Pak. J. Bot., 53(5): 1683-1690, 2021. DOI: http://dx.doi.org/10.30848/PJB2021-5(19)

GENETIC DIVERSITY AND POPULATION STRUCTURE OF CULTIVATED *DENDROBIUM HUOSHANENSE* (C.Z. TANG ET S.J. CHENG) USING SNP MARKERS GENERATED FROM GBS ANALYSIS

MEIRONG YE*, XIAOPENG WANG, YULI ZHOU, SHOUCHENG HUANG AND AIRONG LIU

College of Life and Health Sciences, Anhui Science and Technology University, Donghua Road, Fengyang, Anhui, China
**Corresponding author's email: lindayyyy815@126.com, yemr@ahstu.edu.cn*

Abstract

Dendrobium huoshanense C.Z. Tang et S.J. Cheng, plants have been declared as geographical indicators (GI) of different ecoregions of China, and wild *D. huoshanense* has been listed as one of the rare and endangered medicinal herbs of China, owing to overexploitation, habitat deteriorationand, *et al*., Therefore, *D. huoshanense* plants have been cultivated on a large scale for protecting the endangered species and meeting the great market demands. However, the genetic diversity and population structure of cultivated *D. huoshanense* are still unknown. Therefore, genetic diversity and population structure of 72 cultivated *D. huoshanense* samples were investigated in the study, using single nucleotide polymorphism (SNP) markers generated from genotyping-by-sequencing (GBS) analysis. A total of 100,098,874 clean read pairs were generated, with 1,390,262 clean read pairs per sample on average, and the ratio of mapped reads to total clean reads was 96.45%. A total of 108,557,703 SNPs were detected which included 94,382,323 (87%) homozygous and 14,175,380 (13%) heterozygous loci. Transition SNPs were more frequent than transversed SNPs, and the ratio of transition SNPs to transversion SNPs was 1.47. An unrooted neighbor-joining phylogenetic tree based on 359,521 filtered SNPs classified 72 cultivated *D. huoshanense* samples into two major clusters. According to ADMIXTURE analysis, these samples collection could be divided into two groups, which was in agreement with the NJ tree. According to AMOVA analysis, 96.68% variance was due to variation within populations, and 3.32% among populations, indicating a high gene exchange (or low genetic differentiation) among these populations.These results provided a valuable resource for understanding genetic diversity of cultivated *D. huoshanense.*

Key words: *Dendrobium huoshanense*; GBS-SNPs; Genetic diversity; Population structure; NJ tree; Clean reads

Introduction

Genotyping-by-sequencing (GBS) approach using next generation sequencing (NGS) technologies is a truly revolutionary technology in the research of plant breeding and genetics becasue it is a practical, low cost, a large number of genome-wide single nucleotide polymorphisms (SNP) discovery and SNP genotyping (Elshire *et al.,* 2011; Poland & Rife 2012). Up to now, large numbers and genome-wide availability of SNPs have been the ideal molecular markers of choice in plant genetic research. The applications of genome-wide SNP markers included linkage map construction, genetic diversity analyses, association mapping, and marker assisted selection and so on (Kumar *et al.,* 2012). For example, Baral *et al*., (2018) and Luo *et al*., (2019) applied GBS to analyze genetic diversity of Crested *Agropyron cristatum* (L.) Gaertn (Wheatgrass) and *Camelina sativa* Spring Panel, respectively. Bekele *et al*., (2018) investigated haplotype-based GBS in oat genome; Lin *et al*., (2019) assessed the genetic differentiation and linkage disequilibrium in *Solanum pimpinellifolium* using genome-wide high-density SNP markers. Therefore, SNPs identified by GBS would be increasingly applied in diverse plants genetics research, breeding and others aspects in the future.

Dendrobium is one of the largest genus of the Orchidaceae family and widely distributed in the tropical and subtropical regions of Asia, Europe, and Australia (Burke *et al*., 2008). *Dendrobium* is composed of approximately 1500 species worldwide and in China, it is comprised of 74 species and two varieties (Tsi, 1999; Ye *et al*., 2015), such as *Dendrobium huoshanense* C.Z. Tang and S.J. Cheng, *Dendrobium officinale* Kimura et Migo, *Dendrobium moniliforme* (Linnaeus) Swartz and so on. *D. huoshanense* plants have been declared as Geographical Indication (GI) products of China, for they are endemic to Ta-pieh Mountains of China including Huoshan, Jinzhai, Yuexi and Shucheng County of Anhui Province and Yingshan County of Hubei Province, indicating their narrow geographical distribution (Yuan *et al*., 2018). It has been classified in the "Conservation Program for Wild Plants with Extremely Small Population in China" (Ma *et al*., 2013). *D. huoshanense* is used in traditional Chinese medicines (TCMs) with high medicinal value and is widely utilized as a medicinal and edible herb in health care product for its broad pharmacological effects, such as nourishing the stomach, treating throat inflammation, enhancing immunity, improving eyesight, slowing aging, possessing anticancer properties and so on (Bulpitt *et al*., 2007; Ge *et al*., 2018; Xie *et al*., 2019). Wild *D. huoshanense* has been listed as one of the rare and endangered medicinal herbs of China in 1987 for over-exploitation, habitat deterioration, slow growth rates and high demands of market due to its medicinal and nutritional value (Li *et al*., 2021). Nowadays, *D. huoshanense* plants are in the stage of artificial cultivation for protecting their resources and meeting high demands of market (Niu *et al*., 2020). Thus, *D. huoshanense* plants have been cultured on a large-scale around Ta-pieh Mountain, especially in Anhui, Hubei, and Henan provinces (Niu *et al*., 2020). However, the genetic diversity of cultivated D. *huoshanense* plants is still unknown. A known well about population structure and genetic diversity of a rare and endangered species is

essential for effective development of conservation strategies. Some types of genetic markers have been developed and applied to determine genetic diversity of *D. huoshanense*, including random amplified polymorphic DNA (RAPD) (Liu *et al*., 2007), polymorphic microsatellite loci (SSR) (Wang *et al*., 2012; Zheng *et al*., 2011). Niu *et al*., (2020) have selected 27 mutational hotspot regions and six polymorphic cpSSRs to assess the genetic diversity and distributional patterns of *D. huoshanense* based on the comparative plastomic approaches. Among all types of molecular markers, SNPs, which were discovered by GBS, have been widely employed to quantify genetic variation, to determine parentage relatedness and population structure, due to their ubiquitous presence, uniform distribution, biallelic nature, and high heritability (Verma *et al*., 2015). However, genome-wide SNPs discovered by GBS have never been applied to study genetic diversity and population structure of cultivated *D. huoshanense* plants. Therefore, the goals of the study were to apply GBS technology (1) to study genetic diversity of cultivated *D. huoshanense*, (2) to discover SNPs for cultivated *D. huoshanense*, and (3) to investigate the population structure of cultivated *D. huoshanense*, using *Dendrobium officinale, Dendrobium moniliforme, Dendrobium Fanjingshanense* which are wild species as out groups.

Materials and Methods

Plant material and DNA extraction: A total of 72 cultivated *D. huoshanense* samples used for the study were collected from 10 plant breeders (or populations) in Ta-pieh Mountains of China, 13 *D. officinale* samples, 5 *D. moniliforme* samples and 1 *D. fanjingshanense* sample were gathered as out groups (Table 1). Fresh young leaves of the above each sample were harvested and immediately frozen in liquid nitrogen. These leaves samples were stored at drikold and transported to Wuhan Frasergen Bioinformatics Co., Ltd. for genomic DNA extraction and genotyping by sequencing (GBS) analysis (Yin *et al*., 2020).

DNA extraction, library construction and GBS analysis: Total genomic DNA of each sample was isolated from 0.1 g fresh leaves using modified CTAB. The quality and quantity of isolated DNA samples was measured with 0.8% agarose gel electrophoresis and a Nanodrop spectrophotometer (Thermo Scientific, DE, USA), respectively (Ryu *et al*., 2019). The total concentration of isolated DNA samples which was detected using Qubit 3.0 Fluorometer (China) was greater than or equal to 10 μg for GBS libraries construction. GBS libraries were constructed using a single restriction enzyme (Msel). Genomic DNA from each sample was digested with Msel restriction enzymes. The digested genomic DNAs with a barcode adaptor and a common Illumina sequencing adaptor were combined, and purified using a commercial kit (QIAquick PCR Purification Kit, Qiagen, Valencia, CA, USA), according to the manufacturer's guidelines. Restriction fragments from each library were then multiplexed and amplified by PCR. These amplified sample pools constituted a sequencing "library" (Tanhuanpääetal *et al*., 2016). The resulting library was sequenced on the Illumina PE150 platform at Wuhan Frasergen Bioinformatics Co., Ltd. (Yin *et al*., 2020).

Sequence preprocessing and alignment to reference genome: The raw reads were demultiplexed using GBSx package (Herten *et al*., 2015). Then reads were processed to remove possible adapator sequences and to remove the base at pairs end of sequencing read shorter than 20 bases, discarded read shorter than 50 bases, and filtered low quality regions and only remained pair of reads by using Trimmomatic.0.38 software (Bolger *et al*., 2014). Cleaned reads were mapped to the genome of *Dendrobium catenatum* Lindl (Zhang *et al*., 2016) with Burrows-Wheeler Alignment tool (BWA) (Li & Durbin 2009; Li & Durbin 2010). SNPs of clean reads were detected and filtered by GATK (McKenna *et al*. 2010). The resulting SNPs were further filtered by disregarding the ones with MAF<0.01 and missing in >20% of the samples for the population genetic analyses (Danecek *et al*., 2011; Luo *et al*., 2019).

Table 1. Location, sampling size and population name of 72 *D. huoshanense* and out groups in the study.

Sample name	Sample source/ population name	Population code	Sample size	Location
D. huoshanense	Gaoshanshan	GSS	8	Huoshan Anhui Province
	Liyouzhi	LYZ	10	Huoshan Anhui Province
	Huxiaoyu	HXY	6	Huoshan Anhui Province
	Jiuxianzhai	JXZ	5	Huoshan Anhui Province
	Xiancaozhai	XCZ	12	Huoshan Anhui Province
	Jiucaoshou	JCS	3	Huoshan Anhui Province
	Wanjiahu Ⅰ	WJH Ⅰ	9	Anqing Anhui Province
	Wanjiahu Ⅱ	WJH Ⅱ	4	Anqing Anhui Province
	Shangelao	SGL	7	Huoshan Anhui Province
	Huzhibao	HZB	8	Huoshan Anhui Province
D. officinale	Tiepi	TP	13	4 Shizong Yunnan Province 2 Yandang Zhejiang Province 7 Huoshan Anhui Province
D. moniliforme	Xijing	XJ	5	2 Weijiang Guangxi Province 3 Huoshan Anhui Province
D. fanjingshanense	Fanjingshan	FJS	1	Fanjingshan Guizhou Province

Table 2. Summary of average GBS clean reads per sample and percentage of alignment to reference genome sequence.

Population code	Clean reads pairs	Clean base (bp)	Mapped (%)	Properly pair mapped (%)	Singletons mapped (%)
GSS	1535753	436387047	97.01	79.79	0.79
LYZ	1363626	388041796	97.09	80.02	0.74
HXY	1419248	403715202	96.54	79.21	0.74
JXZ	1513928	431546882	97.01	80.02	0.71
XCZ	1331617	379003091	96.78	79.49	0.73
JCS	1622830	461920246	87.36	71.67	0.71
WJH Ⅰ	1255413	357315732	96.97	81.96	0.68
WJH Ⅱ	1302193	370162593	95.04	78.94	0.74
SGL	1475757	419660339	96.97	81.92	0.69
HZB	1300724	370090025	97.22	82.64	0.63
Average	1,390,262	395,557,949	96.45	80.15	0.72
Total	100,098,874	28,480,172,319			

Note: Mapped (%): reads compared to the reference genome / total clean reads *100%; Properly pair mapped (%): pair of reads compared to the reference genome/total clean reads *100%; Singletons mapped (%): single of reads compared to the reference genome/total clean reads *100%

Table 3. Summary of average SNP per sample detected by GBS.

Population code	SNP num	Transition	Transvertion	Ts/Tv	Heter num	Heter ratio	Hom num	Hom ratio
GSS	1509648	899064	610825	1.47	204940	0.135	1304709	0.865
LYZ	1440572	860667	579761	1.485	192507	0.133	1248065	0.867
HXY	1520897	905288	614905	1.472	186049	0.122	1334848	0.878
JXZ	1662744	988730	673495	1.468	205916	0.124	1456828	0.876
XCZ	1475675	877303	597849	1.467	178078	0.12	1297597	0.88
JCS	1607678	952139	655480	1.45	207784	0.13	1399894	0.87
WJH Ⅰ	1465942	876008	589592	1.486	198625	0.134	1267318	0.866
WJH Ⅱ	1382385	820181	561359	1.463	164356	0.12	1218029	0.88
SGL	1593143	947616	645701	1.469	220768	0.14	1372375	0.86
HZB	1528693	911122	617832	1.474	214277	0.138	1314416	0.863
Average	1507746	897857	609668	1.47	196880	0.13	1310866	0.87
Total	108557703	64645705	43896062		14175380		94382323	

Note: SNP num: the number of SNPs; Transition: the number of transition SNPs; Transvertion: the number of transvertion SNPs; Ts/Tv: The ratio of transition SNPs to transivation SNPs; Heter num: the number of heterozygous SNPs; Heter ratio: the ratio of heterozygous SNPs to total SNPs; Hom num: the number of homozygous SNPs; Hom ratio: the ratio of Homozygous SNPs to total SNPs

Analysis of population structure: Population structure was estimated using ADMIXTURE software (Alexander *et al*., 2009; Li *et al*., 2019). The best K with the lowest cross-validation error was determined following the procedure of cross-validation in the manual (Alexander *et al*., 2009). An unrooted neighbor-joining phylogenetic tree was constructed using TreeBeST (Vilella *et al*., 2008). A principal coordinate analysis (PCoA) was conducted by GCTA software (Yang *et al*., 2011; 2013). Genetic distances between pairs of samples or populations was calculated using GenAlEx v6.5 (Peakall & Smouse. 2012). And AMOVA was analysed by GenAlEx (Peakall & Smouse, 2012), revealing the level of variance among and within populations.

Results

Genotyping-by-sequencing of *D. huoshanense*: The GBS library was constructed from cultivated 72 *D. huoshanense* samples, and sequenced by the Illumina platform (PE150). The sequencing produced 28,889,900,352 bp (28.89 Gb) raw base in total, 100,098,874 clean reads pairs [clean base: 28,480,172,319 bp (28.48Gb)], with 1,390,262 clean reads (clean base: 395,557,949 bp) per each sample on average. The highest average clean reads pairs of the population was JCS (1,622,830 clean reads), and the lowest was WHJ Ⅰ (1,255,413 clean reads). A summary of these sequencing results was presented in Table 2. Raw bases of 91 samples were deposited in NCBI (PRJNA659117). Genome sequences (1.01 Gb) of *Dendrobium catenatum* Lindl. were used as a reference genome in the study (ASM160598v2, Zhang *et al*. 2016).On average, 96.45% of the clean reads were mapped to the reference genome. properly pair mapped was 80.15%, and singletons mapped was 0.72% (see Table 2), Average covering approximately 5.47% of the reference genome. Average mapped sequencing depth of *D. huoshanense* populations ranged from 5.566 to 6.6 (Appendix 1).

Appendix 1. The result of average sequencing depth, coverage (%) of *D. huoshanense* populations.

Population code	Average sequencing depth (x)	Coverage (%)	Coverage at least 4X (%)	Coverage at least 10X (%)	Coverage at least 20X (%)
GSS	6.6	5.474	0.836	0.098	0.026
LYZ	5.96	5.385	0.65	0.072	0.022
HXY	5.941	5.633	0.715	0.08	0.025
JXZ	6.02	5.962	0.628	0.07	0.02
XCZ	5.771	5.473	0.644	0.07	0.02
JCS	5.957	5.79	0.697	0.073	0.02
WJH Ⅰ	5.731	5.234	0.668	0.07	0.019
WJH Ⅱ	5.933	5.078	0.69	0.073	0.023
SGL	6.306	5.53	0.777	0.083	0.023
HZB	5.566	5.445	0.711	0.075	0.021
Average	5.97	5.47	0.70	0.076	0.022

Appendix 2. Genetic distance between pairwise of *D. huoshanense* populations.

	GSS	LYZ	HXY	JXZ	XCZ	JCS	WJH Ⅰ	WJH Ⅱ	SGL	HZB
GSS	0									
LYZ	0.1827	0								
HXY	0.1963	0.1944	0							
JXZ	0.1980	0.1955	0.1869	0						
XCZ	0.1923	0.1913	0.1873	0.1898	0					
JCS	0.1866	0.1859	0.1953	0.1966	0.1927	0				
WJH Ⅰ	0.1965	0.1948	0.1888	0.1900	0.1909	0.1971	0			
WJH Ⅱ	0.1895	0.1881	0.1907	0.1927	0.1907	0.1865	0.1917	0		
SGL	0.1876	0.1871	0.1970	0.1987	0.1953	0.1878	0.1954	0.1891	0	
HZB	0.1871	0.1870	0.1972	0.1991	0.1953	0.1886	0.1947	0.1885	0.1869	0

Table 4. Summary AMOVA table for 10 populations of *D. huoshanense*.

Source of variation	df	SS	MS	Est. Var.	Total variation (%)
10 populations					
Among populations	9	14729.358	1636.595	45.243	3.32
Within populations	62	81573.656	1315.704	1315.704	96.68
Stat	Value	P(rand≧data			
PhiPT	0.033	0.001			

Note: df degrees of freedom; SS Sum of squares; MS Mean squares; Est. Var.: estimation variation; Probability: P(rand ≧ data); for PhiPT is based on 999 permutations across the full data set

Genome-wide discovery of SNPs: The average SNPs number of individual among populations ranged from 1,382,385 to 1,662,744, with 1,507,746 SNPs per each sample on average (Table 3). In total, 108,557,703 SNPs were detected in the cultivated 72 *D. huoshanense* samples, Homozygous SNPs and heterozygous SNPs were 94,382,323 (87%) and 14,175,380 (13%), respectively, with an average of 1,310,866 homozygous SNPs and 196880 heterozygous SNPs per sample. The SNPs were categorized on base of nucleotide substitutions: 64,645,705 SNPs were transitions [A/G (30,387,654), C/T (34,258,051)], 43,896,062 SNPs were transversitions [A/C(11,678,124), C/G (5,206,939), A/T (14,533,519), G/T (12,477,480)] (Table 3 and Fig. 1). The transition/ transversition ratio was 1.47 on average, which was higher than the results (1.34) reported by Zhang *et al*., (2016).

Genetic diversity and population structure based on filtered SNPs: The genetic relationships and population structure of the cultivated *D. huoshanense* samples were inferred with a total of 420445 high qualities filtered SNPs, which included 359521 filtered SNPs from 72 cultivated *D. huoshanense* samples and 60924 filtered SNPs from the out groups. Different analyses of the filtered SNPs were applied to identify the relationships among the cultivated samples. First, an admixture analysis was performed by the ADMIXTURE software. The most favorable number of cluster was 2 (0.3682), indicating that 72 cultivated *D. huoshanense* samples were divided into two clusters (Fig. 2). Second, 72 cultivated *D. huoshanense* samples, except HZB4, were classified into two major clusters in an unrooted neighbor-joining phylogenetic tree (NJ tree) based on filtered SNPs

(Fig. 3). 13 *D. officinale*, 5 *D. moniliforme* and 1 *D. fanjingshanense* samples were strictly listed as out groups, moreover, 13 *D. officinale* samples, 5 *D. moniliforme* samples and 1 *D. Fanjingshanense* sample were strictly distinguished each other. The result was similar to the admixture analysis. Two main clusters strongly supported by high bootstrap value (100%) were revealed in the tree dendrogram, and the NJ phylogenetic tree showed that individuals in LYZ, GSS, JCS populations were distributed into one cluster, individuals in HXY and JXZ populations into the other cluster, and individuals in the remains populations were distributed into different clusters, moreover, individuals of one population in a cluster were not always clustered together, revealing a high proportion of admixture (Fig. 3). HZB4 sample in the study was not listed in the two clusters of *D. huoshanense* and listed in outgroups, revealing it probably underwent recombination or hybridization. Third, Analysis of molecular variance (AMOVA) demonstrated 3.32% variance among and 96.68% variance within populations, and PhiPT was 0.033 (Table 4), revealing a high gene exchange (or low genetic differentiation) among these populations. Principal coordinates analyses (PCoA) validated the results. PCoA based on the filtered SNPs showed that the first two components only revealed 7.38% of variation (4.6% and 2.78%) (Fig. 4), indicating low genetic differentiation among these samples. Genetic distance between GSS and LYZ population was the lowest (0.1827), and the highest (0.1991) was between JXZ and HZB population (Appendix 2), revealing that the difference of genetic distance between populations was not obvious.

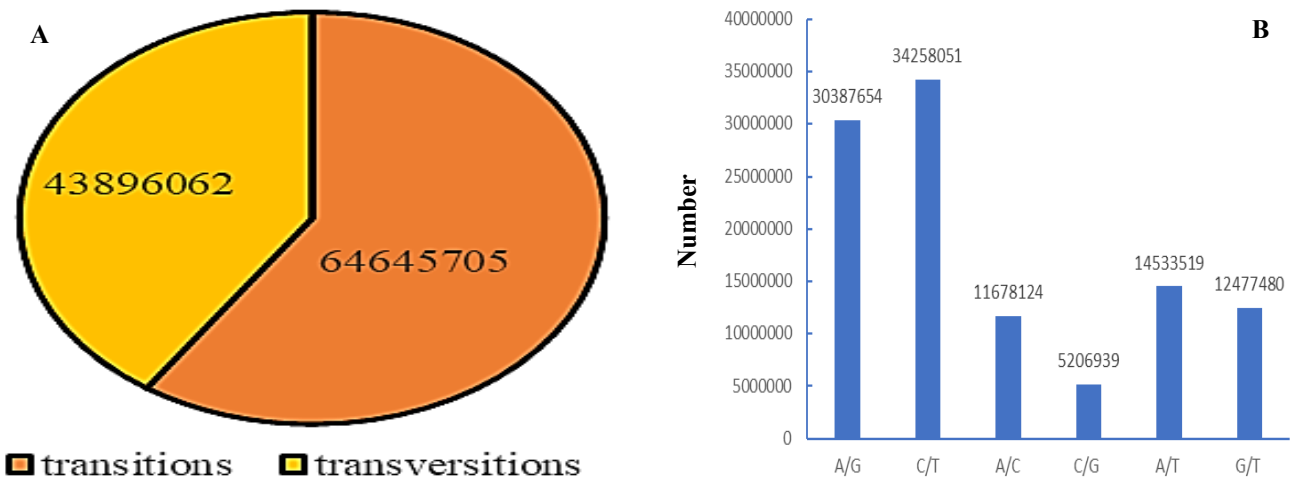

Fig. 1. Number of nucleotide transition and transversion (A) and number of nucleotide substitutions (B).

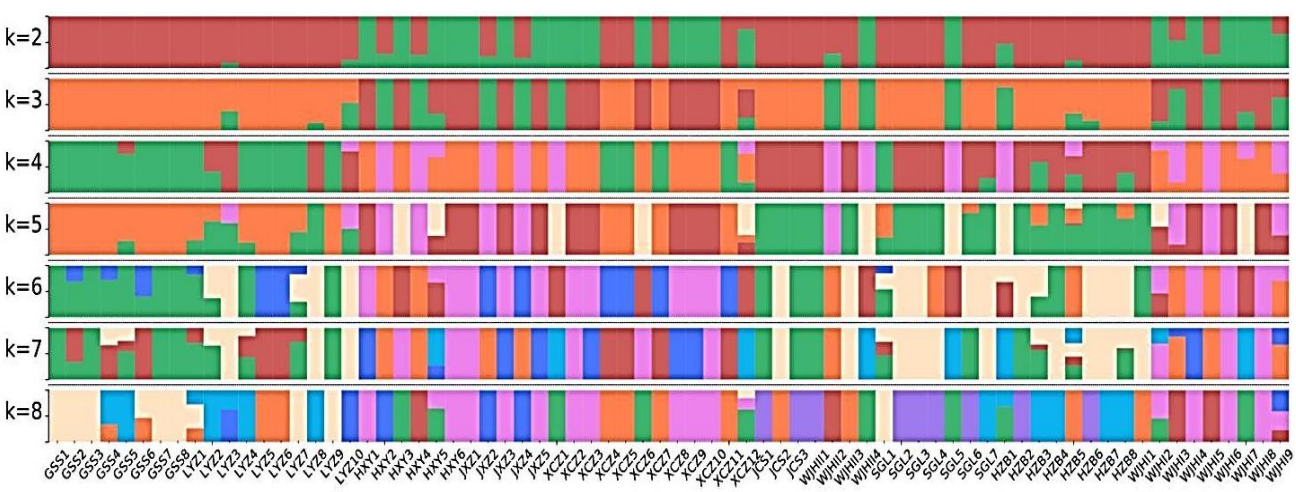

Fig. 2. Genetic structure of cultivated 72 D. *huoshanense* for K=2-8 based on the ADMIXTURE software. K=2 or 3 with cross-validation error (0.3682 or 0.3962, respectively).

Discussion

Wild *Dendrobium* plants have been listed as endangered taxa by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) (Xiang *et al*., 2011; Zhang *et al*., 2018) for overexploitation, habit deterioration, high market demand, their low germination rate and slow growing. Therefore it is of utmost urgent to know well their genetic diversity. For genetic diversity is the core of biological diversity (Chen *et al*., 2013; Tripathi *et al*., 2012). Conservation of genetic diversity for threatened and endangered species is

crucial for their adaptation to changing environments, long term survival, and evolution (Avise & Hamrick, 1996; Frankham, 2012; Munoz *et al*., 2010; Swarts *et al*., 2009). Therefore, genetic diversity and population structure of different Dendrobium species have been investigated by using various molecular markers. For example, Ding *et al*., (2008) revealed genetic diversity across natural populations of *D. officinale* by inter-simple sequence repeat (ISSR) and amplified polymorphic DNA (RAPD) markers. Bhattacharyya *et al*., (2013) used start codon targeted (SCoT) marker to reveal genetic diversity of *Dendrobium nobile* Lindl., Ye *et al*., (2015) applied amplified fragment length polymorphism (AFLP) marker to study genetic diversity of the endangered herb *D. moniliforme*. Nguyen *et al*., (2020) investigated the genetic diversity of 76 Dendrobium samples from Southern Vietnam based on the ITS, ITS2, matK, rbcL and trnH-psbA regions. In the paper, genome-wide SNPs detected by GBS were used to study genetic diversity of cultivated *D. huoshanense* populations, which revealed a high gene exchange (or low genetic differentiation) among populations. Niu *et al*., (2020) revealed that human activities have played key roles in shaping the genetic diversity and distributional patterns of *D. huoshanense*, and population differentiation of *D. huoshanense* based on cpDNA data was lower, which was in accord with a low level of genetic differentiation in the study and in other Dendrobium species (Hou *et al*., 2017; Ye *et al*., 2016). A high mixture among populations have been observed, which was in line with the result of Niu *et al*., (2020), indicating those populations were underwent hybridization or introgression events. *D. huoshanense* plants, collected from the Ta-pieh Mountains, have been artificially cultivated on a large scale for protecting the endangered species and meeting the great market demands for medicinal and health care products. *D. huoshanense* plants have not only been grown in greenhouses, but have also been cultivated in imitating wild conditions, such as on the trunk of pear trees or cliffs in nature which resulted in its natural and/or artificial migration. Therefore, gene exchange would occur between cultivated populations, which would bring about the mixed genetic structure among *D. huoshanense* populations. Cultivated HZB4 in the study was listed outside the two clusters of *D. huoshanense*, revealing it probably underwent recombination or hybridization. *D. officinale*, *D. moniliforme* and *D. fanjingshanense* samples were strictly listed into out groups and were strictly distinguished each other, revealing SNPs generated by GBS were able to be used to identify different species of Dendrobium.

In the past years, next-generation sequencing (NGS) technologies, with the decreasing cost and rapid progress, have been recently used for whole genome sequencing and for sequencing several specimens to discover large numbers of SNPs for studying genetic diversity. Until now, 2 Dendrobium genomes have been sequenced, including *Dendrobium catenatum* Lindl. (Zhang *et al*., 2016) and *D. officinale* (Yan *et al*., 2015), and *D. catenatum* was used as a reference genome in the study. Based on the results, the average coverage of the reference genome in the study (5.47%) was higher than the reported (0.48%) by Ryu *et al*., (2019). Recently, GBS methods have been applied in the molecular research of *Dendrobium*. For example, Ryu *et al*., (2019) applied GBS to discover SNPs for investigating genetic characterization of *Dendrobium* mutants and cultivars. Previously, SNPs discovered by GBS had not been performed to investigate genetic diversity of the cultivated *Dendrobium* samples. The study was the first to study genetic diversity of the cultivated D. *huoshanense* samples with SNPs discovered by GBS and revealed that genome-wide SNPs were the ideal molecular markers for investigating genetic diversity and differentiating *Dendrobium* species.

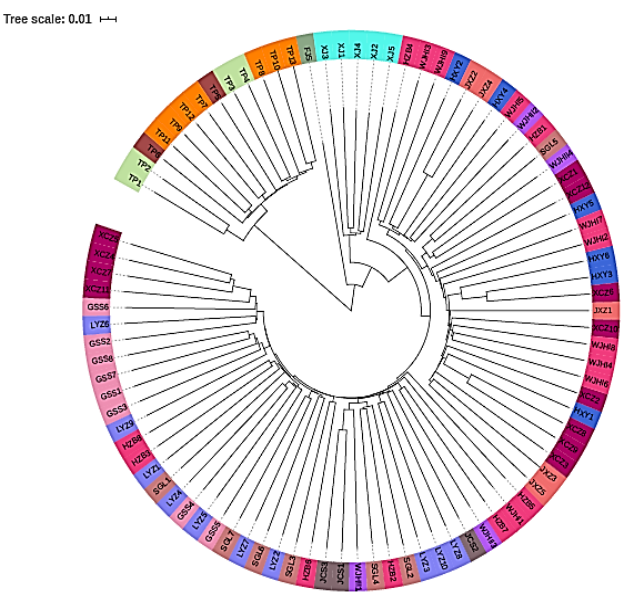

Fig. 3. Neighbor-joining tree based on pairwise distance matrix representing the grouping of the 91 *Dendrobium* plants obtained from 420455 GBS-SNPs.

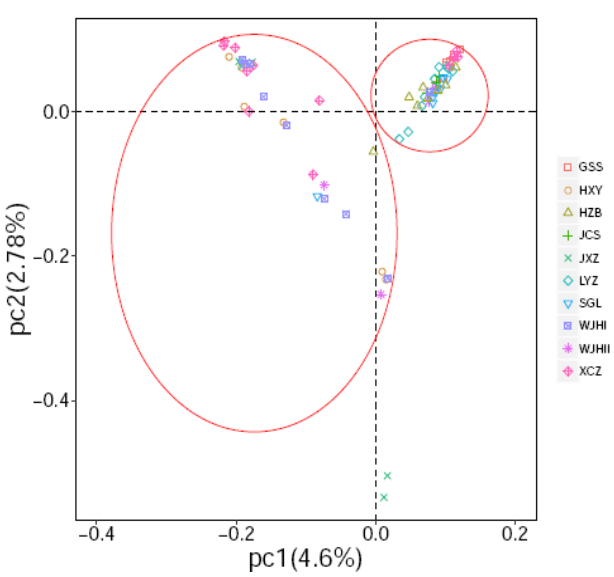

Fig. 4. Principal coordinates analysis (PCoA) of pairwise simple matching dissimilarities between individuals. Axis1 and Axis 2 explain 7.38 % of total variation.

Conclusion

Genetic diversity and population structure of the cultivated *D. huoshanense* samples were investigated using SNPs detected from GBS analysis in the study. 100,098,874 clean reads pairs and 108,557,703 SNPs were detected. 420445 filtered SNPs (359521 from *D. huoshanense* cultivars and 60924 from the out groups) were used to deduce the genetic relationships and population structure of the cultivated *D. huoshanense* samples. The Structure, NJ tree and PCoA analysis indicated that the cultivated *D. huoshanense* samples in the study were obviously divided into 2 clusters, nevertheless, many cultivated *D. huoshanense* samples in some populations were evidently mixed each other together, revealing a high gene exchange (or low genetic differentiation) among these populations. The highest and lowest genetic distance between populations was 0.1991 and 0.1827, respectively, indicating that genetic distance between populations was not distinct. Based on these results, Genetic diversity of *D. huoshanense* was relatively low. To protect *D. huoshanense* species and meet the great market demand, the cultivation size should be raised and its genetic characteristic should be thoroughly investigated in the future.

Acknowledgments

We thank plant breeders for providing cultivated *D. huoshanense* samples and other Dendrobium plants. We thank Wuhan Frasergen Bioinformatics Co., Ltd. for genomic DNA extraction and genotyping by sequencing (GBS) analysis. This work was supported by Natural science fund for colleges and universities in Anhui Province, China (1808085MC78).

References

Alexander, D.H., J. Novembre and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*., 19(9): 1655-1664.

Avise, J.C. and J.L. Hamrick. 1996. Conservation Genetics. *Livestock Prod. Sci.*, 150-152.

Baral, K., B. Coulman, B. Biligetu and Y.B. Fu. 2018. Genotyping-by-Sequencing enhances genetic diversity analysis of crested wheatgrass [*Agropyron cristatum* (L.) Gaertn.]. *Int. J. Mol. Sci*., 19(9): 2587.

Bekele, W.A., C.P. Wight, S. Chao, C.J. Howarth and N.A. Tinker. 2018. Haplotype-based genotyping-by-sequencing in oat genome research. *Plant Biotechnol. J*., 16(8): 1452-1463.

Bhattacharyya, P., S. Kumaria, S. Kumar and P. Tandon. 2013. Start Codon Targeted (SCoT) marker reveals genetic diversity of Dendrobium nobile Lindl., an endangered medicinal orchid species. *Gene*, 529(1): 21-26.

Bolger, A.M., M. Lohse and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15): 2114-2120.

Bulpitt, C.J., Y. Li, P.F. Bulpitt and J. Wang. 2007. The use of orchids in Chinese medicine. *J. R. Soc. Med*., 100(12): 558-563.

Burke, J.M., M.J. Bayly, P.B. Adams and P.Y. Ladiges. 2008. Molecular phylogenetic analysis of Dendrobium (Orchidaceae), with emphasis on the Australian section Dendrocoryne, and implications for generic classification. *Aust. Syst. Bot.*, 21(1): 65-80.

Chen, X., J. Guan, R. Ding, Q. Zhang, X. Ling, B. Qu and L. Zhang. 2013. Conservation genetics of the endangered terrestrial orchid Liparis japonica in Northeast China based on AFLP markers. *Plant Syst. Evol*., 299(4): 691-698.

Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, G. McVean and R. Durbin. 2011. The variant call format and VCFtools. *Bioinformatics*, 27(15): 2156-2158.

Ding, G., D.Z. Zhang, X.Y. Ding, Q. Zhou, W.C. Zhang and X.X. Li. 2008. Genetic variation and conservation of the endangered Chinese endemic herb *Dendrobium officinale* based on SRAP analysis. *Plant Syst. Evol*., 276(3-4): 149-156.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, 6(5): e19379.

Frankham, R. 2012. How closely does genetic diversity in finite populations conform to predictions of neutral theory? Large deficits in regions of low recombination. *Heredity*, 108(3): 167- 178.

Ge, J.C., X.Q. Zha, C.Y. Nie, N.J. Yu, Q.M. Li, D.Y. Peng and J.P. Luo. 2018. Polysaccharides from *Dendrobium huoshanense* stems alleviates lung inflammation in cigarette smoke-induced mice. *Carbohyd. Polym*., 189: 289-295.

Herten, K., M.S. Hestand, J.R. Vermeesch and J.K. Van Houdt. 2015. GBSX: a toolkit for experimental design and demultiplexing genotyping by sequencing experiments. *BMC Bioinform.*, 16(1): 1-6.

Hou, B.W., J. Luo, Y.S. Zhang, Z.T. Niu, Q.Y. Xue and X.Y. Ding. 2017. Iteration expansion and regional evolution: phylogeography of *Dendrobium officinale* and four related taxa in southern China. *Sci. Rep. UK.*, 7: 43525.

Huang, Z.C., Y.Y. Pan, G.Z. Chen, L.J. Chen, X.Y. Wu and J. Huang. 2019. The complete chloroplast genome of *Dendrobium harveyanum* (Orchidaceae). *Mitochon. DNA B Resour.*, 4(2): 3200-3201.

Kumar, S., T.W. Banks and S. Cloutier. 2012. SNP discovery through next-generation sequencing and its applications. *Int. J. Plant Genom.*, 2012: 831460.

Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14): 1754-1760.

Li, H. and R. Durbin. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, 26(5): 589-595.

Li, R., C.Q. Li, H.Y. Chen, X.H. Liu, H. Xiao and S.Y. Chen. 2019. Genomic diversity and admixture patterns among six Chinese indigenous cattle breeds in Yunnan. *Asian-Australas. J. Anim. Sci.*, 32(8): 1069-1076.

Li, Z.Q., H.Q. Zhou, O.Y. Zhen, J. Dai, Q. Yue, Y. Wei and B.X. Han. 2021. Comparison of active ingredients and protective effects of *Dendrobium huoshanense* of different growth years on acute liver injury. *China J. Chin. Mater. Medica*, 46(2): 298-305.

Lin, Y.P., C.Y. Liu and K.Y. Chen. 2019. Assessment of genetic differentiation and linkage disequilibrium in *Solanum pimpinellifolium* using genome-wide high-density SNP markers. *G3-Genes Genom. Genet.*, 9(5): 1497-1505.

Liu, S.Q., X.J. Li, Q.B. Yu, X. Hong and G.Y. Zhou. 2007. Analysis on genetic stability in different development stages of *Dendrobium huoshanense* by RAPD. *China J. Chin. Mater. Medica*, 32(10): 902-905. [Article in Chinese].

Luo, Z., J. Brock, J.M. Dyer, T. Kutchan, D. Schachtman, M. Augustin, Y. Ge, N. Fahlgren and H. Abdel-Haleem. 2019. Genetic diversity and population structure of a *Camelina sativa* spring panel. *Front. Plant Sci.*, 10: 184.

Ma, Y., G. Chen, E.R. Grumbine, Z. Dao, W. Sun and H. Guo. 2013. Conserving plant species with extremely small populations (PSESP) in China. *Biodivers. Conserv.*, 22(3): 803-809.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly and M.A. DePristo. 2010. The genome analysis toolkit: a Map Reduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, 20(9): 1297-1303.

Munoz, M., J. Warner and F.J. Albertazzi. 2010. Genetic diversity analysis of the endangered slipper orchid Phragmipedium longifolium in Costa Rica. *Plant Syst. Evol*., 290(1-4): 217-223.

Nguyen, N.H., H.T. Vu, N.D. Le, T.D. Nguyen, H.X. Duong and H.D. Tran. 2020. Molecular identification and evaluation of the genetic diversity of *Dendrobium* species collected in Southern Vietnam. *Biology (Basel)*, 9(4): 76.

Niu, Z.T., S.Y. Zhu, J.J. Pan, L.D. Li, J. Sun and X.Y. Ding. 2016. Comparative analysis of *Dendrobium plastomes* and utility of plastomic mutational hotspots. *Sci*. *Rep*., 7: 2073.|

Niu, Z.T., Z.Y. Hou, M.T. Wang, M.R. Ye, B.H. Zhang, Q.Y. Xue, W. Liu and X.Y. Ding. 2020. A comparative plastomics approach reveals available molecular markers for the phylogeographic study of *Dendrobium huoshanense*, an endangered orchid with extremely small populations. *Ecol. Evol.*, 10(12): 5332-5342.

Peakall, R. and P.E. Smouse. 2012. GenAlEx 6.5: genetic analysis in excel. Population genetic software for teaching and research -- an update. *Bioinformatics*, 28(19): 2537-2539.

Poland, J.A. and T.W. Rife. 2012. Genotyping-by-Sequencing for plant breeding and genetics. *Plant Genome*, 5(3): 92-102.

Ryu, J., W.J. Kim, J. Im, K.W. Kang, S.H. Kim, Y.D. Jo, S.Y. Kang, J.H. Lee and B.K. Ha. 2019. Single nucleotide polymorphism (SNP) discovery through genotyping-by-sequencing (GBS) and genetic characterization of *Dendrobium* mutants and cultivars. *Sci. Hort.*, 244: 225-233.

Swarts, N.D., E.A. Sinclair, S.L. Krauss and K.W. Dixon. 2009. Genetic diversity in fragmented populations of the critically endangered spider orchid Caladenia huegelii: implications for conservation. *Conserv. Genet.*, 10(5): 1199-1208.

Tanhuanpää, P., M. Erkkilä, T. Tenhola-Roininen, J. Tanskanen and O. Manninen. 2016. SNP diversity within and among *Brassica rapa* accessions reveals no geographic differentiation. *Genome*, 59: 11-21.

Tripathi, N., N. Saini, P. Nair and S. Tiwari. 2012. Lack of genetic diversity of a critically endangered important medicinal plant *Chlorophytum borivilianum* in Central India revealed by AFLP markers. *Physiol. Mol. Biol. Pla.*, 18(2): 161-167.

Tsi, Z.H. 1999. The Flora of China. Science Press, Beijing (Chinese edition).

Verma, S., S. Gupta, N. Bandhiwal, T. Kumar, C. Bharadwaj and S. Bhatia. 2015. High-density linkage map construction and mapping of seed trait QTLs in chickpea (*Cicer arietinum* L.) using Genotyping-by-Sequencing (GBS). *Sci*. *Rep*., 5(1): 17512.

Vilella, A.J., J. Severin, A. Ureta-Vidal, L. Heng, R. Durbin and E. Birney. 2008. EnsemblCompara GeneTrees: Complete, Duplication-Aware Phylogenetic Trees in Vertebrates. *Genome Res.*, 19(2): 327-335.

Wang, H., N.F. Chen, J.Y. Zheng, W.C. Wang, Y.Y. Pei and G.P. Zhu. 2012. Isolation and characterization of eleven polymorphic microsatellite loci for the valuable medicinal plant *Dendrobium huoshanense* and *Cross-Species Amplification*. *Int. J. Mol. Sci.*, 13(12): 16779-16784.

Xiang, X.G., H.U. Hao, W. Wang and X.H. Jin. 2011. DNA barcoding of the recently evolved genus Holcoglossum (Orchidaceae: Aeridinae): a test of DNA barcode candidates. *Mol. Ecol. Resour*., 11(6): 1012-1021.

Xie, S.Z., Z.Z. Shang, Q.M. Li, X.Q. Zha, L.H. Pan and J.P. Luo. 2019. *Dendrobium huoshanense* polysaccharide regulates intestinal lamina propria immune response by stimulation of intestinal epithelial cells via toll-like receptor 4. *Carbohyd*. *Polym*., 222: 115028.

Yan, L., X. Wang, H. Liu, Y. Tian, J.M. Lian, R.J. Yang, S.M. Hao, X.J. Wang, S.C. Yang, Q.Y. Li, S. Qi, L. Kui, M. Okpekum, X. Ma, J.J. Zhang, Z.L. Ding, G.J. Zhang, W. Wang, Y. Dong and J. Sheng. 2015. The genome of *Dendrobium officinale* illuminates the biology of the important traditional Chinese orchid herb. *Mol. Plant*, 8(6): 922-934.

Yang, J., S.H. Lee, M.E. Goddard and P.M. Vissche. 2011. GCTA: a tool for genome-wide complex trait analysis. *Amer. J. Hum. Genet.*, 88(1): 76-82.

Yang, J., S.H. Lee, M.E. Goddard and P.M. Visscher. 2013. Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. *Methods Mol. Biol*., 1019: 215-36.

Ye, M.R., B.W. Hou, J. Luo, W.J. Yan, W. Liu and X.Y. Ding. 2015. Genetic diversity and conservation of the endangered herb *Dendrobium moniliforme* based on amplified fragment length polymorphism markers. *Sci. Hort*., 189: 51-58.

Ye, M.R., W. Liu, Q.Y. Xue, B.W. Hou and X.Y. Ding. 2016. Phylogeography of the endangered orchid *Dendrobium moniliforme* in East Asia inferred from chloroplast DNA sequences. *Mitochond. DNA A.*, 28(6): 880-891.

Yin, M., B.Y. Ye, Y.F. Jin, L. Liu, Y. Zhang, P. Li, Y.H. Wang, Y. Li, Y.P. Han, W.L. Shen and Z.H. Zhao. 2020. Changes in *Vibrio natriegens* growth under simulated microgravity. *Front. Microbiol*., 11: 2040.

Yuan, Y.D., M.Y. Yu, Z.H. Jia, X.E. Song, Y.Q. Liang and J.C. Zhang. 2018. Analysis of *Dendrobium huoshanense* transcriptome unveils putative genes associated with active ingredients synthesis. *BMC. Genom.*, 19(1): 978.

Zhang, G.Q., Q. Xu, C. Bian, W.C. Tsai, C.M. Yeh, K.W. Liu, K. Yoshida, L.S. Zhang, S.B. Chang, F. Chen, Y. Shi, Y.Y. Su, Y.Q. Zhang, L.J. Chen, Y. Yin, M. Lin, H. Huang, H. Deng, Z.W. Wang, S.L. Zhu, X. Zhao, C. Deng, S.C. Niu, J. Huang, M. Wang, G.H. Liu, H.J. Yang, X.J. Xiao, Y.Y. Hsiao, W.L. Wu, Y.Y. Chen, N. Mitsuda, M. Ohme-Takagi, Y.B. Luo, Y. Van de Peer and Z.J. Liu. 2016. The *Dendrobium catenatum* Lindl. Genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. *Sci*. *Rep*., 6: 19029.

Zhang, Y.J., C. Ma, Y. Feng, X. Cheng and J. Song. 2018. The complete chloroplast genome sequence of an endangered orchid species *Dendrobium bellatulum* (Orchidaceae). *Mitochond. DNA B Resour*., 3(1): 233-234.

Zheng, J., N. Chen, H. Wang, P. Gao and G. Zhu. 2011. Isolation and analysis of polymorphic microsatellite loci in *Dendrobium huoshanense*. *China J. Chin. Mater. Medica*, 36: 2926-2931.

(Received for publication 20 May 2019)