNA fragments which are the same among cultivars, a 618 bp DNA sequence alignment corresponding to bases 24 to 641 in the entire sequence of the amplified DNA fragment from the 5’ end was used for cultivar identification (Tables 1 and 2 and Figure S1). The position number of each variable base site used in the formula was determined according to the newly generated 618 bp sequence alignment. The six polymorphic base sites used in the NMF of the cultivars for the genus

<table>
<thead>
<tr>
<th>No.</th>
<th>Cultivar</th>
<th>Group#</th>
<th>Petal Major color</th>
<th>Petal Edge color</th>
<th>Petal claw color</th>
<th>Inner surface color of calyx</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>‘Baoqing Zi’</td>
<td>Amabilis Group</td>
<td>Bluish purple</td>
<td>Bluish purple</td>
<td>A gradual transition of the claw color from reddish purple to purplish blue, occurred between the lower and the upper parts of the claw.</td>
<td>Colored ring absent. The inner surface color of calyx is in uneven, deep purplish red.</td>
</tr>
<tr>
<td>2</td>
<td>‘Caixia Mantian’</td>
<td>Rubra Group</td>
<td>Purplish pink</td>
<td>Purplish pink</td>
<td>Purplish red below the mid-claw, with a gradual transition to pinkish red upward</td>
<td>Color ring outline is unobvious. The inner surface color of calyx is in deep purplish red.</td>
</tr>
<tr>
<td>3</td>
<td>‘Cuipan Hongfen’</td>
<td>Rubra Group</td>
<td>Pinkish red</td>
<td>Pinkish red</td>
<td>Light yellow green below the mid-claw, with a gradual transition to light pink upward</td>
<td>Color ring area is in light yellow green of the background color or in uneven light pink</td>
</tr>
<tr>
<td>4</td>
<td>‘Hongzhao Shenzi’</td>
<td>Amabilis Group</td>
<td>Purplish red</td>
<td>Purplish red</td>
<td>Red below the mid-claw, with a gradual transition to reddish purple upward</td>
<td>Light purple ring, with an upward uneven diffusion of the color</td>
</tr>
<tr>
<td>5</td>
<td>‘Dahua Erba Nianhua’</td>
<td>Sajin Group</td>
<td>Pinkish white</td>
<td>Pinkish white</td>
<td>Light yellowish green below the mid-claw, with a gradual transition to shallow white color upward</td>
<td>With an obvious and discontinuous outline of a pinkish white ring</td>
</tr>
<tr>
<td>6</td>
<td>‘Duchua Ziyun’</td>
<td>Amabilis Group/</td>
<td>Light purplish blue</td>
<td>Light purplish blue</td>
<td>Purplish red of the lower part of the claw gradually become light purplish red upward</td>
<td>With an obvious deep purplish red ring, slight upward diffusion of deep purplish red from the base of the inner surface of calyx.</td>
</tr>
<tr>
<td>7</td>
<td>‘Furongmian’</td>
<td>Sajin Group</td>
<td>Light pink</td>
<td>Light pink</td>
<td>A gradual transition of the claw color from pinkish purple to light pinkish purple upward occurred between the lower and the upper parts of the claw</td>
<td>With a light pink ring</td>
</tr>
<tr>
<td>8</td>
<td>‘Xiaohua Yinwei’</td>
<td>Alba Group</td>
<td>White</td>
<td>White</td>
<td>A gradual transition of the claw color from purplish red to light purple, and then to white, occurred between the lower and the upper parts of the claw</td>
<td>No color ring was observed. There is uneven, purplish red on the inner surface of calyx.</td>
</tr>
</tbody>
</table>

Table 1: Major floral traits of the eight crape myrtle cultivars used in this study.

<table>
<thead>
<tr>
<th>Cultivar name</th>
<th>Positions of the variable base sites generated based on the 618 bp DNA sequence alignment (Figure: S1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>17</td>
</tr>
<tr>
<td>'Caixia Mantian'</td>
<td>G</td>
</tr>
<tr>
<td>'Baoqing Zi'</td>
<td>G</td>
</tr>
<tr>
<td>'Cuipan Hongfen'</td>
<td>A</td>
</tr>
<tr>
<td>'Dahua Erba Nianhua'</td>
<td>R</td>
</tr>
<tr>
<td>'Duchua Ziyun'</td>
<td>R</td>
</tr>
<tr>
<td>'Furongmian'</td>
<td>G</td>
</tr>
<tr>
<td>'Hongzhao Shenzi'</td>
<td>R</td>
</tr>
<tr>
<td>'Xiaohua Yinwei'</td>
<td>G</td>
</tr>
</tbody>
</table>

*, indicates the five key variable base sites and one simple sequence repeat locus used for construction of nucleotide molecular formulas. R=A/G, Y=T/C.

**K=GT, M=A/C, and W=A/IT Doe recorded respectively from A, G, or T are heterozygous base sites, with a double peak each in the chromatography.

**, indicates the base deletion in the sequence alignment. The position number of the polymorphic base site refers to the base position in the aligned sequence matrix from the 5’ end to the 3’ end. Bases 533 to 541 represent a simple sequence repeat locus with three repeats of GGG repeat unit.
Lagerstroemia are No. 17, 205, 214, 298, 347 and 533-541 (Table 2 and Figure S1). For instance, "SBMP-SNP_aln_618bp_" was used to refer to the DNA region employed in the NMF with "aln_618 bp" referring to the aligned sequence length (618 bp) of the eight representative crape myrtle cultivars (Figure 1). As a result, "SBMP-SNP_aln_618 bp_GC205 T214GC298TGGCGCGC233-541" can be constructed as an NMF for molecularly characterizing Lagerstroemia indica 'Caixia Mantian,' with the figure following the nucleotide character indicating the position of the corresponding polymorphic base site from the 5' end of the aligned sequence. The NMF can be constructed in a similar way for the rest of the samples of the crape myrtles. "SBMP-SNP_aln_618bp_" is omitted to save space in the description below. "Type T214", for example, in the following taxonomic key, refers to the cultivar with T214-typed base mutation, i.e., nucleotide T can be detected at base position 214 from the 5' end in the amplified DNA region. Other types of base mutation are indicated in the same way. As shown in Figure 2, a novel taxonomic key based on nucleotide molecular formulae is constructed by which the molecular feature of each cultivar is given.

Figure 1: Photos showing floral characters of the eight crape myrtle cultivars. 1a, 'Hongzhao Shenzi'; 1b, 'Cuipan Hongfen'; 1c, 'Caixia Mantian'; 1d, 'Baoqing Zi'; 1e, 'Furongmian'; 1f, 'Dahua Erba Nianhua'; 1g, 'Xiaohua Yinwei'; 1h, 'Duohua Ziyun.'
No genetic variation was detected in the sequence of the amplified DNA region among individual plants within any one cultivar, but large genetic variations were found among the crape myrtle cultivars at DNA level (Table 2, Figures 1 and 2 and Figure S1).

Discussion

Genetic uniformity within cultivar

Outstanding individual plants are initially selected from thousands of seedlings (or individual plants introduced from natural populations) according to the excellence/distinctness in aspects of floral characters and drought/water-logging/cold/heat resistance. Then, they are propagated asexually by cuttings or division in order to obtain several dozens of individuals for further observations for years. After passing a test of Distinctness, Uniformity and Stability (DUS) of the plants, they will be given a formal cultivar name for utilization and conservation. Our results are identical to the actual situation that no genetic variation exists among individual plants within each cultivar as far as the sequence of this amplified DNA region is concerned.

Genetic variation among cultivars

In crape myrtle, the 618 bp SBM-SNP marker developed in this study successfully discriminated eight crape myrtle cultivars that are highly similar in morphology, using only 35.29% of the total seventy variable sites. It is worthy of further study. As the ubiquitin-proteasome system may provide different resolutions in detection of the smallest genetic variation unit with good stability and high sensitivity. This study presents a good complementary methodology for deeper understanding of plant genetic diversities. It will be of significant value for plant breeding, protection of Plant Breeders’ Rights, and evaluation, protection and utilization of plant germplasm resources.

Conclusion

Direct use of DNA sequence as DNA marker can effectively overcome the disadvantages in aspects of experimental stability and detection. Our study further confirmed that highly informative SSRs and SNPs are the most sought key molecular features in DNA identification of plant germplasm characterization. The SBM-SNP marker utilized more rapidly evolving DNA regions and detected the smallest genetic variation unit with good stability and high sensitivity. This study presents a good complementary methodology for deeper understanding of plant genetic diversities. It will be of significant value for plant breeding, protection of Plant Breeders’ Rights, and evaluation, protection and utilization of plant germplasm resources.

Acknowledgement

This study was financially supported by the National Natural Science Foundation of China (Grant No. 30972412), by the program of Development of DNA Molecular Markers and Genetic Analysis of Crape Myrtle Cultivars (70009C1036), by the joint project no. 70009C1036 and by the National Forest Genetic Resources Platform (2005DKA21003). Jianmin Huang, Boxin Hou and Qiangfa Yang identified cultivars using morphological traits and made discussions during the study.

References
