Genetic diversity of *Begonia versicolor* (Begoniaceae), a narrow endemic species in southeast Yunnan of China

Chun Li\(^1\)\(^2\), Jun-Lin CHEN\(^3\), Yong-Yong LI\(^1\)\(^4\) and Dai-Ke TIAN\(^1\)\(^2\)

1. Shanghai Chenshan Plant Science Research Center, Chinese Academy of Sciences, 3888 Chenhua Road, Shanghai, China.
2. Shanghai Key Laboratory of Plant Functional Genomics and Resources, Shanghai Chenshan Botanical Garden, 3888, Chenhua Road, Shanghai, China.
3. College of Humanities, Sichuan Agricultural University, 46 Xinkang Road, Ya'an, Sichuan, China.
4. School of Humanities and Social Sciences, Peking Union Medical College, Beijing, China.

* Corresponding author’s email: dktian@sibs.ac.cn; † These authors contributed equally.

(Manuscript received 26 September 2017; accepted 2 March 2018; online published 12 March 2018)

**ABSTRACT**: *Begonia versicolor* Irmscher, a narrow endemic *Begonia* species in southeast Yunnan of China, is a wonderful ornamental plant with huge diversity in colored foliage. To investigate its variations, the genetic diversity and population structure were studied based on 56 individuals sampled from four localities using 12 polymorphic microsatellite loci transferred from other species of *Begonia*. The results showed a relatively low level of genetic diversity in *B. versicolor* comparing with other species of *Begonia* using microsatellite. Positive inbreeding coefficient (F\(_{IS}\)) values were found in three populations (SWC, XPZ and DSD). AMOVA analysis indicated that genetic variations occurred mainly within populations (55.9%) rather than among populations (9.7%) and among groups (34.4%). Four populations were grouped into two clusters based on STRUCTURE. AMOVA and STRUCTURE analysis showed a high level and significant genetic differentiation in the populations of *B. versicolor*. Based on its genetic status and rarity in the wild, the sustainable *in-situ* and *ex-situ* conservation strategies should be urgently carried out to protect this species with high horticultural and scientific values.

**KEY WORDS**: *Begonia versicolor*, Conservation strategies, Genetics diversity, Population structure, Yunnan.

**INTRODUCTION**

*Begonia* L., containing more than 1800 species, is the sixth largest genus of vascular plants (Attawade and Yadav, 2012; Hoover et al., 2004; Thomas, 2010; Kiew et al., 2015). Astonishing diversity makes it a genus with enormous horticultural appeal for flamboyant and colorful leaves and spectacular blooms (Tebbitt, 2005). The first living plant in *Begonia* was introduced to Europe during the eighteenth century, and thereafter over 400 natural species have been introduced for horticulture and plenty of hybrid cultivars have been bred (Tebbitt, 2005). However, a majority of wild begonias with high ornamental value have yet to be utilized.

*Begonia versicolor* Irmscher is a wonderful foliage plant for its extremely high variation in leaf colors, but only a few of its resource have been introduced for ornamental use. This species is endemic to the southeast of Yunnan province in China, and grows in a shady moist environment under forest, on slope or along stream bank at the altitudes of 1800-2100 m (Gu et al., 2007). *Begonia versicolor* was named for its beautiful foliage featured by diverse colors and spot patterns on leaf even within a very tiny population. However, as an endemic and rare species with high horticultural value, little is known about its population genetic diversity, which is fundamental for conservation strategies and resources development (Rao and Hodgkin, 2002; Schaal et al., 1991).

Simple sequence repeats (SSR) or microsatellites, the most popular molecular marker in study of population genetics, are constituted by tandemly repeated motifs of 1-6 bases, and widely exist in coding and non-coding region of all prokaryotic and eukaryotic genomes (Morgante and Olivieri, 1993; Zane et al., 2002). High levels of allelic diversity, co-dominant inheritance and widespread existence in genomes make microsatellite a powerful tool in genetic research (Gupta and Varshney, 2000; Peakall et al., 1998). Several studies have been published on developing polymorphic microsatellites markers of genus *Begonia* (Chap et al., 2014; Hughes et al., 2002a; Hughes et al., 2002b; Nakamura et al., 2012; Twyford et al., 2013; Tseng et al., 2017; Wiesner and Wiesnerova, 2008). To evaluate the population genetic diversity of *B. versicolor*, 136 Expressed Sequence Tags SSR (EST-SSR) markers of *B. plebeja* developed by Twyford et al. (2013) were screened. A suggestion on conservation and use of this species was provided based on genetic analysis.

**MATERIALS AND METHODS**

**Sample collection and DNA extraction**

Four populations, totaling 56 samples, were collected in September 2013 at southeast Yunnan of China (Fig. 1 and Table 1). The voucher specimens of each population were stored at the Herbarium of
Taiwania
Vol. 63, No. 1

Table 1. Estimates of population genetic variability of *B. versicolor* in Yunnan province of China based on 12 microsatellite loci

<table>
<thead>
<tr>
<th>Pop.</th>
<th>Voucher</th>
<th>Locality</th>
<th>N</th>
<th>N_a</th>
<th>A_p</th>
<th>H_o</th>
<th>H_e</th>
<th>F_is</th>
<th>A_r</th>
<th>G_d</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWC</td>
<td>TDK-1287</td>
<td>Shuiweicheng, Dawei Mountain</td>
<td>22</td>
<td>3.083</td>
<td>6</td>
<td>0.384</td>
<td>0.390</td>
<td>0.039</td>
<td>2.553</td>
<td>0.400</td>
</tr>
<tr>
<td>JCW</td>
<td>TDK-1298</td>
<td>Jiuchen Waterfall, Dawei Mountain</td>
<td>7</td>
<td>2.583</td>
<td>0</td>
<td>0.440</td>
<td>0.386</td>
<td>-0.068</td>
<td>2.583</td>
<td>0.414</td>
</tr>
<tr>
<td>XZP</td>
<td>TDK-1345</td>
<td>Xiaopingzai, Malipo</td>
<td>18</td>
<td>3.083</td>
<td>4</td>
<td>0.322</td>
<td>0.391</td>
<td>0.207</td>
<td>2.490</td>
<td>0.406</td>
</tr>
<tr>
<td>DSD</td>
<td>TDK-1375</td>
<td>Doushidian, Malipo</td>
<td>9</td>
<td>2.25</td>
<td>1</td>
<td>0.222</td>
<td>0.241</td>
<td>0.137</td>
<td>2.095</td>
<td>0.258</td>
</tr>
</tbody>
</table>

Notes: N: population size; N_a: number of alleles; A_p: number of private alleles; H_o: observed heterozygosity; H_e: expected heterozygosity; F_is: inbreeding coefficient; A_r: allelic richness; G_d: gene diversity.

Shanghai Chenshan Botanical Garden (CSH). All samples in the same population were collected at least 1 m apart from each other to reduce the chance of sampling clones (Hughes and Hollingsworth, 2008). Fresh leaf material of each sample was collected and stored dried with silica gel. Total genomic DNA was extracted from ca. 20 mg silica-dried leaves using DNAsecure Plant Kit (Tiangen Biotech, Beijing, China) according to the manufacturer’s protocols.

**Microsatellite genotyping**

One hundred and thirty-six EST-SSR markers developed by Twyford *et al.* (2013) were assessed in a pilot experiment with 8 samples (two samples each population). Finally, 12 EST-SSR markers with clear and polymorphic bands were chosen to genotype all samples. The method of amplification reactions was adopted from Li *et al.* (2015).

**Data Analysis**

Microchecker was used to detect genotyping errors resulting from stutterting, short allele dominance and null alleles (Van Oosterhout *et al.*, 2004). Number of alleles (N_a), number of private alleles (N_p), allelic richness (A_r), gene diversity (G_d), observed heterozygosity (H_o), expected heterozygosity (H_e), genetic differentiation (F_st) and inbreeding coefficient (F_is) were calculated by using GenAlEx (Peakall and Smouse, 2006) and Fstat (Goudet, 2001).

Genetic structure was investigated with a model-based Bayesian clustering method implemented in STRUCTURE (Pritchard *et al.*, 2000). The admixture model, the option of correlated allele frequencies between populations and 20 times for each K value from 1 to 4 were implemented in this program with each run comprising a burn-in period of 10 000 iterations and a length of 30 000 iterations. STRUCTURE HARVESTER (Earl and vonHoldt, 2012) was used to estimate the number of genetic clusters (K) based on the method of Evanno *et al.* (2005). Analysis of molecular variance (AMOVA) was performed with Arlequin (Excoffier and Lischer, 2010) to evaluate the partition of genetic variation within and among populations, and among groups based on the result of genetic structure. The effective population size was estimated by the program LDNe, which measures linkage disequilibrium at presumably unlinked loci (Waples and Do 2008).

**RESULTS**

**Genetic variation across populations**

Observed and expected heterozygosities per population ranged from 0.222 to 0.440 and 0.241 to 0.391, respectively. The JCW population had the greatest allelic richness (2.583), while DSD showed the least (2.095). The XZP population showed the most gene diversity (0.414) and DSD had the least (0.258). Inbreeding coefficient (F_is) among populations varied from -0.065 to 0.207. Three populations (DSD, XZP and SWC) had private alleles ranging from 1 (DSD) to 6 (SWC). Detailed information was showed in Table 1.

**Genetic variation across loci**

Totally, 53 alleles were detected in 56 individuals from four populations using 12 SSR markers, ranging from 3 (B07, B15 and B37) to 7 (B80) with an average number of 4.417. The mean observed heterozygosity (H_o) was 0.346 (range: 0.018-0.600) and the expected heterozygosity (H_e) ranged from 0.149 to 0.707 (average 0.542). The mean allelic richness (A_r) was 4.387 (range: 2.929-6.997) and mean gene diversity (G_d) per locus was 0.549 (range: 0.152-0.716). Inbreeding coefficient (F_is) ranged from -0.116 to...
Fig. 2. Results of Bayesian analysis of 56 individuals from 4 populations of *B. versicolor* in Yunnan province of China based on 12 microsatellite loci using STRUCTURE. Different colors representation of the different genetic pools or clusters for *K* = 2. Populations are separated by vertical bars.

Table 2. Genetic diversity for 12 microsatellite loci in 4 populations of *B. versicolor* in Yunnan province of China

<table>
<thead>
<tr>
<th>Locus</th>
<th><em>N</em></th>
<th><em>H</em></th>
<th><em>H</em></th>
<th><em>F</em></th>
<th><em>A</em></th>
<th><em>G</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>B07</td>
<td>3</td>
<td>0.018</td>
<td>0.149</td>
<td>0.882</td>
<td>2.929</td>
<td>0.152</td>
</tr>
<tr>
<td>B15</td>
<td>3</td>
<td>0.232</td>
<td>0.460</td>
<td>0.502</td>
<td>2.995</td>
<td>0.466</td>
</tr>
<tr>
<td>B17</td>
<td>4</td>
<td>0.339</td>
<td>0.618</td>
<td>0.458</td>
<td>4.00</td>
<td>0.626</td>
</tr>
<tr>
<td>B20</td>
<td>5</td>
<td>0.418</td>
<td>0.707</td>
<td>0.416</td>
<td>3.929</td>
<td>0.528</td>
</tr>
<tr>
<td>B29</td>
<td>4</td>
<td>0.589</td>
<td>0.524</td>
<td>-0.116</td>
<td>3.929</td>
<td>0.528</td>
</tr>
<tr>
<td>B32</td>
<td>4</td>
<td>0.411</td>
<td>0.658</td>
<td>0.384</td>
<td>3.929</td>
<td>0.666</td>
</tr>
<tr>
<td>B37</td>
<td>3</td>
<td>0.161</td>
<td>0.311</td>
<td>0.49</td>
<td>2.995</td>
<td>0.315</td>
</tr>
<tr>
<td>B43</td>
<td>6</td>
<td>0.321</td>
<td>0.701</td>
<td>0.548</td>
<td>5.929</td>
<td>0.711</td>
</tr>
<tr>
<td>B80</td>
<td>7</td>
<td>0.600</td>
<td>0.691</td>
<td>0.141</td>
<td>6.997</td>
<td>0.698</td>
</tr>
<tr>
<td>B94</td>
<td>4</td>
<td>0.250</td>
<td>0.542</td>
<td>0.548</td>
<td>4.00</td>
<td>0.551</td>
</tr>
<tr>
<td>B98</td>
<td>4</td>
<td>0.321</td>
<td>0.591</td>
<td>0.463</td>
<td>4.00</td>
<td>0.599</td>
</tr>
<tr>
<td>B106</td>
<td>6</td>
<td>0.491</td>
<td>0.548</td>
<td>0.114</td>
<td>5.945</td>
<td>0.554</td>
</tr>
<tr>
<td>Mean</td>
<td>4.417</td>
<td>0.346</td>
<td>0.542</td>
<td>0.402</td>
<td>4.387</td>
<td>0.549</td>
</tr>
</tbody>
</table>

Notes: *N*: number of alleles; *H*: observed heterozygosity; *H*<sub>E</sub>: expected heterozygosity; *F*<sub>ST</sub>: inbreeding coefficient; *A*: allelic richness; *G*: gene diversity.

0.882 (average 0.402). Detailed information was showed in Table 2.

Genetic structure among populations

The most likely number of genetic clusters was identified *K* = 2 using STRUCTURE HARVESTER. One cluster grouped two populations (DSD and XPZ), the remaining two populations grouped another cluster (Fig. 2). The partition of variances within and among populations, and among groups in the AMOVA was 55.9%, 9.7% and 34.4%, respectively (Table 3). Genetic differentiation coefficient (*F*<sub>ST</sub>) between populations, ranging from 0.0644 (SWC and JCW) to 0.5641 (SWC and DSD) suggested a high level of significant genetic differentiation (Table 4).

The effective population size

Estimates of effective population size based on linkage disequilibrium for four populations were given in Table 5. Only one population (SWC) had estimates more than 30 (31.3-62.7) under three level of lowest allele frequency (FL = 0.01-0.05). XPZ have estimates more than 10 under two level of lowest allele frequency (FL = 0.01 and 0.02) and lower than 10 under FL = 0.05.

Table 3. AMOVA of 4 populations of *B. versicolor*

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Sum of squares</th>
<th>Variance components</th>
<th>Percentage variation(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among groups</td>
<td>85.5</td>
<td>1.28</td>
<td>34.4</td>
</tr>
<tr>
<td>Among populations</td>
<td>20.48</td>
<td>0.36</td>
<td>9.7</td>
</tr>
<tr>
<td>Within populations</td>
<td>223.83</td>
<td>2.08</td>
<td>55.9</td>
</tr>
<tr>
<td>Total</td>
<td>329.81</td>
<td>3.72</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 4. Genetic differentiation between populations (above diagonal) and its significance (below diagonal)

<table>
<thead>
<tr>
<th></th>
<th>SWC</th>
<th>JCW</th>
<th>XPZ</th>
<th>DSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWC</td>
<td>0.0644</td>
<td>*</td>
<td>0.3711</td>
<td>0.5641</td>
</tr>
<tr>
<td>JCW</td>
<td>*</td>
<td>0.3192</td>
<td>0.5351</td>
<td></td>
</tr>
<tr>
<td>XPZ</td>
<td>*</td>
<td>*</td>
<td>0.2418</td>
<td></td>
</tr>
<tr>
<td>DSD</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Notes: * means values significant at P < 0.05.

Table 5. The effective population size of each populations of *B. versicolor*

<table>
<thead>
<tr>
<th>Ne (F&lt;sub&gt;L&lt;/sub&gt;=0.05)</th>
<th>Ne (F&lt;sub&gt;L&lt;/sub&gt;=0.02)</th>
<th>Ne (F&lt;sub&gt;L&lt;/sub&gt;=0.01)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWC</td>
<td>31.3</td>
<td>62.7</td>
</tr>
<tr>
<td>JCW</td>
<td>3.1</td>
<td>3.1</td>
</tr>
<tr>
<td>XPZ</td>
<td>7.0</td>
<td>13.2</td>
</tr>
<tr>
<td>DSD</td>
<td>3.2</td>
<td>3.2</td>
</tr>
</tbody>
</table>

Notes: Ne: the effective population size; F<sub>L</sub>: lowest allele frequency.

The other two populations had the same estimates under three level of lowest allele frequency, respectively, which were lower than 10 (3.1 for JCW, 3.2 for DSD).

DISCUSSION

The extremely high diversity of the wild begonias have increasingly attracted attentions of the botanists, and several studies have been done on evaluation of the genetic status of *Begonia* (Hughes and Hollingsworth, 2008; Hughes et al., 2003, 2015; Matolweni et al., 2000; Twyford, et al. 2014, 2015; Tseng et al., 2017). Previous studies showed high genetic diversity and significant differentiation at the species level in *Begonia*. Our study is the first attempt to investigate the population genetics of *B. versicolor* in order to evaluate
its genetic status. Compared with the reports on genetic diversity analysis using microsatellite in other Begonia species, our study showed low genetic diversity in B. versicolor. Two possible reasons may account for this consequence. Firstly, most of inbreeding coefficient in populations (Table 1) and loci (Table 2) were positive, which means a heterozygote deficiency due to inbreeding. Inbreeding in populations would cause loss of genetic variation and low genetic diversity. The microsatellites used in our research were developed from transcriptome sequence by Twyford et al. (2013) and may be the second reason for low genetic diversity. Transcriptome sequences are highly correlated with protein-coding genes with important functions. It has shown that microsatellites derived from transcriptome have lower genetic diversity than microsatellites from genome (Li et al., 2004; Martin et al., 2010).

A high level and significant genetic differentiation was observed in the populations of B. versicolor. Four populations of B. versicolor were grouped into two clusters based on STRUCTURE, this result was almost clustered according to their geographical distributions, and there were few gene exchanges between two clusters (Fig. 2). The AMOVA analysis indicated that percentage of variation among groups (33.4%) is higher than among populations within groups (9.7%). This differentiation can also be seen in the genetic structure analysis, clustering based on geographic regions. Previous studies revealed little gene exchange among populations of Begonia, even with short distance (only a few kilometers apart) or small spatial scale within the same habitat (Dewitte et al., 2011; Hughes and Hollingsworth, 2008; Hughes et al., 2003; Matolweni et al., 2000). Begonias were thought to not attract specialist pollinators but generalist pollinators such as small bees and flies leading to a low seed set and releases the seeds, which passively resulted in localized dispersal patterns (Agren and Schemske, 1991; De Lange and Bouman, 1999; Matolweni et al., 2000). Both reasons lead to the restricted gene flow and isolation between populations in Begonia. Strong and long-term isolation between populations generated extensive genetic divergence and high potential for speciation, which may be one of the explanations for this mega-diverse genus.

Genetic variation is important for a species to maintain its evolutionary potential to cope with ever-changing environments (Frankham et al., 2002), and the status of genetic diversity is important for designing a conservation strategy for threatened and endangered species (Francisco-Ortega et al., 2000; Hamrick and Godt, 1989). Our study showed that relatively low level of genetic diversity, significant genetic differentiation and clear genetic structure in populations of B. versicolor. An immediate action must be carried out for its conservation. Genetic drift and restricted gene flow due to fragmentation is the main reason for losing genetic diversity. Firstly, the habitats of natural populations must be well protected and the populations with high genetic diversity and low effective population size should be given more concerns. Based on this principle, conservation of population JCW with the highest genetic diversity and the lowest effective population size must be a top priority in B. versicolor, followed by population DSD, SWC, and finally population XPZ. Fortunately, SWC and JCW populations are currently located in Yunnan Daweishan National Nature Reserve. Secondly, ex situ conservation strategies should be also developed to conserve this species. The common strategy is to establish a germplasm bank, which should cover individuals from each population of this species. Populations with high genetic diversity should be given a priority. Finally, further research should be conducted to translocate or reintroduce this species.

ACKNOWLEDGEMENTS

This work is supported by the funds from Shanghai Administration Bureau of Landscape and City Appearance (funded project code: F122416), the specific project for strategically biological resources and technology supporting system from Chinese Academy of Sciences (CAS) (funded project code: CZBZX-1), and The National Natural Science Foundation of China (funded project code: 31570199).

LITERATURE CITED

Excoffer, L. and H.E. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics


