Medicinal Plant Research 2022, Vol.12, No.4, 1-10
http://hortherbpublisher.com/index.php/mpr

Research Article Open Access

The Differences and Genetic Relationships of 4 Dendrobium Species Based on High Throughput GBS-SNP Markers

Meirong Ye, Xiaopeng Wang, Yuli Zhou
College of Life and Health Sciences, Anhui Science and Technology University, Fengyang, 233100, China
Corresponding author email: lindayyyy815@126.com
Medicinal Plant Research, 2022, Vol.12, No.4 doi: 10.5376/mpr.2022.12.0004
Received: 03 Aug., 2022
Accepted: 09 Aug., 2022
Published: 15 Aug., 2022

Preferred citation for this article:
Ye M.R., Wang X.P., and Zhou Y.L., 2022, The differences and genetic relationships of 4 Dendrobium species based on high throughput GBS-SNP markers, Medicinal Plant Research, 12(4): 1-10 (doi: 10.5376/mpr.2022.12.0004)

Abstract The differences and genetic relationships of 4 Dendrobium species (*Dendrobium huoshanense, Dendrobium officinale, Dendrobium moniliforme, Dendrobium Fanjingshanense*) were analyzed based on high throughput SNP markers obtained by GBS, using *Dendrobium catenatum* as the reference genome, for providing reference for the study of these 4 Dendrobium species. The results showed that the number of clean reads per each sample in *D. huoshanense*, *D. officinale*, *D. moniliforme*, *D. Fanjingshanense* were 1 300 724, 1 286 162, 1 380 009 and 1 170 337, respectively, and the number of SNPs per each sample were 1 507 746, 893 333, 1 364 605 and 1 227 006, respectively. The genetic structure based on Admixture software revealed that the samples of *D. officinale* were clustered into one branch, and the remaining Dendrobium samples were clustered into the other branch, when the best K value was 2; when k = 3, the samples of *D. officinale* were grouped into one cluster, some samples of *D. huoshanense*, all samples of *D. moniliforme* and *D. fanjingshanense* were grouped into the second cluster, and some samples of *D. huoshanense* were grouped into the third cluster; When k = 4, it was divided into four branches: the samples of *D. officinale* were grouped one branch, the samples of *D. moniliforme* and *D. fanjingshanense* were grouped one branch, the samples of *D. huoshanense* were divided into two branches, which was similar to the result of an unrooted neighbor-joining phylogenetic tree (NJ tree); Furthermore, principal coordinates analyses (PCoA) revealed that 91 Dendrobium samples were classed into three cluster. All samples of *D. officinale* and *D. huoshanense* were grouped together, respectively, and the samples of *D. moniliforme* and *D.* fanjingshan*ense* were grouped together. According to the results of this study, the genetic relation of *D. officinale* and the other 3 species of Dendrobium are relatively distant, and the genetic relationship of *D. moniliforme* and *D. fanjingshanense* were relatively close, 72 samples of *D. huoshanense* in the study were divided into two branches, the genetic relationship in one of which was relatively close to *D. moniliforme* and *D. fanjingshanense*. The study verified that the genetic relationship between different Dendrobium species or different species of plants was investigated using SNPs detected by GBS.
Keywords SNP; Dendrobium; Clean reads; NJ tree

Genotyping by sequencing (GBS) is a simplified genome sequencing technology based on second-generation sequencing technology with the advantages of high throughput and low cost, which could detect a large number of variation sites of single nucleotide polymorphism (SNP) for the study of genetic relationship, population structure, genetics and reproduction of animals and plants (Poland and Rife, 2012). SNP refers to DNA sequence polymorphism caused by single nucleotide variation in the genome, which has the advantages of high density, wide distribution, high stability, simple typing, rapid and high throughput (Kumar et al., 2012). At present, SNP, as a third-generation molecular marker, is more and more widely used in the research of animal and plant genetics and reproduction. For example, Baral et al. (2018) used the SNP obtained by GBS to study the genetic diversity of crested wheatgrass (*Agropyron cristatum* (L.) Gaertn.), revealing the potential of GBS application in the characterization of non-model polyploid plants with complex genomes. Luo et al. (2019) used SNPs generated by GBS to analyze the genetic diversity and genetic structure of *Camelina sativa* Spring Panel.

Dendrobium is the third largest genus in the family of Orchid. There are about 74 species and 2 varieties in China, of which more than 40 species are traditional Chinese medicine, especially *Dendrobium huoshanense* C.Z. Tang and S.J. Cheng and *Dendrobium officinale* Kimura et Migo, which are valuable Chinese medicinal

herbs. There is a recorded in "Ben-Cao Gang-mu (Compendium of Materia Medica)", that "Dendrobium has the effects of 'Chubi Xiaqi (Terms of traditional Chinese Medicine, means removing bi (A TCM disease refers to the abnormal symptoms of joints and muscles caused by external adverse stimulation) and descending qi (A TCM disease refers to the most fundamental and subtle substances that constitute the human body and maintain life activities))', tonify the 'Wuzang (means heart, liver, spleen, kidney and lung)', strong body, nourish Yin (Terms of TCM) and benefit qi. Taking it for a long time can enhance the function of the spleen and stomach, so as to make people flexible, enhance immunity and prolong life." Modern pharmacological research shows that Dendrobium has the functions of anti-oxidation, anti-cancer, antithrombotic, hypoglycemic, enhancing human immunity, anti-aging and so on (Song et al., 2014). Dendrobium is also a kind of plant "Yaoshi Tongyuan" (means that the homology of medicine and food). It is a valuable health product, and the market demand is very large. Due to over excavation, habitat destruction, large market demand, and the slow growth of Dendrobium itself, low seed germination rate, strict requirements on habitat and other reasons, all kinds of Dendrobium species have been listed as endangered species. Therefore, to meet the market demand, Dendrobium species, especially *Dendrobium huoshanense* and *Dendrobium officinale*, are cultivated on a large scale. And the research on various aspects of Dendrobium has also been widely valued, such as the identification of Dendrobium species (Duan et al., 2019), genetic relationship (Ren et al., 2013), genetic diversity (Song et al., 2016), resistance gene research (Li et al., 2013), gene cloning and identification (Li et al., 2018), medicinal components and medicinal value (Si et al., 2016). At present, the whole genome of *Dendrobium officinale* (Yan et al., 2015) and *Dendrobium catenatum* (Zhang et al., 2016) has been sequenced, so researchers can use it as a reference genome to make it easier to use GBS technology to study Dendrobium plant genetic diversity, reproduction, gene cloning and identification, as well as medicinal components. For example, Ryu et al. (2019) used the whole genome of *Dendrobium officinale* as the reference genome to analyze the genetic characteristics of Dendrobium mutants and cultivated species using SNPs produced by GBS. It is rarely reported to analyze the differences and genetic relationships of different species of Dendrobium by using SNPs produced by GBS. Therefore, this study analyzed the differences and genetic relationships of cultivated *Dendrobium huoshanense*, *Dendrobium officinale*, *Dendrobium moniliforme* and *Dendrobium fanjingshanense* by using genome-wide SNP markers obtained by GBS technology, to provide references for the study of Dendrobium.

1 Results and Analysis

1.1 Data filtering and alignment to reference genomes

The number of clean reads per each sample in *Dendrobium huoshanense* (abbreviation: HS), *Dendrobium officinale* (abbreviation: TP), *Dendrobium moniliforme* (abbreviation: XJ) and *Dendrobium fanjingshanense* (abbreviation: FJS) were 1 390 262, 1 286 162, 1 380 009 and 1 170 337, respectively, and the clean bases were 39 557 949 bp, 366 277 414 bp, 391 592 914 bp and 333 221 985 bp, respectively. The TP reads mapped to reference genome of *D. catenatum* accounted for 99.08% of the total reads, which was the highest. And there was no significant difference between HS, XJ and FJS, which were 96.45%, 96.98% and 96.91%, respectively. The proportion of clean reads pairs mapped to the reference genome to the total clean reads and the sequencing length meets a certain threshold to the total number of reads is also the highest in TP, which was 91.01%. There was no significant difference between HS, XJ and FJS, which were 80.15%, 80.08 and 79.21%, respectively (Table 1). The highest coverage of sequencing was TP, which was 6.09%, and the lowest was FJS, which was 4.74% (Table 2).

1.2 Variation detection of SNP

GATK (McKenna et al., 2010) software was used to detect that the average number of SNPs in each sample of HS, TP, XJ and FJS were 1 507 746, 893 333, 1 364 605 and 1 227 006, respectively. The number of transition SNP and transvertion SNP were 897 857 and 609 668, 546 045 and 344 283, 810 614 and 551 968, 727 566 and 495 436, respectively. And only the ratio of transition SNPs to transvertion SNPs of TP was 1.58, and the rest were 1.47. The average number of heterozygous SNPs produced by each sample of HS, TP, XJ, and FJS were

196 880, 229 222, 168 830 and 136 823, respectively. The proportion of heterozygous SNPs in the total number of SNPs was 0.257 for TP, which was the highest, and 0.13 for HS, 0.12 for XJ and 0.11 for FJS. The average number of homozygous SNPs produced by each sample of HS, TP, XJ and FJS were 1 310 866 (highest), 664 110 (lowest), 1 195 774, and 1 090 183, respectively. The lowest proportion of homozygous SNP number in the total SNP number was TP, which was 0.74, the highest was FJS (0.89). HS and XJ were 0.87 and 0.88 (Table 3).

Table 1 Summary of GBS clean reads per sample and percentage of alignment to reference genome sequence

Dendrobium species abbreviation	Clean reads pairs per sample	Clean base per sample (bp)	The proportion of clean reads mapped to the reference genome to the total clean reads (%)	The proportion of clean reads pairs mapped to the reference genome to the total clean reads (%)	The proportion of clean reads singletons mapped to the reference genome to the total clean reads (%)
HS	1 390 262	395 557 949	96.45	80.15	0.72
TP	1 286 162	366 277 414	99.08	91.01	0.27
XJ	1 380 009	391 592 914	96.98	80.08	0.82
FJS	1 170 337	333 221 985	96.91	79.21	0.85

Table 2 The result of average sequencing depth, coverage (%) of 4 species of *Dendrobium*

Dendrobium species abbreviation	Average sequencing depth (X)	Coverage of clean reads mapped to the reference genome (%)	Coverage of clean reads mapped to the reference genome at least 4X (%)	Coverage of clean reads mapped to the reference genome at least 10X (%)	Coverage of clean reads mapped to the reference genome at least 20X (%)
HS	5.97	5.47	0.70	0.076	0.022
TP	5.46	6.09	0.65	0.04	0.01
XJ	6.13	5.32	0.77	0.09	0.03
FJS	5.86	4.74	0.47	0.05	0.01

Table 3 Summary of average SNPs per sample detected by GBS

Dendrobium species abbreviation	The number of SNP	The number of transition SNP	The number of transvertion SNP	The ratio of transition SNPs to Transition SNPs	The number of heterozygous SNP	The ratio of heterozygous SNPs to the total SNPs	The number of homozygous SNP	The ratio of homozygous SNPs to the total SNPs
HS	1 507 746	897 857	609 668	1.47	196 880	0.13	1 310 866	0.87
TP	893 333	546 045	344 283	1.58	229 222	0.26	664 110	0.74
XJ	1 364 605	810 614	551 968	1.47	168 830	0.12	1 195 774	0.88
FJS	1 227 006	727 566	495 436	1.47	136 823	0.11	1 090 183	0.89

1.3 Analysis of population structure and genetic distance

The total number of SNPs for population structure analysis obtained from Clean reads after filtration was 420 445, including 359 521 in *D. huoshanense*, 118 484 in *D. officinale*, 153 325 in *D. moniliforme* and 65 231 in *D. fanjingshanense*. The genetic distance matrix was constructed with Tree BeST software (Vilella et al., 2009), and the Neighbor-joining tree of the system was constructed with the adjacency method (Figure 1). It showed that the samples of *D. officinale* were clustered into one branch, as well as *D. moniliforme*. *D. fanjingshanense* was clustered between *D. moniliforme* and *D. officinale* alone, and the samples of *D. huoshanense* were divided into two branches except one sample, with a support rate of 100%. It also revealed that the samples from different cohabitation groups of *D. huoshanense* were obviously mixed, that was, the samples from cohabitation groups were not completely clustered together, or even divided into different branches, indicating that there was a phenomenon of gene hybridization or gene transfer among *D. huoshanense* samples.

Figure 1 Neighbor-joining tree based on pairwise distance matrix representing the grouping of the 91 Dendrobium plants obtained from 420 455 SNPs-GBS

The genetic structure based on Admixture software (Alexander et al., 2009) revealed that the best K value was 2, samples of *D. officinale* were clustered into one branch, and the remaining Dendrobium samples were clustered into the other branch, that was, *D. huoshanense*, *D. moniliforme* and *D. fanjingshanense* were clustered together, suggesting that their genetic relationship was closer. When K = 3, the samples of *D. officinale* were grouped into one cluster, and the remaining Dendrobium samples were clustered into two branches, some samples of *D. huoshanense*, all samples of *D. moniliforme* and *D. fanjingshanense* were grouped into the second cluster, and some samples of *D. huoshanense* were grouped into the third cluster; When K = 4, the samples of *D. officinale* were grouped one branch, the samples of *D. moniliforme* and *D. fanjingshanense* were grouped one branch, and the samples of *D. huoshanense* were divided into two branches, which was similar to the result of an unrooted neighbor-joining phylogenetic tree (NJ tree) (Figure 2; Figure 3).

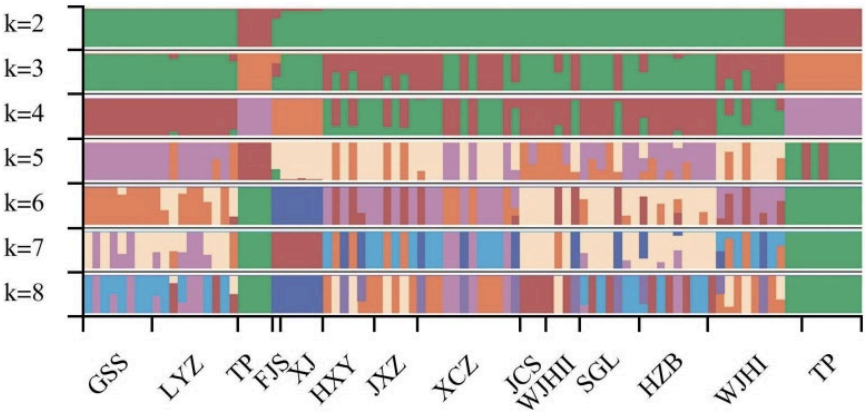

Figure 2 Genetic structure of 91 Dendrobium samples for K=2-8 based on the Admixture software

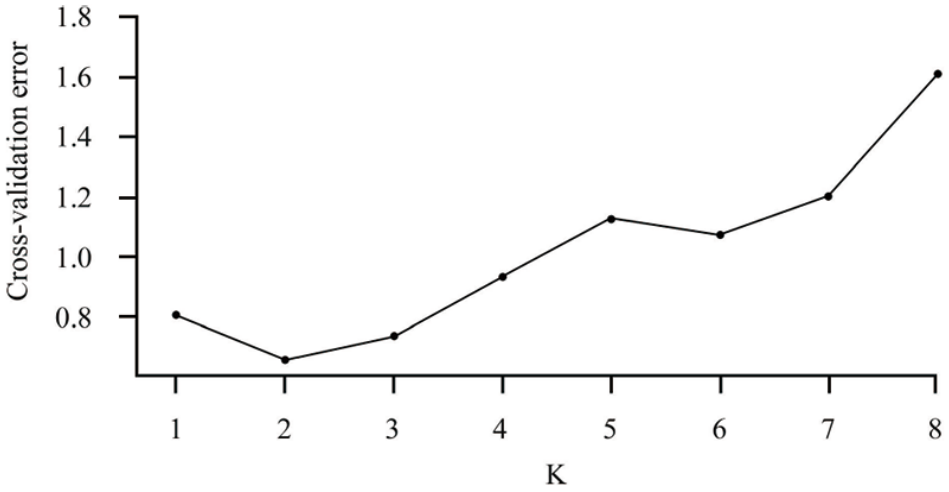

Figure 3 Cross-validation error and K value

Principal coordinates analysis (PCoA) (Yang et al., 2011) was not exactly the same as Structure analysis, which showed that *D. officinale* were clustered together, *D. moniliforme* and *D. fanjingshanense* were clustered together, while *D. huoshanense* were clustered together, with the total genetic variation of 22.14%, while PC1 and PC2 were 18.79% and 3.35%, respectively (Figure 4).

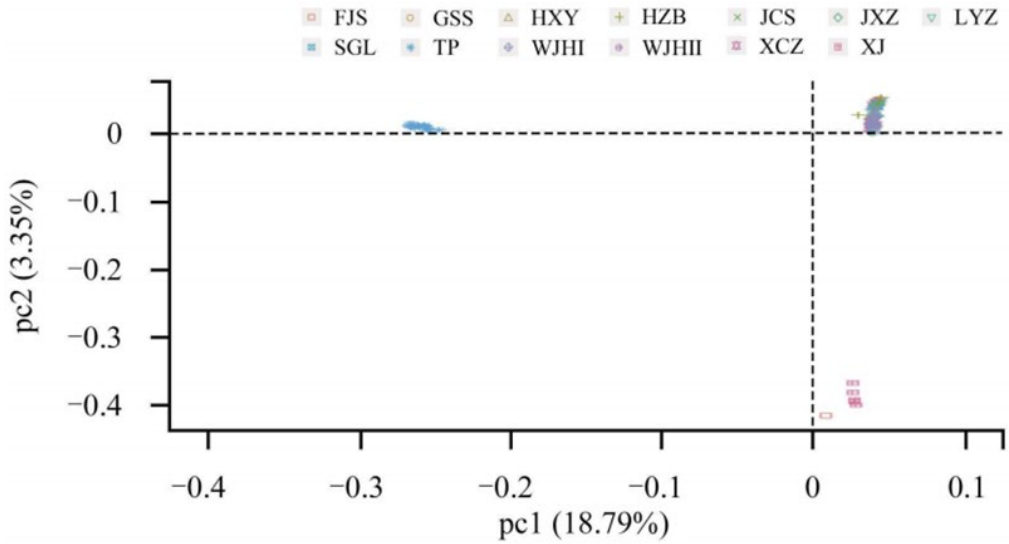

Figure 4 Principal coordinates analysis (PCoA) of pairwise simple matching dissimilarities between individuals

Note: Axis1 and Axis 2 explain 22.14% of total variation

According to the genetic distance between populations calculated by GenAlEx software (Table 4), it was found that the maximum genetic distance was 0.306 2 between SGL and TP3, and the minimum was 0.129 3 between TP2 and TP3. Among them, the genetic distance between *D. huoshanense* and *D. officinale* populations was the largest, with a range of 0.290 1~0.306 2, followed by *D. huoshanense* and *D. fanjingense* populations with a range of 0.214 0~0.2233, followed by *D. huoshanense* and *D. moniliforme* with a range of 0.206 9~0.214 2, while the genetic distance between *D. huoshanense* populations was small, with a range of 0.182 7~0.199 1. *D. moniliforme* and *D. officinale* were 0.259 1～0.279 2, *D. moniliforme* and *D. fanjingense* were 0.201 2, *D. officinale* and *D. fanjingense* were 0.257 2～0.259 9. *D. officinale* populations are the smallest, with a range of 0.129 3~0.133 1. According to the genetic distance, it could be concluded that if the genetic distance between *D. officinale* populations was smaller than that between *D. huoshanense* populations, *D. officinale* has a far genetic relationship with the other three Dendrobium species, and in order of distance, *D. huoshanense*> *D. fanjingense*>*D. moniliforme*.

Table 4 The Genetic distance between populations

	GSS	LYZ	HXY	JXZ	XCZ	JCS	WJH Ⅰ	WJH Ⅱ	SGL	HZB	TP1	TP2	TP3	FJS	XJ
GSS	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LYZ	0.182 7	0	-	-	-	-	-	-	-	-	-	-	-	-	-
HXY	0.196 3	0.194 4	0	-	-	-	-	-	-	-	-	-	-	-	-
JXZ	0.198 0	0.195 5	0.186 9	0	-	-	-	-	-	-	-	-	-	-	-
XCZ	0.192 3	0.191 3	0.187 3	0.189 8	0			-	-	-	-	-	-	-	-
JCS	0.186 6	0.185 9	0.1953	0.196 6	0.192 7	0.0		-	-	-	-	-	-	-	-
WJH Ⅰ	0.196 5	0.194 8	0.188 8	0.190 0	0.190 9	0.197 1	0	-	-	-	-	-	-	-	-
WJH Ⅱ	0.189 5	0.188 1	0.190 7	0.192 7	0.190 7	0.186 5	0.191 7	0	-	-	-	-	-	-	-
SGL	0.187 6	0.187 1	0.197 0	0.198 7	0.195 3	0.187 8	0.195 4	0.189 1	0	-	-	-	-	-	-
HZB	0.187 1	0.187 0	0.197 2	0.199 1	0.195 3	0.188 6	0.194 7	0.188 5	0.186 9	0	-	-	-	-	-
TP1	0.302 1	0.298 2	0.292 9	0.296 7	0.293 2	0.299 4	0.297 6	0.290 1	0.301 9	0.297 1	0	-	-	-	-
TP2	0.304 2	0.300 7	0.294 3	0.298 5	0.294 5	0.300 1	0.301 0	0.292 3	0.305 4	0.300 1	0.132 9	0	-	-	-
TP3	0.305 2	0.301 4	0.295 4	0.299 2	0.295 6	0.301 7	0.302 2	0.293 5	0.306 2	0.301 5	0.133 1	0.129 3	0	-	-
FJS	0.222 9	0.219 6	0.215 7	0.217 6	0.216 4	0.220 4	0.219 1	0.214 0	0.223 3	0.220 8	0.257 2	0.259 1	0.259 9	0	-
XJ	0.214 1	0.211 3	0.209 1	0.211 0	0.209 5	0.212 5	0.212 7	0.206 9	0.214 2	0.212 6	0.277 0	0.259 1	0.279 2	0.201 2	0

2 Discussion

Since the whole genome determination of *D. officinale* and *D. catenatum* in 2015 and 2016, it has provided convenience for the future research on the evolutionary history, genetic relationship, genetic reproduction, gene cloning and identification, gene function and so on of Orchid or Dendrobium Plants. Botanists could use abundant genomic data and genomic tools to better understand the physiological, molecular and genetic mechanisms of Dendrobium plants, and conduct research on comparative genomics, gene editing and so on. For example, Zhang et al. (2016) believed that the extensive ecological niche of Dendrobium is related to the expansion of resistance related genes, and the synthesis of medicinal polysaccharides seems to be related to the extensive replication of the gene encoding glucomannan synthase. In this study, SNP analysis produced by GBS was used to compare the differences and genetic relationships of four Dendrobium species. The results showed that the percentage of *D. officinale* clean reads mapped to *D. catenatum* reference genes was the highest, and the number of SNPs of *D. officinale* was the lowest, suggesting that *D. officinale* and *D. catenatum* had the closest genetic relationship, which was consistent with the research results of Li et al. (2013). Principal coordinates analysis (PCoA) found that *D. moniliforme* and *D. fanjingense* clustered together, suggesting that they were closely related, while *D. officinale*, *D. huoshanense* and *D. moniliforme* were far separated, suggesting that their genetic relationship was far away. Structure analysis showed that when K=2, samples of *D. officinale* were clustered into one branch, *D. moniliforme*, *D. fanjingense* and *D. huoshanense* gathered into another branch, indicating that *D. officinale* was far away from these three kinds of Dendrobium. When K=3, it was divided into three branches, one branch of *D. officinale*, and the rest were divided into two branches. Some *D. huoshanense*, *D. moniliforme* and *D. fanjingense* gather into one branch, and some *D. huoshanense* gather into another branch, which was consistent with the results of Li et al. (2013) that *D. huoshanense* and *D. moniliforme* have a relatively close genetic distance. When K=4, it was divided into four branches, *D. moniliforme* and *D. huoshanense* were separated, *D. moniliforme* and *D. fanjingense* were grouped into one branch, *D. officinale* gather into one branch. *D. huoshanense* was divided into two branches, and there was confusion among the populations of *D. huoshanense*, indicating that *D. moniliforme* and *D. fanjingense* were closely related, and there were certain differences between *D. huoshanense* and *D. moniliforme*. NJ tree analysis showed that *D. officinale*, *D. moniliforme* and *D. fanjingense* were gather into one branch, respectively, and *D. fanjingense* was between *D. officinale* and *D. moniliforme*. Except for one sample, *D. huoshanense* samples were divided into two branches, of which some population samples were divided into two branches, and the same population samples were not completely gathered, that was, the samples between populations were mixed. It was found that *D. huoshanense* was divided into *D. huoshanense* 1 and *D. huoshanense* 2, that was, *D. huoshanense* has two species, which was consistent with the results of this study that that *D. huoshanense* was divided into two branches. However, in this study, it was found that some populations have these two branches, indicating that these two branches were closely related and have a mixed situation. According to the genetic distance analysis, it was confirmed that *D. officinale* was far away from the other three kinds of Dendrobium, and that *D. moniliforme* was close to *D. fanjingense*, *D. moniliforme* and *D. huoshanense*. In addition, there was little difference in genetic distance between *D. huoshanense* populations.

In this study, the differences and genetic relationships of 4 Dendrobium species were analyzed based on SNPs obtained by GBS with the *Dendrobium catenatum* as the reference genome. The results of NJ, Structure and PCoA analysis were similar. *D. officinale* was far away from the other three kinds of Dendrobium. *D. moniliforme* was close to *D. fanjingense* and a branch of *D. huoshanense*. *D. huoshanense* could be divided into two branches in general, but there was confusion between the two branches, and the genetic distance analysis verified the genetic relationship of four species of Dendrobium. This study confirmed that the high throughput SNP markers obtained by GBS could be used to study the genetic relationship of different Dendrobium species or other different plants.

3 Materials and Methods

3.1 Experimental materials

A total of 72 samples were collected from 10 *Dendrobium huoshanense* populations that have been cultivated for more than three years in Huoshan (Table 5). Seven, four and two samples of *D. officinale* were collected in Huoshan, Anhui, Yunnan and Zhejiang, respectively. Three and two samples of *D. moniliforme* were collected in Huoshan and Guangxi, respectively. One sample of *D. fanjingense* was collected in Guizhou. The fresh leaves collected from these samples were stored in liquid nitrogen, and then transported to Frasergen with dry ice for DNA extraction and GBS analysis.

Table 5 Location, sampling size and population code of 4 species of Dendrobium

Species name	Population name	Population abbreviation	Sample size	Location
D.huoshanense	Gaoshanshan	GSS	8	Huoshan, Anhui
	Liyouzhi	LYZ	10	Huoshan, Anhui
	Huxiaoyu	HXY	6	Huoshan, Anhui
	Jiuxianzhai	JXZ	5	Huoshan, Anhui
	Xiancaozhai	XCZ	12	Huoshan, Anhui
	Jiucaoshou	JCS	3	Huoshan, Anhui
	Ⅰ Wanjiahu Ⅰ	WJH Ⅰ	9	Anqing, Anhui
	Ⅱ Wanjiahu Ⅱ	WJH Ⅱ	4	Anqing, Anhui
	Shangelao	SGL	7	Huoshan, Anhui
	Huzhibao	HZB	8	Huoshan, Anhui
D.huoshanense (total)	Huoshan	HS	72	Huoshan, Anhui
D. officinale	Tiepi	TP1	13	Shizong Yunnan (4 samples)
		TP2		Yandang Zhejiang (2 samples)
		TP3		Huoshan Anhui (7 samples)
D. moniliforme	Xijing	XJ	5	Weijiang Guangxi (2 samples)
				Huoshan Anhui (3 samples)
D. Fanjingshanense	Fanjingshan	FJS	1	Fanjingshan Guizhou

3.2 DNA extraction, library construction and GBS analysis

Took 0.1 g of each sample and extracted DNA with the improved CTAB method. 0.8% agarose gel electrophoresis was used for the quality inspection of DNA to detect whether the sample has degradation and the size of DNA fragments. Nanodrop micro spectrophotometer was used to detect the purity of DNA (A260/A280 is between 1.8~2.0). Qubit 3.0 was used to accurately quantify the concentration of DNA. DNA concentration required for the construction of each sample library≧10 μg. The qualified DNA samples were randomly interrupted by Covaris ultrasonic crusher, degraded by Msel enzyme, repaired at the end, added A, sequenced splice, purification, PCR diffusion and other steps to complete the construction of the whole library. The constructed library was sequenced by Illunina sequencer PE15.

3.3 Sequence processing and alignment to the reference genome

The original image data file obtained by the sequencing platform was transformed into the original sequencing sequence through base recognition and analysis, that was, Raw data (All Raw data have been uploaded to NCBI, GenBank: PRJNA659117). Raw data was filtered by the Trimmomatic.0.38 software to remove the sequencing fragments containing joints, the continuous quality of both ends was less than 20 and the length of base was less than 50, and the paired sequencing fragments were retained, that was, the final clean sequencing fragment was called clean data or clean reads. Burrows-Wheeler Alignment tool (BWA) (Li and Durbin, 2009; 2010) was used to map the clean data to *D. catenatum*reference genome. GATK (McKenna et al., 2010) software was used to detect the number of SNPs in clean data, remove MAF<0.01 and SNPs not found in 20% of the samples, and the obtained SNPs could be further used for genetic structure analysis.

3.4 Analysis of population genetic structure and genetic distance

ADMIXTURE software (Alexander et al., 2009) was used for the analysis of population genetic structure, and the optimal K value was the value with the minimum cross validation error. TreeBeST (Vilella et al., 2009) was used for the construction of NJ tree. GCTA software (Yang et al., 2011) was used for PCoA to explore the genetic structure relationship of four Dendrobium species. The genetic distance between populations was calculated with the GenAlEx software to further analyze the genetic relationship of four Dendrobium species.

Authors' Contributions

ZYL is the executor of the experimental design and research of this study. WXP participated in the experimental design, experimental result analysis and data analysis. YMR is the designer and person in charge of the project, guiding experimental design, data analysis, paper writing and revision. All authors read and approved the final manuscript.

Acknowledgements

This study was supported by the Natural Science Foundation Project of Anhui Province (1808085MC78), Talent Stabilization Program of Anhui Science and Technology University (SKWD201601). We would like to thank Frasergen company for its DNA extraction and GBS sequencing analysis of Dendrobium samples, and the Dendrobium cultivator for providing samples of four species of Dendrobium.

References

Alexander D.H., Novembre J., and Lange K., 2009, Fast model-based estimation of ancestry in unrelated individuals, Genome Research, 19(9): 1655-1664.
https://doi.org/10.1101/gr.094052.109
PMid:19648217 PMCid:PMC2752134.

Baral K., Coulman B., Biligetu B., and Fu Y.B., 2018, Genotyping-by-Sequencing enhances genetic diversity analysis of crested wheatgrass [*Agropyron cristatum* (L.) Gaertn.], International Journal of molecular Sciences, 19(9): 2587.
https://doi.org/10.3390/ijms19092587
PMid:30200310 PMCid:PMC6163524.

Duan Y.Y., Lu C., Guo J., Guo X.L., Zhou W.X., and You J.M., 2019, Establishment and optimization of ISSR identification system of *Dendrobium huoshanense* and *Dendrobium officinale*, Fenzi Zhiwu Yuzhong (Molecular Plant Breeding), 17(13): 4360-4366

Kumar S., Banks T.W., and Cloutier S., 2012, SNP discovery through next-generation sequencing and its applications, International Journal of Plant Genomics, 831460: 1-15.
https://doi.org/10.1155/2012/831460
PMid:23227038 PMCid:PMC3512287.

Li D.B., Gao Y.H., Si J.P., and Zhu Y.Q., 2013, Cloning and expression analysis of HSP70 gene from *Dendrobium officinale* under low temperature stress, Zhongguo Zhongyao Zazhi (China Journal of Chinese Materia Medica), 38(20): 3446-3452

Li H., and Durbin R., 2009, Fast and accurate short read alignment with Burrows–Wheeler transform, Bioinformatics, 25(14): 1754- 1760.
https://doi.org/10.1093/bioinformatics/btp324
PMid:19451168 PMCid:PMC2705234

Li H., Durbin R., 2010, Fast and accurate long-read alignment with Burrows–Wheeler transform, Bioinformatics, 26(5): 589-595.
https://doi.org/10.1093/bioinformatics/btp698
PMid:20080505 PMCid:PMC2828108.

Li G.L., Zhang J.X., Zeng S.J., Wu K.L., Duan J., 2013, Systematic position of *Dendrobium huoshanense* inferred from ITS, Nad intron2 and psbA-trnH DNA sequences, Guangdong Nongye Kexue (Guangdong Agricultural Sciences), 40(13): 145-147, 168

Li Y.M., Zang N., Shen X., Li H., Hei X.B., Guo S. X., and Zhang G., 2018, Identification and expression analysis of four CIPKs genes from a rare and medicinal plant, *Dendrobium officinale*, Yaoxue Xuebao (Acta Pharmaceutica Sinica), 53(2): 304-312

Luo Z.N., Brock J., Dyer J.M., Kutchan T., Schachtman D., Augustin M., Ge Y.F., Fahlgren N., and Abdel-Haleem H., 2019, Genetic diversity and population structure of a *Camelina sativa* spring panel, Frontiers in Plant Science, 10(184): 1-12.
https://doi.org/10.3389/fpls.2019.00184
PMid:30842785 PMCid:PMC6391347.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., and DePristo M.A., 2010, The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, Genome Research, 20(9): 1297-1303.
https://doi.org/10.1101/gr.107524.110
PMid:20644199 PMCid:PMC2928508.

Poland J.A., and Rife T.W., 2012, Genotyping-by-sequencing for plant breeding and genetics, The Plant Genome, 5: 92-102.
https://doi.org/10.3835/plantgenome2012.05.0005.

Ren Y., Wang C.D., Xu S.S., Huang S.H., Zhang Y.D., 2013, Analysis of genetic relationship of ornamental Dendrobium by SSR marker. Chinese Journal of Tropical Crops, 34(7): 1252-1256

Ryu J., Kim W.J., Im J.H., Kang K.W., Kim S.H., Jo Y.D., Kang S.Y., Leed J.H., and Ha B.K., 2019, Single nucleotide polymorphism (SNP) discovery through genotyping-by-sequencing (GBS) and genetic characterization of Dendrobium mutants and cultivars, Scientia Horticulture, 244: 225-233. https://doi.org/10.1016/j.scienta.2018.09.053.

Si H.Y., Chen N.D., and Chen N.F., 2016, Review on isolation and pharmacological activities of polysaccharides from *Dendrobium huoshanense* C.Z.Tang et S.J.Cheng, Tianran Chanwu Yanjiu yu Kaifa (Natural Product Research and Development), 28(3): 467-470,445

Song G.Q., Liu X.M.,Wang Q., and Liu J., 2014, Research progress on pharmacological activities of medical plants from Dendrobium Sw., Zhongcaoyao (Chinese Traditional and Herbal Drugs), 45(17): 2576-2580

Song S., Zhou Y.F., Liu Z.J., Zhao M.F., Yang J., Xu S.Z., and Wen G.S., 2016, Genetic diversity among Dendrobium with known origins based on the ISSR and AFLP markers, Yunnan Nongye Daxue Xuebao (Journal of Yunnan Agricultural University (Natural Science Edition)), 31(4): 688-695

Vilella A.J., Severin J., Ureta-Vidal A., Heng L., Durbin R., and Birney E., 2009, Ensembl compara genetrees: complete, duplication-aware phylogenetic trees in vertebrates, Genome Research, 19(2): 327-335.
https://doi.org/10.1101/gr.073585.107
PMid:19029536 PMCid:PMC2652215.

Yan L., Wang X., Liu H., Tian Y., Lian J. M., Yang R.J., Hao S.M., Wang X.J., Yang S.C., Li Q.Y., Qi S., Kui L., Okpekum M., Ma X., Zhang J.J., Ding Z.L., Zhang G.J., Wang W., Dong Y., and Sheng J., 2015, The genome of *Dendrobium officinale* illuminates the biology of the important traditional Chinese orchid herb, Molecular Plant, 8(6): 922-934.
https://doi.org/10.1016/j.molp.2014.12.011
PMid:25825286

Yang J., Lee S.H., Goddard M.E., and Visscher P.M., 2011, GCTA: a tool for genome-wide complex trait analysis, American Journal Human Genetics, 88(1): 76-82.
https://doi.org/10.1016/j.ajhg.2010.11.011
PMid:21167468 PMCid:PMC3014363.

Zhang G.Q., Xu Q., Bian C., Tsai W.C., Yeh C.M., Liu K.W., Yoshida K., Zhang L.S., Chang S.B., Chen F., Shi Y., Su Y.Y., Zhang Y.Q., Chen L.J., Yin Y., Lin M., Huang H., Deng H., Wang Z.W., Zhu S.L., Zhao X., Deng C., Niu S.C., Huang J., Wang M., Liu G.H., Yang H.J., Xiao X.J., Hsiao Y.Y., Wu W.L., Chen Y.Y., Mitsuda N., Ohme-Takagi M,. Luo Y.B., Van de Peer Y., and Liu Z.J., 2016, The *Dendrobium catenatum* Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. Scientific Reports, 6: 19029.
https://doi.org/10.1038/srep19029
PMid:26754549 PMCid:PMC4709516.