Development of 19 universal microsatellite loci for three closely related *Ficus* species (Moraceae) by high-throughput sequencing

Shuqiong Li¹,², Jianfeng Huang¹*, Clive T. Darwell³ and Yanqiong Peng¹†

¹CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming 650223, China
²University of Chinese Academy of Sciences, Beijing 100049, China
³Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa 904-0412, Japan

(Received 4 September 2019, accepted 2 December 2019; J-STAGE Advance published date: 12 March 2020)

The genus *Ficus* is striking for its species diversity, ecological significance, and its often species-specific relationship with coevolved pollinating fig wasps, which has long fascinated biologists. The three closely related and generally co-distributed dioecious species *Ficus hispida*, *F. heterostyla* and *F. squamosa* provide an ideal system for the study of speciation, hybridization (caused by pollinator sharing) and comparative phylogeography to infer historical biogeography. We aimed to develop microsatellite markers for these allied species to facilitate the outlined study investigations. A DNA library was constructed from one *F. heterostyla* sample, and 19 microsatellite loci were developed based on high-throughput sequencing. These markers showed relatively high polymorphism in all three fig species. The mean number of alleles per locus was 3.594–5.286, and the mean observed and expected heterozygosity were 0.469–0.546 and 0.467–0.528, respectively. Principal coordinate, STRUCTURE and AMOVA analyses revealed different degrees of genetic differentiation within species, and, despite some observed genetic admixture, indicated the presence of clear boundaries between different species. In summary, we successfully developed universal microsatellite markers for three closely related *Ficus* species. These markers will be of great value for investigating patterns of biodiversity among the species in this model system for coevolutionary studies.

**Key words:** *Ficus*, fig, microsatellite loci, polymorphism

INTRODUCTION

With approximately 800 species, *Ficus* (Moraceae) is arguably the most diverse woody plant genus and occurs globally in the tropics and subtropics. *Ficus* are viewed as keystone species within tropical forest ecosystems due to their provisioning of fig resources for frugivores (birds, mammals, etc.) (Herre et al., 2008). All *Ficus* species are involved in an obligate pollination mutualism with mostly species-specific pollinating fig wasps (Hymenoptera, Chalcidoidea, Agaonidae), and the incumbent coevolutionary dynamics between *Ficus* and agaonids are considered to be the main driver of fig diversity (Cruaud et al., 2012). However, numerous cases of host shift and pollinator sharing have been reported (Haine et al., 2006; Cornille et al., 2012; Liu et al., 2015), potentially leading to host fig hybridization, which has been posited as an additional mechanism generating fig diversity (Machado et al., 2005).

*Ficus hispida* forms the outgroup lineage to the sister species *F. heterostyla* and *F. squamosa* (Cruaud et al., 2012). Despite close phylogenetic relationships, they display differences in phenotype or habitat preference. *Ficus hispida* is a pioneer shrub or small tree, widely distributed over tropical Asia and Australasia. It is mainly found along roads or secondary habitats as a pioneer species. Berg and Chantarasuwan (2007) proposed *F. heterostyla* as distinct from *F. hispida* because of differences in fruiting position and hair color on leaves and stems. *Ficus heterostyla* grows under the forest canopy or along roadsides (Berg and Chantarasuwan,
However, *F. squamosa* is a short dioecious shrub (1–3 m tall) with a creeping stem, typically found growing along riverbanks or near fast-flowing streams in tropical forests (Pothasin et al., 2016) and occurring from Nepal to Thailand. These three species co-occur in Yunnan (China) and the northern part of Mainland Southeast Asia. The complementary fruiting phenologies between *F. heterostyla* and *F. squamosa* facilitate the sharing of one pollinating fig wasp by these two figs, which may result in hybridization. Furthermore, experimental hybrid seeds have been generated (Liu et al., 2015). Therefore, *F. hispida*, *F. heterostyla* and *F. squamosa* comprise an ideal system to study species diversification, the maintenance of species boundaries, hybridization and comparative phylogeography.

Due to the usefulness of estimating parameters of population genetic diversity, identifying individuals or populations, and investigating phylogeographic patterns (Wei et al., 2014; Zhao et al., 2015), developing microsatellite loci for these three species is an appropriate approach for the studies mentioned above. Although Dev et al. (2011) published eight microsatellite loci for *F. hispida*, they often fail to display polymorphisms among these species and the number of markers is insufficient for high-resolution studies. Thus, there is a need for the development of additional microsatellite markers. In the current work, we aimed to develop a set of universal polymorphic microsatellite primers for these three *Ficus* species using high-throughput sequencing. These new primers will enable us to evaluate population dynamics and connectivity of these three species and will facilitate further comparative phylogeographic studies.

**MATERIALS AND METHODS**

**Isolation of microsatellite loci** Genomic DNA was extracted from leaves of all three species collected from six natural populations in Southwestern China and Southeast Asia (Fig. 1A), using the Tiangen Plant Genomic DNA Kit (Tiangen Biotech, Beijing, China). The DNA was quantified using a NAS-99 spectrophotometer (ATCGene, Piscataway, NJ, USA) and we followed the manufacturer’s protocols for the KAPA HyperPrep kit (KAPA Biosystems, Wilmington, MA, USA) to construct a DNA library using one sample of *F. heterostyla*. The sequencing processes were performed by Novogene (Beijing, China).
on a HiSeq 2500 (Illumina, San Diego, CA, USA), using reagents from Illumina and with paired read lengths of 2 × 150 bp. All raw data were quality-filtered by trimming adapter sequences and by removing reads with quality scores < 20. Finally, high-quality cleaned paired reads were aligned by FLASH (Magoc and Salzberg, 2011) and we obtained 10,952,088 tags that were deposited in the NCBI Sequence Read Archive (BioProject ID PRJNA542260; accession no. SRP197518). The MICRO-SATELLITE (MISA) identification tool (Thiel et al., 2003) was adopted to identify reads with microsatellite loci, and Primer 3.0 (Rozen and Skaletsky, 1999) was used to design microsatellite primers using default settings.

To develop universal polymorphic microsatellite primers for the three species, a two-step PCR amplification approach was employed. The first PCR amplification was tested on six samples randomly selected from the six populations to initially screen for primers. An M13 tag (5′-TGTAAACGACGCGCCAGT-3′) was added to the 5′ end of the primer of each locus. This amplification step used the following protocol and conditions: 15 μl of PCR amplification reaction volume containing 1.5 μl of 10×Buffer, 1.2 μl of dNTP (2.5 mM), 1.2 μl of DNA, 0.8 μl of TP-M13 (5′ M), 1 μl of specific primer, 0.1 μl of Takara Taq HS polymerase and 9.2 μl of deionized water; initial denaturation at 95 °C for 5 min, 30 cycles at 94 °C for 30 s, at 56 °C for 30 s and at 72 °C for 30 s, 10 cycles at 94 °C for 30 s, at 53 °C for 30 s and at 72 °C for 30 s, and final extension at 60 °C for 30 min. Microsatellite loci that were successfully amplified and displayed polymorphisms in all three species were selected for the next PCR amplification stage. For this, we used multiplex PCR and focused on the applicability of the filtered primers in exploring genetic diversity within populations. In this step, a total of 95 samples from the six populations of the three species were analyzed using 5′ fluorophore primers labeled with 6-FAM, HEX, ROX or TMR instead of M13-tailed primers. PCR amplification was based on the following protocol and conditions: 10 μl of PCR amplification reaction volume containing 2.0 μl of 5 × Buffer, 0.3 μl of dNTP (10 mM), 1 μl of DNA, 0.4 μl of primer mixture (5 μM), 0.1 μl of Taq polymerase, and 6.2 μl of deionized water; initial denaturation at 95 °C for 8 min, 30 cycles at 94 °C for 30 s, at 58 °C for 40 s and at 72 °C for 40 s, and final extension at 60 °C for 30 min. All PCR products were analyzed by capillary electrophoresis on an ABI 3730XL DNA analyzer (Applied Biosystems, Foster City, CA, USA) with the GeneScan 500 ROX Size Standard. Microsatellite fragment sizes were determined using GeneMapper version 3.2.

**Data analysis** First, we checked null alleles using MICRO-CHECKER 2.2 (van Oosterhout et al., 2004), employing 95% confidence intervals. The population genetic parameters of diversity, comprising the number of alleles per locus (A), the observed heterozygosity (Hₒ), the expected heterozygosity (Hₑ) and the inbreeding coefficient (Fₛ), were calculated by GenAlEx version 6.5 (Peakall and Smouse, 2012). We tested for linkage disequilibrium and Hardy–Weinberg equilibrium in the three species using GENEPOP 4.7.0 (Rousset, 2008) and used Bonferroni correction (at P < 0.05) with default parameters in MYRIADS 1.1 (Carvajal-Rodriguez and de Uña-Alvarez, 2011). We then tested for population differentiation within individual species using analysis of molecular variance (AMOVA) in ARLEQUIN 3.5 (Excoffier and Lischer, 2010) with 1,000 permutations to test for significance. STRUCTURE 2.3.4 (Pritchard et al., 2000) was used to assign individuals into clusters of different numbers of demes (K; at values between 1–6) with membership coefficients estimated under an admixture model with correlated allele frequencies. We set 10 independent runs for each value of K and 100,000 MCMC repetitions with a burn-in of 100,000 iterations. We used the HARVESTER online program (Earl and vonHoldt, 2012) to plot the values of Delta K and variance per K value, and to determine the optimal K value (Evanno et al., 2005). For the optimal value of K, CLUMPP 1.1.2 (Jakobsson and Rosenberg, 2007) was used to summarize the membership coefficients into clusters, after which the program DISTRACT 1.1 (Rosenberg, 2004) was launched to draw the results as colored box-plots. Results were also plotted on a distribution map. Finally, a principal coordinate analysis (PCoA) was carried out in GenAlEx version 6.5 based on a pairwise, individual-by-individual dominant genotypic distance matrix.

**RESULTS AND DISCUSSION**

MISA found 9,265 contigs containing at least one microsatellite, along with a total of 9,616 microsatellite loci. Among these loci, 2-, 3-, 4- and 5-nucleotide motifs were found at 90.34%, 8.53%, 0.83% and 0.3%, respectively. A total of 14,505 primer pairs were recovered, and 158 primer pairs featuring different repetitions were then selected randomly for subsequent PCR tests. Across the 158 selected primer pairs, 34 microsatellite loci were successfully amplified from the first six samples of the three species, with 19 loci displaying relatively high polymorphism along with appropriate fragment size in the pretest experiment (Table 1). For these 19 applicable loci, we tested polymorphism with the three species. Several loci displayed signals of null alleles (6-N104, 19-N530, 11-N346, 16-N106 and 27-N150 in *F. hispida*; 12-N102, 32-N125, 10-N381, 16-N106, 27-N150 and 29-N105 in *F. heterostyla*; 21-N197, 6-N104 and 29-N105 in *F. squamosa*). Over populations, per locus, the A, Hₒ, Hₑ and Fₛ ranged from 1 to 10, 0 to 0.871, 0 to 0.830, and −0.233 to 0.912, respectively (Table 2). The relatively high Fₛ values that were shown for some loci likely result...
from the presence of null alleles (Table 2). After correcting, we found a total of seven locus pairs (12-N102 vs. 21-N197, 3-N173 vs. 7-N245 and 30-N457 in *F. hispida*; 6-N104 vs. 21-N197, 7-N245 vs. 26-N180, 5-N108 vs. 11-N346 and 21-N197 in *F. heterostyla*) that significantly deviated from linkage equilibrium, but no loci that significantly deviated from Hardy–Weinberg equilibrium (*P* < 0.05), across all species. AMOVA tests revealed a much higher population differentiation in *F. squamosa* (*F*<sub>ST</sub> = 0.43, *P* < 0.01) than in *F. hispida* (*F*<sub>ST</sub> = 0.071, *P* < 0.01).

Table 1. Characteristics of 19 microsatellite loci developed for *Ficus hispida*, *F. heterostyla* and *F. squamosa*

<table>
<thead>
<tr>
<th>Locus</th>
<th>Primer sequences (5′–3′)</th>
<th>Repeat motif</th>
<th>Allele size range (bp)</th>
<th><em>T</em>&lt;sub&gt;a&lt;/sub&gt; (°C)</th>
<th>Fluorescent label</th>
<th>GenBank accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-N173</td>
<td>F: AATGTCCTTCGGCTCTCAGA R: GGGTGCACACTTTTCCTATTG</td>
<td>(GA)&lt;sub&gt;30&lt;/sub&gt;</td>
<td>100–137</td>
<td>60.1</td>
<td>HEX</td>
<td>MK937077</td>
</tr>
<tr>
<td>4-N101</td>
<td>F: AAAATGTCGGGATGTATTCAAGA R: TTTCCGCCCATATCCTATTTCAC</td>
<td>(AC)&lt;sub&gt;27&lt;/sub&gt;</td>
<td>72–155</td>
<td>59.8</td>
<td>ROX</td>
<td>MK937078</td>
</tr>
<tr>
<td>5-N108</td>
<td>F: CCGAAGACATTTAACCACAA R: AGTACGGTTTCCTTTTCTTGTC</td>
<td>(TG)&lt;sub&gt;23&lt;/sub&gt;</td>
<td>128–177</td>
<td>60.7</td>
<td>HEX</td>
<td>MK937097</td>
</tr>
<tr>
<td>6-N104</td>
<td>F: AAGCAGGCTTGTTGGGTAACAT R: GCACACCGCAGAGGAAGTA</td>
<td>(TC)&lt;sub&gt;26&lt;/sub&gt;</td>
<td>126–164</td>
<td>59.7</td>
<td>HEX</td>
<td>MK937080</td>
</tr>
<tr>
<td>7-N245</td>
<td>F: CACAAAGGCTCTTGGTTCTTTC</td>
<td>(TC)&lt;sub&gt;30&lt;/sub&gt;</td>
<td>125–170</td>
<td>60.4</td>
<td>ROX</td>
<td>MK937081</td>
</tr>
<tr>
<td>10-N381</td>
<td>F: GTCCGAGGAGTATTTCCAGTG R: CCGACACAAACTCAGCACCATA</td>
<td>(GA)&lt;sub&gt;35&lt;/sub&gt;</td>
<td>120–161</td>
<td>59.8</td>
<td>FAM</td>
<td>MK937082</td>
</tr>
<tr>
<td>11-N346</td>
<td>F: GGGAAATAGGGCCGTTAGGA R: GGGCTTTTTTGTTCTCTAGAT</td>
<td>(AG)&lt;sub&gt;39&lt;/sub&gt;</td>
<td>104–155</td>
<td>61.5</td>
<td>HEX</td>
<td>MK937083</td>
</tr>
<tr>
<td>12-N102</td>
<td>F: TCTCAGAATCGAATCGAAGC</td>
<td>(TG)&lt;sub&gt;25&lt;/sub&gt;</td>
<td>136–308</td>
<td>59.0</td>
<td>FAM</td>
<td>MK937084</td>
</tr>
<tr>
<td>16-N106</td>
<td>F: GGAAAAACACAGATAGAGGCA R: GTGTTAAACACCTAGGCGCAGGA</td>
<td>(CT)&lt;sub&gt;33&lt;/sub&gt;</td>
<td>142–230</td>
<td>60.7</td>
<td>FAM</td>
<td>MK937085</td>
</tr>
<tr>
<td>19-N530</td>
<td>F: CAGTGGCACAAGATTTCTCCT</td>
<td>(GA)&lt;sub&gt;39&lt;/sub&gt;</td>
<td>163–178</td>
<td>58.8</td>
<td>TMR</td>
<td>MK937086</td>
</tr>
<tr>
<td>20-N291</td>
<td>F: AATCTGATTCACCGGCTCAGA R: CAGCACCATCGCACCAGGAA</td>
<td>(GA)&lt;sub&gt;38&lt;/sub&gt;</td>
<td>148–210</td>
<td>61.2</td>
<td>TMR</td>
<td>MK937087</td>
</tr>
<tr>
<td>21-N197</td>
<td>F: GGGGAGGATCAGGAGGAGGACG R: CTCATATCGCCCAACGCAT</td>
<td>(TC)&lt;sub&gt;31&lt;/sub&gt;</td>
<td>165–225</td>
<td>59.7</td>
<td>FAM</td>
<td>MK937088</td>
</tr>
<tr>
<td>23-N724</td>
<td>F: AGGGTGAGGCTGTTTTCGCCA</td>
<td>(TCT)&lt;sub&gt;32&lt;/sub&gt;</td>
<td>167–191</td>
<td>59.9</td>
<td>TMR</td>
<td>MK937089</td>
</tr>
<tr>
<td>26-N180</td>
<td>F: TGTTCAGTATCCATCTTGCGGAAG R: AATACGGTGACCGAGCGAAG</td>
<td>(TTC)&lt;sub&gt;15&lt;/sub&gt;</td>
<td>133–168</td>
<td>59.9</td>
<td>ROX</td>
<td>MK937090</td>
</tr>
<tr>
<td>27-N150</td>
<td>F: CATTTTCGCGTTGCTTGTGATGT R: GIAACGACACCATATACCAACACT</td>
<td>(GA)&lt;sub&gt;30&lt;/sub&gt;</td>
<td>141–161</td>
<td>60.0</td>
<td>HEX</td>
<td>MK937091</td>
</tr>
<tr>
<td>28-N247</td>
<td>F: GCTGCAAAAGCATAACAAACGC</td>
<td>(CA)&lt;sub&gt;34&lt;/sub&gt;</td>
<td>150–182</td>
<td>59.6</td>
<td>FAM</td>
<td>MK937092</td>
</tr>
<tr>
<td>29-N105</td>
<td>F: GAAGCTAAATCCACCCAAGCA R: CTCTTCCTTTGATCGTCCTCAG</td>
<td>(TC)&lt;sub&gt;19&lt;/sub&gt;</td>
<td>161–230</td>
<td>60.4</td>
<td>TMR</td>
<td>MK937093</td>
</tr>
<tr>
<td>30-N457</td>
<td>F: AGGAAGGCGAAAGCTCAGATTC R: TCCCTCTTCTCCCTCTCTC</td>
<td>(CAA)&lt;sub&gt;17&lt;/sub&gt;</td>
<td>179–226</td>
<td>61.5</td>
<td>FAM</td>
<td>MK937094</td>
</tr>
<tr>
<td>32-N125</td>
<td>F: GCAGCCTGACCGACGACCTGA R: TAGGCAGCAATACCTCATCT</td>
<td>(TG)&lt;sub&gt;22&lt;/sub&gt;</td>
<td>166–203</td>
<td>59.3</td>
<td>TMR</td>
<td>MK937095</td>
</tr>
</tbody>
</table>

*T*<sub>a</sub>, annealing temperature.
and *F. heterostyla* (*F*$_{ST}$ = 0.079, *P* < 0.01). STRUCTURE results indicated an optimal K value of three among three species (Fig. 1B, 1C). PCoA showed patterns consistent with those of STRUCTURE (Figs. 1C and 2), which indicated the presence of clear boundaries between the three species. STRUCTURE results also displayed some signatures of admixture between paired species (Fig. 1A, 1C), indicating interspecific gene flow.

Pollinator sharing and interspecific hybridization caused by pollinator sharing have been documented repeatedly for *Ficus* species (e.g., Machado et al., 2005; Liu et al., 2015). Complementary flowering phenologies between *F. heterostyla* and *F. squamosa* have facilitated host shift of the pollinator, which may cause interspecific hybridization. Hybrid seeds have been experimentally generated between these two figs, indicating the potential for introgression (Liu et al., 2015). Additionally, there is phylogenetic evidence of a high degree of pollinator sharing based on ultra-conserved elements (unpublished results) from pollinators of these three figs. Thus, the presence of weak gene flow among them indicated in the present study is not unexpected. However, our discovery that two *F. heterostyla* individuals of population “A” showed identical genetic constitution with *F. hispida*...
should be explored deeply in future (Fig. 1C), including by enhanced sampling and morphological identification. Furthermore, more detailed studies are required to establish the exact genetic patterns of species interaction, mechanisms of introgression between different pairs of species, and individual species’ evolutionary histories.

Within species, *F. squamosa* shows the highest genetic differentiation, followed by *F. heterostyla* and *F. hispida*. Pollen dispersal of dioecious figs is often restricted, which is likely a result of high population densities, fruiting patterns, and growth forms, since high densities of receptive host trees will nullify the requirement for long-distance dispersal of tiny pollinating wasps (Harrison, 2003; Wang et al., 2009; Chen et al., 2011). Among our studied dioecious figs, *F. squamosa* shows higher local abundances and is found near streams or rivers, which could account for the high population differentiation if water courses act as barriers to dispersal of pollinators. Despite the shorter inter-population sampling distance (*F. heterostyla*: 215 km; *F. hispida*: 661 km), *F. heterostyla* shows slightly higher differentiation than *F. hispida*. The figs of *F. heterostyla* are located in rooting stolons near or under the soil (Liu et al., 2015), which may affect pollinator dispersal and thus have resulted in increased genetic differentiation. This pattern was also documented for *F. tikoua* (Chen et al., 2011).

We successfully developed 19 polymorphic microsatellite loci for the target species. These new microsatellite markers will be highly valuable for future studies on species diversification, the maintenance of species boundaries, hybridization and comparative phylogeography.

The authors are grateful to X. X. Song for technical assistance, and to D. R. Yang, B. G. Miao and P. Yang for sample collection. This research was funded by the National Natural Science Foundation of China (31800313, 31672373), the Light of West China Program of the Chinese Academy of Sciences, and the Program of the Ministry of Science and Technology of China (2018FY100406).

REFERENCES


Microsatellite loci for three *Ficus* species

Mol. Biol. 6, 365–386.


