Identification of the Geographic Origin of *Dendrobium thyrsiflorum* on Chinese Herbal Medicine Market Using Trinucleotide Microsatellite Markers

Ying Hui YUAN, Bei Wei HOU, Hui Jun XU, Jing LUO, and Xiao Yu DING*

College of Life Sciences, Nanjing Normal University; Nanjing 210046, People’s Republic of China.

Received March 7, 2011; accepted September 13, 2011; published online September 28, 2011

The stems of *Dendrobium thyrsiflorum* RCHB.F. ex ANDRÉ can be processed into an important class of Traditional Chinese Medicine named “Huangcao Shihu,” which has diverse curative effects, such as nourishing yin and clearing away unhealthy heat, benefiting the stomach, and promoting the production of body fluid. The identification of the geographical origin of *D. thyrsiflorum* is vital for preserving its natural resource and ensuring the quality of “Huangcao Shihu.” In order to identify the origin of *D. thyrsiflorum* on Chinese herbal medicine market, 14 *D. thyrsiflorum*-specific microsatellite markers were developed in this study. Assignment tests were performed by the microsatellite marker analysis coupled with three new statistical approaches (partially Bayesian, frequency-based, and fully Bayesian methods) to determine the origin populations of 12 commercial samples of “Huangcao Shihu” collected from a medicine market in Nanjing, Jiangsu Province, China. Their genotypes were compared with those of 136 individuals belonging to five wild *D. thyrsiflorum* populations from China, Thailand, India, Myanmar, and Laos. Comparisons of the probabilities of 12 unknown individuals originating from each candidate population indicated that most of them appeared to originate from Myanmar and Laos. This suggests that the two countries may be the predominant sources of *D. thyrsiflorum* on the medicine market in Nanjing. In addition, the 14 microsatellite markers developed in this study may be an effective tool for identification of the origin of commercial available “Huangcao Shihu” and play an important role in its quality control.

Key words *Dendrobium thyrsiflorum*; microsatellite; assignment test; Bayesian

The genus *Dendrobium* (Orchidaceae) confirmed by Olaf Swartz in 1799 consists of more than 1400 species, which are distributed throughout Asia, Europe, and Australia. Herba Dendrobii, one of the most precious classes of Traditional Chinese Medicine (TCM), is made of the stems of various *Dendrobium* species. Based on their morphological characteristics and processing methods, Herba Dendrobii is classified into fresh *Dendrobium*, “Fengdu Shihu,” and “Huangcao Shihu.” “Huangcao Shihu” is the predominant form of Herba Dendrobii on Chinese herbal medicine market due to its extensive usage and a high demand for it. As a tonic medicine, “Huangcao Shihu” can enhance the immune function, nourish yin and clear away heat-evil, and moisten the lung and relieve a cough. Approximately 27 species of *Dendrobium*, such as *Dendrobium thyrsiflorum* RCHB.F. ex ANDRÉ, *Dendrobium capillipes* RCHB.F., *Dendrobium chrysotoxum* LINDL., and *Dendrobium fimbriatum* LINDL., can be processed into “Huangcao Shihu.” Among them, *D. thyrsiflorum* is regarded as one of the most representative sources of “Huangcao Shihu” because of its relatively abundant wild resources, stronger reproductive capacity, and higher yields. *D. thyrsiflorum* is rich in scoparone and coumarins, which have effects of relaxing smooth muscle, expanding vessels, and anti-coagulating blood. Besides, this medicinal herb is a good immunomodulator due to its abundant content of polysaccharides. *D. thyrsiflorum* is also used as an ornamental plant because of its graceful flowers. *D. thyrsiflorum* is mostly distributed in South China, Thailand, India, Myanmar, Laos, and Vietnam. Because of the uniqueness of its habitat, this herb is vulnerable to the destruction of the natural ecosystem, even a tiny change. However, the wide usage of *D. thyrsiflorum* in medicine and horticulture has caused the over-exploitation of its natural resources over the past few decades, resulting in a substantial decrease in its stock.

The demand for “Huangcao Shihu” is growing in China, which leads to massive imports of *D. thyrsiflorum* from foreign countries. Preliminary market surveys show that *D. thyrsiflorum* on Chinese herbal medicine market mainly comes from a number of Southeast Asian countries. Chinese herbal medicine practitioners hold a consensus that herbs grown in different places vary in curative effect. Thus, choosing good populations of medicinal herbs has become a key factor in the modernization of TCM. Identification of the origins of medicinal herbs is vital for controlling the quality of TCM. Accordingly, elucidating the origin of *D. thyrsiflorum* is critical for the quality control of commercially available “Huangcao Shihu.” In addition, once the geographic origin of collected *D. thyrsiflorum* individuals is identified, effective measures can be taken to prohibit its over-exploitation and protect its wild resources. Therefore, an accurate, convenient, and sensitive method is urgently needed for the identification of the origin of *D. thyrsiflorum*.

Microsatellites, also referred to as simple sequence repeats (SSRs), are tandem repeated units of short nucleotide motifs. Numerous lines of evidence have demonstrated that SSRs exist in any region of eukaryotic genomes, including coding and non-coding regions. Microsatellite markers have been used to identify the geographic origins of commercially available herbs and shown to be an effective tool for individual assignment as well as genetic variation analysis for various species due to their features: composite genotypes across several loci, relatively high polymorphism, and co-dominant inheritance. On the other side, microsatellite markers are quite species-specific, which limits their use as a universal tool for testing multiple species. To date, the development and use of SSR markers of *D. thyrsiflorum* has not been reported yet. Therefore, it is extremely urgent to develop a set of *D. thyrsiflorum*-specific SSR markers by means...
of an efficient library enrichment method.

In the present study, using microsatellite marker analysis and new statistical methods, assignment tests were performed for identification of the origin of *D. thyrsiflorum*. The objectives of our study were (a) to develop a set of highly informative SSR markers of *D. thyrsiflorum*; (b) to investigate the genetic diversity and differentiation of five wild *D. thyrsiflorum* populations, which served as candidate populations for assignment tests; and (c) to identify the origin (i.e., the harvesting place) of processed *D. thyrsiflorum* on Chinese herbal medicine market.

**MATERIALS AND METHODS**

**Sampling** In July 2009, 12 “Huangcao Shihu” samples made of the stems of *D. thyrsiflorum* were randomly collected from a local medicine market in Nanjing, Jiangsu Province, China. From 2007 to 2009, 136 wild individuals of *D. thyrsiflorum* were sampled from five countries (China [Yunnan Province, South China], Thailand, India, Myanmar and Laos), which represented five candidate populations of *D. thyrsiflorum* for assignment tests. All samples were authenticated by Xiaoyu Ding and kept in College of Life Sciences, Nanjing Normal University. Five locations for sampling are marked in Fig. 1, and their detailed geographical information is shown in Table 1.

**DNA Extraction** Genomic DNA was extracted from each “Huangcao Shihu” sample and the leaves of each fresh *D. thyrsiflorum* sample according to the standard cetyltrimethyl ammonium bromide (CTAB) method. The extracted DNA was loaded into 1% agarose gel containing 0.5 mg/ml of ethidium bromide and then subjected to electrophoresis; its relative purity and quantity were crudely estimated by visually comparing it with a DNA sample with known concentration. Finally, each DNA sample was adjusted to 20 ng/µl for subsequent use.

**Development of SSR Genetic Markers** The genomic DNA used for SSR primer development was extracted from a fresh *D. thyrsiflorum* sample collected from Yunnan Province, China; the extracted DNA was digested with *Sau3AI* (TaKaRa, Japan) and separated in 1.5% agarose gel. Size-selected fragments (300—700 bp) were extracted from the gel and purified with a gel extraction kit (Qiagen, U.S.A.). Subsequently, super SNX24 linkers (AP11, 5'-GATCGTCGACGTTACCGAATTCT-3' and AP12, 5'-GTCAAGAATTCCGTACGCGACGAC-3') serving as priming sites for polymerase chain reaction (PCR) were ligated onto the ends of the fragments. The enrichment of DNA fragments was performed by their hybridization with biotinylated oligonucleotide probes (GAA)_{10}, and then streptavidin-coated magnetic beads (Promega, U.S.A.) were used to capture biotinylated oligos. After stringent washing, the captured DNA fragments were amplified using SNX24 primers and inserted into pMD18-T using the TA cloning kit (TaKaRa) and subsequently transformed into chemically competent cells (DH5a).

Subsequently, 158 positive colonies containing an inserted microsatellite sequence were picked, and their DNA was amplified via PCR. After the PCR products were sequenced by an ABI 3100 DNA sequencer with a BigDye Terminator kit (Applied Biosystems, Foster City, CA, U.S.A.), 60 appropriate microsatellite sequences were chosen for primer design. Ultimately, 24 pairs of primers were designed with DNA-MAN Version 5.2.9 and Primer 5.0. They were tested for the effectiveness in generating clear amplification products with evident polymorphism in length by means of PCR with the genomic DNA of five fresh *D. thyrsiflorum* samples collected from Yunnan Province, China as template DNA. Of the 24 primer pairs, 14 pairs were finally chosen for analysis of the genetic characteristics of the 14 SSR loci of *D. thyrsiflorum* (Table 2).

**PCR Amplification and Microsatellite Genotyping**

Using PCR, all “Huangcao Shihu” and fresh *D. thyrsiflorum* samples were genotyped at 14 microsatellite loci (see Table 2). PCR reactions of all samples were carried out in a 10 µl reaction volume consisting of 20 ng of template DNA, 2 mM MgCl_{2}, 200 μM deoxyribonucleotide triphosphate (dNTP) (TaKaRa), 1 µl of 10× PCR buffer containing 50 mM KCl and 10 mM Tris–HCl (TaKaRa), 0.2 μM primer, and 0.5 U *Taq* DNA polymerase (TaKaRa). PCR was performed in Mastercycler Pro (Eppendorf) according to the following program:

**Table 1. Genetic Variation of Five Candidate Populations of *D. thyrsiflorum***

<table>
<thead>
<tr>
<th>Population code</th>
<th>Population name</th>
<th>Population size</th>
<th>Location</th>
<th>Longitude E</th>
<th>Latitude N</th>
<th>Na</th>
<th>Ho</th>
<th>He</th>
<th>Loci departing from HW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CN</td>
<td>28</td>
<td>Yunnan Province, China</td>
<td>99°83’</td>
<td>23°46’</td>
<td>7.6</td>
<td>0.64</td>
<td>0.76</td>
<td>YYH001, YYH006</td>
</tr>
<tr>
<td>2</td>
<td>MVA</td>
<td>30</td>
<td>Mong Si, Myanmar</td>
<td>98°39’</td>
<td>23°68’</td>
<td>7.2</td>
<td>0.69</td>
<td>0.76</td>
<td>None</td>
</tr>
<tr>
<td>3</td>
<td>LAO</td>
<td>28</td>
<td>Sen Sai, Laos</td>
<td>102°37’</td>
<td>21°69’</td>
<td>7.9</td>
<td>0.79</td>
<td>0.78</td>
<td>None</td>
</tr>
<tr>
<td>4</td>
<td>THA</td>
<td>23</td>
<td>Mae Suek, Thailand</td>
<td>98°15’</td>
<td>18°79’</td>
<td>8.4</td>
<td>0.69</td>
<td>0.80</td>
<td>YYH006, YYH011</td>
</tr>
<tr>
<td>5</td>
<td>IND</td>
<td>27</td>
<td>Maredumil, India</td>
<td>81°75’</td>
<td>17°56’</td>
<td>8.9</td>
<td>0.74</td>
<td>0.82</td>
<td>YYH001</td>
</tr>
</tbody>
</table>

Na: mean number of alleles per locus, Ho: mean observed heterozygosity, He: mean expected heterozygosity, HW: expected Hardy–Weinberg proportions.
Genetic Diversity and Genetic Structure Analysis To estimate the level of the genetic diversity of candidate populations, the software ARLEQUIN3.20 was used to calculate the number of alleles per locus (Na), observed heterozygosity (Ho), and expected heterozygosity (He). Tests for Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were also performed with ARLEQUIN3.1. Furthermore, a sequential Bonferroni correction was performed to determine significance of the above tests.21)

Using ARLEQUIN3.1, an analysis of molecular variance (AMOVA) was conducted to assess the genetic variation of candidate populations. By means of ARLEQUIN3.1, the values of pairwise FST were calculated to analyze the extent of genetic differentiation between any two of the five candidate populations.

The genetic structure of each D. thyrsiflorum population was determined by means of the STRUCTURE program with the admixture model.22) This model is based on the following two assumptions. The first is that the genome of individuals is a mixture of genes originating from K unknown ‘ancestral’ populations, which may have undergone introgression events; the second is that the K unknown ‘ancestral’ populations are at HW equilibrium. Six independent runs were performed for each K value (number of clusters) ranging from 2 to 7. Each run consists of a burn-in period of 100000 steps and subsequent 200000 iterations. The choice of the appropriate K was made according to the ad hoc ln Pr(X | K) method,22,23) and the inferred K value with the highest ln Pr(X | K) values was chosen. Averaging the estimated membership coefficients of the individuals and ancestry estimates yields the proportion of ancestry of each population in each cluster.

Assignment Test In the current assignment test, we used the following three statistical methods: (I) partially Bayesian method coupled with an exclusion-simulation test; (II) frequency-based approach combined with the above exclusion-simulation test; (III) fully Bayesian method.22) The above three statistical methods have been available for individual assignment based on its multi-locus genotype.22)

A common feature of the above three statistical methods is that the assignment of an unknown individual to its most likely origin population is based on the probability of its genotype in each candidate population.11) Additionally, these methods share the assumption that there are Hardy–Weinberg proportions within populations as well as independence and linkage equilibrium between loci.22) The assumption is a prerequisite for computing the probability of a genotype in each population based on allele frequencies since deviations from Hardy–Weinberg may affect the accuracy of assignment tests.

On the other hand, the above three statistical methods differ in the principle underlying the performance of the assignment test. For the method I, the probability of an individual belonging to a population is calculated with the Monte Carlo simulations, which simulate 10000 independent individuals.

Table 2. Genetic Characteristics of 14 SSR Loci

<table>
<thead>
<tr>
<th>Locus code</th>
<th>Forward primer (5’—3’)</th>
<th>Reverse primer (5’—3’)</th>
<th>Repeat motif</th>
<th>Tm (°C)</th>
<th>Allele size range (bp)</th>
<th>Number of alleles</th>
<th>GenBank accession No.</th>
<th>Ho</th>
<th>He</th>
</tr>
</thead>
<tbody>
<tr>
<td>YYH001</td>
<td>CTTCCACCACCTGCAAGTTTCA</td>
<td>TGAAGAACCAGCAGTTAACC</td>
<td>(CTT)11</td>
<td>57</td>
<td>220—275</td>
<td>12.0</td>
<td>HQ610596</td>
<td>0.78</td>
<td>0.87</td>
</tr>
<tr>
<td>YYH002</td>
<td>GACAGTGCACAAACCACACAC</td>
<td>GAGATACACTACAGACACACTA</td>
<td>(GAA)16</td>
<td>60</td>
<td>263—325</td>
<td>11.4</td>
<td>HQ610597</td>
<td>0.82</td>
<td>0.85</td>
</tr>
<tr>
<td>YYH003</td>
<td>TGCTCAGCACAATACACAC</td>
<td>GTCTTGTGAGACCCCTG</td>
<td>(GAA)2</td>
<td>57</td>
<td>228—257</td>
<td>6.4</td>
<td>HQ610598</td>
<td>0.82</td>
<td>0.78</td>
</tr>
<tr>
<td>YYH004</td>
<td>AGTTCTGGTGAGCCCTGTA</td>
<td>GCCATCAACAAACACCTCTT</td>
<td>(TTC)5</td>
<td>62</td>
<td>212—257</td>
<td>7.0</td>
<td>HQ610599</td>
<td>0.80</td>
<td>0.78</td>
</tr>
<tr>
<td>YYH005</td>
<td>CGACCTCCTATTTAGAGCA</td>
<td>TACCTTGACCGCCACCTCAT</td>
<td>(CTT)7</td>
<td>62</td>
<td>178—224</td>
<td>8.6</td>
<td>HQ610600</td>
<td>0.66</td>
<td>0.75</td>
</tr>
<tr>
<td>YYH006</td>
<td>TCTACCTTGGTCTCTG</td>
<td>CGCGATCCATTGAGAGA</td>
<td>(CTT)12</td>
<td>57</td>
<td>395—430</td>
<td>6.2</td>
<td>HQ610601</td>
<td>0.56</td>
<td>0.77</td>
</tr>
<tr>
<td>YYH007</td>
<td>ATCGTCTCCTCCTTTCTTC</td>
<td>TGAGCTCATGCGTTTCCAC</td>
<td>(CTT)8</td>
<td>59</td>
<td>234—267</td>
<td>6.6</td>
<td>HQ610602</td>
<td>0.65</td>
<td>0.78</td>
</tr>
<tr>
<td>YYH008</td>
<td>CGACCTTACCAGGAAGCC</td>
<td>TGGTGAACGCAAAGAAAGCT</td>
<td>(TTC)10</td>
<td>62</td>
<td>136—175</td>
<td>8.4</td>
<td>HQ610603</td>
<td>0.81</td>
<td>0.85</td>
</tr>
<tr>
<td>YYH009</td>
<td>CGGTAGCAGCAGACAGACAG</td>
<td>GGAGGAGAAGCTAGGAG</td>
<td>(TTC)7</td>
<td>62</td>
<td>341—390</td>
<td>8.4</td>
<td>HQ610604</td>
<td>0.55</td>
<td>0.70</td>
</tr>
<tr>
<td>YYH100</td>
<td>CTCCTGCACCTTGGCCACA</td>
<td>TCCACCCCGATCCATTCCTCC</td>
<td>(GAA)11</td>
<td>66</td>
<td>489—520</td>
<td>5.8</td>
<td>HQ610605</td>
<td>0.63</td>
<td>0.76</td>
</tr>
<tr>
<td>YYH101</td>
<td>GAAATCTCGTGCAGTCCA</td>
<td>CCCCCAGATTCAAACCAT</td>
<td>(TTC)6</td>
<td>62</td>
<td>489—530</td>
<td>9.2</td>
<td>HQ610606</td>
<td>0.75</td>
<td>0.78</td>
</tr>
<tr>
<td>YYH102</td>
<td>CGTGAACCCAACGGAAACC</td>
<td>CCAAGAGGGAGAAGGGAGAAG</td>
<td>(TTC)12</td>
<td>62</td>
<td>292—330</td>
<td>6.2</td>
<td>HQ610607</td>
<td>0.59</td>
<td>0.75</td>
</tr>
<tr>
<td>YYH103</td>
<td>CCTAACTGCTGCCGATA</td>
<td>GAAAGGAGCATCACAT</td>
<td>(TTC)6</td>
<td>56</td>
<td>250—289</td>
<td>7.2</td>
<td>HQ610608</td>
<td>0.70</td>
<td>0.81</td>
</tr>
<tr>
<td>YYH104</td>
<td>AGAAATGGAAAGCAACCAT</td>
<td>GACCCCTTTCCTCAGTGA</td>
<td>(GAA)6</td>
<td>60</td>
<td>235—257</td>
<td>8.4</td>
<td>HQ610609</td>
<td>0.80</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Tm: annealing temperature, Ho: mean heterozygosity observed (directly calculated estimate), He: mean heterozygosity expected (unbiased estimate Nei, 1987). All estimates resulted from the data of five candidate populations.
for each candidate population,\(^{15}\) while the method II is carried out without simulation.\(^{13}\) Moreover, in the software GENECLASS2.0, neither approach I nor II requires the assumption that the true origin population had been sampled.\(^{25}\) Method III, an alternative to the method I, is essentially a clustering method and used to identify populations (clusters) and to assign each individual to each cluster probability. Based on the assumption that the true origin population had been sampled, method III implemented in the software STRUCTURE2.2\(^{22}\) was used to compute the probabilities of an individual originating from each population.

RESULTS

The detected number of alleles per locus ranged from 5.8 to 12 with an average of 8.0. The averages of the observed and expected heterozygosities across loci were 0.71 and 0.79, respectively. Among 14 loci, only YYH001 and YYH006 revealed significant HW departure in 40% of candidate populations, and YYH011 showed HW deviation in one population. No significant linkage disequilibrium was observed between any pair of loci. These values suggested that 11 of the 14 loci were available for analysis of the genetic diversity and structure of \(D.\ thyrsiflorum\), and for subsequent assignment tests.

The above 11 microsatellite markers were employed to analyze the genetic diversity of five \(D.\ thyrsiflorum\) populations. As shown in Table 1, the mean number of alleles per locus was 8.0 for the five candidate populations, and mean values of Ho and He within populations were 0.71 and 0.78, respectively.

Genetic variations among \(D.\ thyrsiflorum\) populations were estimated using hierarchical \(F\)-statistics. The \(F_{ST}\) values of pairwise comparisons of the five \(D.\ thyrsiflorum\) populations ranged from 0.06 to 0.22, exhibiting that the genetic differentiation among \(D.\ thyrsiflorum\) populations was significant (Table 3).

Using fully Bayesian method, the genetic structure of each \(D.\ thyrsiflorum\) population was determined with the STRUCTURE program. According to the ad hoc \(\ln Pr(X\mid K)\) method of Pritchard et al.,\(^{22}\) \(\ln Pr(X\mid K)\) attained the maximal approximate value at \(K=5\), which implied that it was perfect to divide the 136 analyzed individuals of \(D.\ thyrsiflorum\) into 5 clusters. For this reason, those 136 individuals were classified into 5 clusters of a bar plot (Fig. 2). Table 5b demonstrates the proportions of the membership of 5 \(D.\ thyrsiflorum\) populations in each cluster, indicating that 78.1% of \(D.\ thyrsiflorum\) individuals in Cluster 1, 65.4% in Cluster 2, 69.3% in Cluster 3, 78.6% in Cluster 4, and 64.4% in Cluster 5, originated from China, Thailand, Laos, Myanmar, and India, respectively (Table 5a).

Based on the above three statistical methods (partially Bayesian method, the frequency-based method, and fully Bayesian method), the probability of each unknown individual originating from each candidate population was calculated. The resulting probabilities are presented in Table 4 and Table 5b.

As shown in Table 4, the probabilities of 12 unknown individuals originating from the CN population were 0.239, 0.578, 0.084, 0.018, 0.034, 0.330, 0.018, 0.145, 0.000, 0.026, 0.000, 0.000, which were computed by the partially Bayesian method. Summing the above 12 probabilities yielded the expectation value of 1.454 for 12 unknown individuals assigned to the CN population. Similarly, the expectation values of 12
unknown individuals assigned to the MYA, LAO, THA, and IND populations were 2.341, 2.96, 1.391, and 1.091, respectively. The sum of all the expectation values of 12 unknown individuals assigned to each candidate population was less than 12 since some individuals could not be assigned to any candidate population and considered to have ‘mixed origin.’

The frequency-based method was used to calculate the probabilities of each unknown individual belonging to each candidate population. Summing the resulting probabilities yielded the expectation values of 2.06409, 3.93581, 4.00010, 1.00000, and 1.00000 for 12 unknown individuals assigned to the CN, MYA, LAO, THA, and IND populations, respectively (Table 4).

Analogously, by summing the probabilities of each unknown individual originating from each candidate population that were computed with the fully Bayesian clustering method, we obtained the expectation values of 2.334, 3.897, 3.417, 1.316, and 1.036 for 12 unknown individuals assigned to the CN, MYA, LAO, THA, and IND populations, respectively (Table 5a).

**DISCUSSION**

During recent years, excessive exploitation has caused a rapid decrease in the wild stock of *D. thyrsiflorum*. Furthermore, *D. thyrsiflorum* grown or cultivated in different regions vary in contents of active ingredients of this herb, which leads to the difference in its therapeutic potency. The determination of the geographic origin of *D. thyrsiflorum* on Chinese herbal medicine market is crucial for the quality control of “Huangcao Shihu” as well as the protection of the natural resources of *D. thyrsiflorum*.

Microsatellite marker analysis coupled with recently emerging statistical methods plays an important role in tracking the origin of living plants or their processed products. The function of SSRs appears to be determined by their locations. Trinucleotide repeats are the most common repeats present in coding regions, playing a role in controlling the activity of a gene and truncation of the protein product, while dinucleotide repeats mainly exist in non-coding regions. Although trinucleotide motifs are less frequently used for the analysis of genetic structure and diversity of plant populations than dinucleotide repeats, trinucleotide motifs are less likely to suffer from slippage errors and can be scored less ambiguously. Up to now, the use of trinucleotide microsatellite markers to investigate the population structure of *D. thyrsiflorum* has not been reported. In addition, it has been demonstrated that GAA is the most common triplet motif in many plants. In this study, using a genomic library enriched with GAA repeats, we developed 14 new polymorphic *D. thyrsiflorum*-specific microsatellites, which enabled us to detect multiple alleles at each of these loci with a large number of repeat motifs. This discovery coincided with that found in *Dendrobium officinale*. In order to increase the number of polymorphic SSRs and their discriminatory power, microsatellites containing abundant repeat motifs should be chosen after high-throughput cloning and sequencing.

The high values of the mean number of alleles per locus (Na=8.0) and the mean expected heterozygosity across loci (He=0.79) of *D. thyrsiflorum* implicate that the SSRs developed herein may be a useful tool for genotype identification, germplasm conservation, and evaluation of its genetic diversity and population structure. In light of previous studies, markers with higher heterozygosity are able to achieve higher assignment accuracy; highly polymorphic loci are
thus advisable (He=0.60—0.80). Moreover, the number of loci strongly influences assignment accuracy, so the availability of a large number of polymorphic loci is instrumental in improving the efficacy of the established assignment tests in identification of the sources of samples.\textsuperscript{11,25,31} Eleven polymorphic loci were available for the assignment tests in the present study, which is in accordance with that recommended by previous studies.\textsuperscript{25,32} Average expected heterozygosity (gene diversity) is a good indicator of genetic variation within populations. The mean expected heterozygosity across loci within populations of \textit{D. thyrsiflorum} ranged from 0.76 to 0.82, suggesting that this species has a high level of genetic diversity comparable to that of other \textit{Dendrobium} species.\textsuperscript{10,33,34} Based on a published study,\textsuperscript{12} a high level of genetic diversity of \textit{D. thyrsiflorum} may increase the power of the statistical analysis for identifying the origin of living or processed \textit{D. thyrsiflorum} samples.

The results of this study suggest that the endangered status of \textit{D. thyrsiflorum} may be caused by recent over-collection of its wild populations, rather than a decrease of overall genetic diversity due to environmental factors. Therefore, the collection of wild \textit{D. thyrsiflorum} populations should be controlled, and cultivated \textit{D. thyrsiflorum} plants can be used as the raw material for producing “Huangchao Shihu.”

Earlier studies have indicated that significant genetic differentiation among populations is an important parameter for achieving high assignment accuracy.\textsuperscript{25} However, encouraging results of population assignment tests were still obtained even when overall differentiation was low.\textsuperscript{24,35} In addition, the value of within-population heterozygosity affects the maximum value of \(F_{ST}\) when highly variable loci are applied.\textsuperscript{36} If the within-population heterozygosity is high, an \(F_{ST}\) of 0.05 for highly variable microsatellite loci may indicate a biologically significant level of population differentiation. In the present report, the \(F_{ST}\) values ranging from 0.06 (IND vs. LAO, IND vs. THA) to 0.22 (MYA vs. THA) indicate a moderate level of population differentiation. Although these \(F_{ST}\) values are somewhat lower than that recommended by Koskinen,\textsuperscript{32} the significance of differentiation among populations implies that accurate assignments may be feasible.

By means of the softwares GENECCLASS2.0 and STRUCTURE2.2, all unknown individuals have been assigned to their most likely origin populations with high confidence. Although the partially Bayesian exclusion and the frequency-based method were less accurate than the fully Bayesian method,\textsuperscript{15} the three methods yielded the consistent results that most of the unknown individuals were assigned to the MYA and the LAO population, implying that the natural resources of \textit{D. thyrsiflorum} in Myanmar and Laos may have been undergoing over-harvesting. On the other side, the substantial import of foreign \textit{D. thyrsiflorum} through illegal trade has seriously affected the interest of Chinese farmers in planting this herb.

Comparisons of the expectation values of 12 unknown individuals originating from each candidate population suggest that Myanmar and Laos may be the predominant sources of \textit{D. thyrsiflorum} on the medicine market in Nanjing, Jiangsu Province, China. The results of our assignment tests are preliminary. In the further research, we need to significantly improve the performance of assignment tests by taking account of the following factors: size of reference samples, number of polymorphic markers, and genetic differentiation of reference populations. All of them may influence the accuracy of assignment tests.\textsuperscript{33} Thus, the accuracy of assignment tests may be greatly improved by sampling a large number of sub-populations of \textit{D. thyrsiflorum} from China, Thailand, India, Myanmar, Laos, and Vietnam. It is urgent to develop more polymorphic microsatellite markers of \textit{D. thyrsiflorum} for significantly improving the reliability of assignment tests in the future.

In view of the extensive distribution of \textit{D. thyrsiflorum}, we suggest establishing its database and DNA registration in order to effectively identify its source. For this purpose, individual samples of this species should be collected from each distribution area and genotyped, and then the resulting genotypes are inputted into the database. Thus, the origin of any fresh \textit{D. thyrsiflorum} sample or its processed product may be identified by comparing its genotype with those in the database, which would make the tracing and verifying the origin of \textit{D. thyrsiflorum} much easier as well as play an important role in constraining illegal collection of wild \textit{D. thyrsiflorum}.

This study provides valuable baseline data on the population genetics of \textit{D. thyrsiflorum}, which are useful for developing strategies aimed at conserving its germplasm resources and optimizing the system for its artificial propagation. For protecting the natural resources of \textit{D. thyrsiflorum}, we firstly recommend its \textit{in situ} conservation due to its special habitat. A significant change in its habitat can affect the survival of the coexisting organisms, such as fungi and pollinators, finally threatening its survival. Moreover, improving the tissue culture techniques of \textit{D. thyrsiflorum} is also beneficial to its germplasm conservation.

Acknowledgments The study was funded by the National Natural Science Foundation of China (No. 30870234) and the Natural Science Foundation of Jiangsu Province (BK20081496).

REFERENCES