

# Development of Genic Simple Sequence Repeat Panels for Population Classification of Chinese *Cymbidium* Species

Xiaobai Li<sup>1,2</sup>

Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China; and International Atomic Energy Agency Collaborating Center, Department of Applied Bioscience, Zhejiang University, Hangzhou, 310029, China

Weirui Li<sup>1</sup> and Chenlu Di<sup>1</sup>

International Atomic Energy Agency Collaborating Center, Department of Applied Bioscience, Zhejiang University, Hangzhou, 310029, China

Ming Xie and Liang Jin

Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China

Cheng Huang

Shaoxing County Agricultural Technology Extension Stations, Shaoxing 312000, China

Dianxing Wu

International Atomic Energy Agency Collaborating Center, Department of Applied Bioscience, Zhejiang University, Hangzhou, 310029, China

**ADDITIONAL INDEX WORDS.** microsatellite, population structure, population assignment, discriminatory power

**ABSTRACT.** Chinese cymbidiums (*Cymbidium* sp.) are important ornamental plants because of their foliage, flower shape, and fragrance. Well-known Chinese cymbidiums mainly include *Cymbidium goeringii*, *Cymbidium faberi*, *Cymbidium ensifolium*, *Cymbidium kanran*, and *Cymbidium sinense*. The population genetics of Chinese cymbidiums can be efficiently analyzed using small-scale marker panels with high discriminatory power. In this study, we tested several genic simple sequence repeats (SSRs) and built six genic SSR panels. The panels included several robust markers, which can rapidly assign Chinese cymbidium accessions to their source species. Fifty-three accessions of Chinese cymbidiums were analyzed using 25 markers, which exhibited polymorphism among five species. These markers were ranked according to their discriminatory scores (*D* scores). The program selected six markers to build an “overall” panel for all *Cymbidium* classifications and yielded 95.16% population assignment accuracy. Considering one species as the “critical” population and the four other species as one population, we built five genic SSR panels: *C. ensifolium* panel (four markers, 98.05% accuracy), *C. faberi* panel (six markers, 95.90% accuracy), *C. goeringii* panel (six markers, 95.15% accuracy), *C. sinense* panel (six markers, 96.35% accuracy), and *C. kanran* panel (five markers, 96.10% accuracy). Genetic distance matrices calculated using the “overall” panels and those derived with the 25 markers were compared. Results showed a high correlation ( $R = 0.807$ ) with statistical significance ( $P = 0.042$ ). Moreover, “all panels” revealed higher genetic variations among populations than “all markers.” Hence, the developed panels are suitable for efficient population classification of Chinese cymbidiums.

Chinese cymbidiums refer to species and hybrids derived from *Cymbidium goeringii*, *C. faberi*, *C. ensifolium*, *C. kanran*, and *C. sinense*. These are terrestrial species and most cultivars of these species present superior flower shape, foliage, and fragrance (Li et al., 2014c; Schneitz, 1999). Chinese cymbidiums have been cultivated for several centuries in Asia, especially in China, Korea, and Japan, because of the high ornamental and economic value of these plants (Moea and Parka, 2012). To conserve the genetic resources of Chinese cymbidiums, scholars have focused on conserving the genetic

diversity of these species. Genetic diversity and population structure have to be preliminarily characterized before germplasm conservation. The genetic diversity of *Cymbidium* species can be analyzed using DNA markers, including restriction enzyme polymorphism markers (Obara-Okeyo and Kako, 1998); random amplified polymorphic DNA markers (Choi et al., 2006; Obara-Okeyo and Kako, 1998); amplified fragment length polymorphism markers (Wang et al., 2004); polymorphisms of internal transcribed spacers of nuclear ribosomal DNA and plastids, as well as inter-SSR (Lu et al., 2011); and SSR (Capesius, 1976; Moe et al., 2010; Moea and Parka, 2012). Novel *Cymbidium* SSRs derived from genes (genic SSRs) are developed from genic regions and can be applied for genetic analysis (Li et al., 2013, 2014a). SSRs are ubiquitous, locus specific, codominant, highly polymorphic, and easily operated, thereby requiring lower cost compared with other molecular markers (Honma and Goto, 2001;

Received for publication 20 Nov. 2015. Accepted for publication 30 Dec. 2015. This research was supported by the National Basic Research Program funded by the Nature Science Foundation of China (no. 31201648) and the Qianjiang Talents Project (no. 2013R10081).

<sup>1</sup>These authors contributed equally to this work.

<sup>2</sup>Corresponding author. E-mail: hufanfan1982815@gmail.com.

Li et al., 2006; Schwarz-Sommer et al., 1990; Tsai et al., 2004; Venkateswarlu et al., 2006; Xu et al., 2006). In addition, genic SSRs exhibit high transferability across species. These features render genic SSRs as suitable for discriminating closely related species.

The use of numerous marker loci in population genetic analysis is an expensive and inefficient method for genotyping large germplasm collections. As such, scholars developed informative panels by using a minimum number of marker locus combinations. The panel can be considered a DNA bar code and can be used for research on taxonomy and population genetics. Minimum combination of highly ranked loci for population assignment can be determined through two basic procedures. First, simulated test data are generated using the observed allele frequencies for each locus, and then the number of times each test genotype is correctly assigned to the appropriate source population is determined and markers are then ranked by comparison of individual marker scores. Second, another round of iterations yield ranked loci by increasing one locus at each run until the assignment score reaches or exceeds the accuracy criterion. This strategy is applied for locus selection in numerous species during population assignment; such species include chinook salmon [*Oncorhynchus tshawytscha* (Greig et al., 2003)], asian seabass [*Lates calcarifer* (Yue et al., 2012)], trematomids [*Trematomus* sp. (Van de Putte et al., 2009)], humans [*Homo sapiens* (Rosenberg et al., 2003)], switchgrass [*Panicum virgatum* (Okada et al., 2011)], and rice [*Oryza sativa* (Agrama et al., 2012)].

In the present study, we aim to develop informative panels by using a minimum number of genic SSRs. The panels can be used to assign Chinese cymbidiums to five populations with high accuracy. The proposed strategy could be an efficient tool for population classification.

## Materials and Methods

### Plant samples

Fifty-three accessions from *C. goeringii*, *C. faberi*, *C. ensifolium*, *C. kanran*, and *C. sinense* were used for testing genetic markers (Table 1). Fifty-three accessions were selected from 105 accessions (Supplemental Fig. 1) based on their ancestries derived from STRUCTURE (Pritchard et al., 2000), which had maximum ancestry index higher than 0.6. The plants were grown and maintained in a greenhouse under natural light. Fresh leaves were collected from two or three seedlings of each accession for genomic DNA extraction.

### Genotyping

Genomic DNA was extracted from leaf samples as previously described (Li et al., 2007). Twenty-five genic SSRs were selected from the genic SSR collection (Li et al., 2014a) for panel development. Polymerase chain reaction (PCR) primers were synthesized by Life Technologies (ABI and Invitrogen, Shanghai, China). PCR experiments were conducted, and the products were separated using polyacrylamide gel electrophoresis gel (Li et al., 2007). The gel was silver stained according to the procedure reported by Li et al. (2001), and the results were recorded using a scanner. The genotype was determined by analyzing band patterns (Li et al., 2014a). The accessions were labeled by 1, 2, 3, and so on, according to the number and position of bands, and then the data were converted

Table 1. Fifty-three Chinese cymbidium accessions for genetic analysis. *Cymbidium ensifolium*, *Cymbidium faberi*, *Cymbidium sinense*, *Cymbidium goeringii*, and *Cymbidium kanran* had 10, 10, 10, 12, and 11 accessions, respectively.

Species	Name	Species	Name
<i>C. ensifolium</i>	Bailian	<i>C. goeringii</i>	Chunjiansuxin
	Baodao		Hexingmei
	Jinhe		Hongmeigui
	Minnanlvdie		Huangmei
	Qingshanyuquan		Jiuxianmudan
	Shisantaibao		Puchunhong
	Tiegusu		Silan
	Yandongsu		Songmei
	Yinsimawei		Wenyi
	Zhongshanmei		Yipin
<i>C. faberi</i>	Dayipin	<i>C. kanran</i>	Yuhudie
	Duanhuimei		Yumei
	Jiepeimei		Changuihong
	Laoranzi		Danxia
	LvlanA		Dongmingmei
	LvlanB		Jihanmei
	LvlanC		Jinwan
	Panlv		Mengzhongmeiren
	Xianhejipin		Xianhe
	Xianlv		Yingcaihe
<i>C. sinense</i>	Baimo		Yipinhong
	Damo		Yisudie
	Fuguiminglan		Zhizunhongyan
	Piaoyi		
	Qiebang		
	Qihei		
	Taoji		
	Wuzicui		
	Yingbianmolan		
	Yutao		

into binary matrix using “compute frequency” procedure in PowerMarker version 2.7 (Liu and Muse, 2005).

### Statistical analysis

**POLYMORPHISM OF MARKERS.** Genetic distances were calculated using Nei’s distance (Nei and Takezaki, 1983). Phylogenetic reconstruction was performed based on neighbor-joining (NJ) method (with 1000 bootstrap replicates) and implemented in PowerMarker version 2.7. The phylogenetic tree was visualized in MEGA version 4 (Tamura et al., 2007). PowerMarker was also used to calculate the average number of marker alleles and polymorphism information content (PIC).

**ANALYSIS OF POPULATION STRUCTURE.** Population structure was initially analyzed using the software STRUCTURE as described in previous studies (Li et al., 2010, 2011, 2012, 2014b). Genetic variations within and among populations were then calculated by analysis of molecular variance (AMOVA) in Arlequin V2.000 (Schneider and Excoffier, 1999). Genetic distances among species were calculated using Nei’s distance (Nei and Takezaki, 1983). A phylogenetic tree was constructed by NJ method. Both calculations were performed in PowerMarker version 2.7. The phylogenetic tree was visualized using MEGA version 4 (Tamura et al., 2007). The correlation of genetic distances based on total markers or panels was assessed using Mantel test in PowerMarker.

**DISCRIMINATORY POWER OF MARKERS.** Twenty-five genic SSRs were employed for WHICHLOCI analysis. Accessions with clear origins were used to test discriminatory powers. Using the WHICHLOCI v.1.0 software package (Banks et al., 2003), we ranked each of the 25 markers according to their *D* scores, which indicate discriminatory powers in population analysis. Discriminatory power was assessed by conducting 10,000 simulations (population size  $N = 2000$ ) based on the allele frequency of each population. The threshold of the assignment accuracy was set to 95% and the stringency of limit of detection (LOD) was designated as 3.0 for all simulations. For the “critical” population, we determined which locus is necessary for identification of a specific species. These methods were applied in each species, where all species that do not belong to the “critical” population were treated as one population. Finally, reliability examination was performed on panel markers by comparing the results derived from panels with those obtained using all markers.

## Results

**POPULATION STRUCTURE AND GENIC SSR PROFILE.** Fifty-three of 105 accessions were selected based on their ancestries derived from STRUCTURE (Table 1). Accessions with severely mixed ancestries (maximum ancestry index lower than 0.6) were excluded. A total of 12, 11, 10, 10, and 10 accessions from *C. goeringii*, *C. kanran*, *C. faberi*, *C. ensifolium*, and *C. sinense*, respectively, were used for further analysis. The AMOVA results revealed 19.98% genetic variation among species and 80.02% genetic variation within species. Polygenetic analysis also indicated that the largest genetic distance was found between *C. ensifolium*

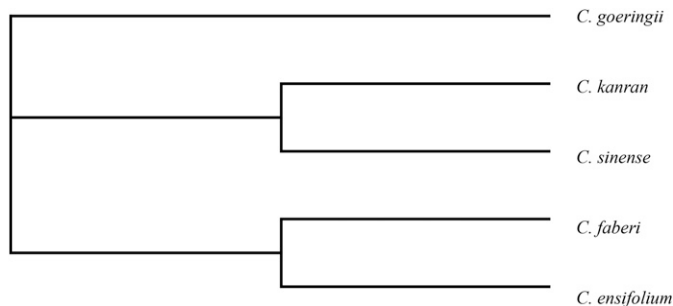


Fig. 1. Neighbor-joining tree of five Chinese cymbidiums based on Nei's distance. These species included *Cymbidium goeringii*, *Cymbidium faberi*, *Cymbidium ensifolium*, *Cymbidium kanran*, and *Cymbidium sinense*. Bootstrap value (out of 1000) is indicated at the branch point.

and *C. sinense* (0.2734), and the smallest distance was detected between *C. goeringii* and *C. kanran* (0.2008) (Fig. 1; Table 2).

In the genic SSR collection, 25 genic SSRs exhibited polymorphism within the five species. These SSRs were selected to test assignment ability because equal numbers of effective markers (polymorphic markers) are required within each population for WHICHLOCI. The polymorphisms of the markers are shown in Table 3. The PIC values varied from 0.204 for SSR68 to 0.855 for SSR73, with an average content of 0.528. The allele number of these markers ranged from two for SSR79 to 14 for SSR08, with an average value of 6.88.

**EVALUATION OF MARKER DISCRIMINATORY POWER AND CONSTRUCTION OF PANELS.** *D* scores were used to measure the ability of markers for correct assignment of accessions to their original species. Among the 25 markers, SSR73 showed the highest assignment score (4.941), whereas SSR68 displayed the lowest score (0.941) (Fig. 2; Table 4). The correlation coefficients between *D* scores and allelic numbers and those between *D* scores and PICs were 0.795 and 0.924, respectively. With increasing number of markers employed, high portions of simulated individuals were accurately assigned (Fig. 2). The percentage of correct assignments markedly increased in response to the first six markers and then gradually increased upon successive addition of the seventh to the 10th markers. The percentage then reached 99.31%, and the “overall” panel was generated by WHICHLOCI without defining the critical population. The panel included six markers: SSR73, SSR21, SSR08, SSR64, SSR53, and SSR03. In this panel, the *D* scores of the markers varied from 3.57 for SSR03 to 4.94 for SSR73, with an average value of 4.15 (Table 4). About 95.16% of the simulated accessions were accurately assigned to their original species by using the most effective markers, which were selected based on *D* scores at an LOD threshold of 3.0 (Table 4).

In critical population method, other populations were treated as one population. When *C. ensifolium* was selected as the critical population, a panel was created; the developed panel generated 98.05% correct assignments. The panel included four markers with *D* scores ranging from 2.07 to 2.67 (Table 4). The *C. faberi* panel consisted of six markers, with *D* scores ranging from 1.71 to 2.77, and resulted in 95.90% correct assignments. The *C. goeringii* panel comprised six markers, with *D* scores ranging from 1.52 to 2.67, and resulted in 95.15% correct assignments. The *C. sinense* panel was composed of six markers, with *D* scores ranging from 1.79 to 2.57, and resulted in 96.35% correct assignments. The *C. kanran* panel included five markers, with *D* scores ranging from 1.94 to 2.79, and resulted in 96.10% correct assignments.

Table 2. Comparison of pairwise genetic distance above the diagonal based on the “overall” panel comprising six markers and below the diagonal based on all 25 markers among five Chinese *Cymbidium* species.<sup>z</sup>

Species	<i>Cymbidium ensifolium</i>	<i>Cymbidium faberi</i>	<i>Cymbidium goeringii</i>	<i>Cymbidium sinense</i>	<i>Cymbidium kanran</i>
<i>C. ensifolium</i>	0.0000	0.5323	0.5603	0.6421	0.5349
<i>C. faberi</i>	0.2081	0.0000	0.5374	0.6392	0.5654
<i>C. goeringii</i>	0.2113	0.2258	0.0000	0.5554	0.5280
<i>C. sinense</i>	0.2734	0.2619	0.2511	0.0000	0.5690
<i>C. kanran</i>	0.2250	0.2097	0.2008	0.2145	0.0000

<sup>z</sup>Genetic distances between species assessed by overall panel comprising six markers with most discriminatory power were highly and significantly correlated with those assessed by all 25 markers (Mantel test:  $R = 0.807$ ,  $P = 0.042$ ).

Table 3. Twenty-five polymorphic genic simple sequence repeats (SSRs) applied for analysis of 53 Chinese cymbidium accessions.

Name	Primer	Alleles	PIC <sup>z</sup>	D score <sup>z</sup>
SSR01	F: AACGCCATGTCCAATACCC R: GGAGGGCTTATTTGCAGCG	5	0.510	3.119
SSR03	F: CTCGGTTCATTTGCAGCCC R: GGGTGGGTATGGCGAAATC	6	0.689	3.571
SSR04	F: AGAATCTGCCAACCTTGATAC R: GCAGATGCCAGTTAGAATGGG	5	0.662	3.024
SSR08	F: TGCTGGAATACATGCGAGAC R: GTTTGCCGAAGCCAGTGC	14	0.768	4.512
SSR11	F: AACTGACAAGCATCTGCAAG R: CTGCTGCATTGGCCTTACC	6	0.458	1.869
SSR14	F: CACAGCAGCTCACAATCCTG R: TACAGCCCTGTTTACCGCC	6	0.441	1.488
SSR16	F: GCCCACAGCAATCCATCTG R: GCAGTCGAAGAAACCGTGG	6	0.301	1.143
SSR17	F: GGATCACCAACAGCATGGG R: TCCACCAAGAGCAAGGATG	4	0.418	1.631
SSR18	F: TGAAACGGTTGGCTCTAGTTC R: AGCAAGCACTGACCTGAAAC	9	0.468	2.881
SSR21	F: TGGGCGACAGATCGAGTTC R: ACATGGACCACAGCATTC	13	0.777	4.512
SSR22	F: TATGCGTCTCTCCCAACCG R: AAGCTAGTGGCCTTTGGTG	10	0.625	2.857
SSR26	F: GCTTTATGCGACATCTGCG R: CGTCGGTTCCATGCACATC	8	0.612	3.250
SSR42	F: GACGACATATCGCGTTCGG R: CTCAGCCACACCCAAGAGG	6	0.558	2.767
SSR43	F: GGAGCTGCATACGCAAGTG R: AGCTTCTCACTGCCTCCAG	7	0.572	2.095
SSR45	F: GCCTTACACATCCCTTCCAAC R: TGCCTGCTGATAGTTTGCC	5	0.374	1.500
SSR49	F: AGAGGGCCACCTGCTTTC R: GCCAATTGCCAGATGGACG	5	0.452	2.417
SSR53	F: GCTGAAGTTCCGGTCCTC R: TCCGCCTCTTTAAGCCGAC	8	0.698	3.643
SSR54	F: ATCTTCCCTCCACATCGGC R: TGGAGAAGAGTCGACCAGC	4	0.387	1.298
SSR64	F: CGCTCAAAGAGATGGCACG R: TAGTACGGCGCTGCTTGAG	10	0.644	3.750
SSR68	F: CCAGATCGAATGGCTACGC R: CAAGGAGCTCGTCGAAGG	4	0.204	0.941
SSR70	F: CCAACGCAGAACGATAGC R: CGGTGGCACAATGGAACG	5	0.495	3.036
SSR73	F: GGACACAATGGAGACGAAGG R: TGCATGAAACCACATGGC	13	0.855	4.941
SSR76	F: AGACAGAGAGTCCCTAAAGGC R: CAGGGATGTTAAGTGGGCTG	6	0.523	2.738
SSR77	F: TTTGTGGCAGTGGAAGCG R: TGATACCAATGGCAAGCG	5	0.470	2.550
SSR79	F: AGGATTCATGTAGCCGACCTC R: TCCCTGAAGGAGGCAAACC	2	0.226	1.548

<sup>z</sup>PIC = polymorphism information content; D score = discriminatory score.

**APPLICATION OF MARKER PANELS IN POPULATION STRUCTURE ANALYSIS.** The discriminatory powers of marker panels in the population assignment were assessed. Genetic variation among the five populations derived using the “overall” panel was 29.92%, whereas that obtained using all marker was 19.98%. In critical population analysis for each Chinese cymbidium

species, genetic variation between *C. ensifolium* and the four other populations derived using the *C. ensifolium* panel was 30.46%, which was higher than the 14.00% variation obtained using all markers. Similar results were observed in the analysis of the four other species (Table 5). In addition, pairwise genetic distances between species derived from each panel or all markers were compared through Mantel analysis. The results showed a significant correlation, with a coefficient of 0.807 ( $P = 0.042$ ), between the two techniques (Table 2). This high correlation suggested that the panels developed can be effectively used to reveal population differentiation patterns.

## Discussion

**GENIC SSRs AS MARKERS FOR ANALYSIS OF POPULATION DIFFERENTIATION.** Genic SSRs show higher transferability across related species than anonymous SSR markers (Scott et al., 2000). Genic SSRs can be used to directly compare different taxa and prevent the risk of locus-specific differences (genetic diversity among accessions), which may mask true species-level differences (Ellis and Burke, 2007). As such, genic SSRs are suitable for discriminating species. In the present study, genic SSRs were used to create highly efficient panels for Chinese cymbidiums. As the complex relationships between accessions can influence the discriminating power of DNA markers (Hayes et al., 2005), accessions with ambiguous ancestries were not included in marker selection. Finally, 53 accessions with clear sources were analyzed in WHICHLOCI. To avoid bias associated with the number of markers used within each species, we selected 25 markers that showed polymorphisms within the five species for further analysis.

**FACTORS INFLUENCING THE DISCRIMINATORY POWER OF MARKERS.** The informative value of markers is attributed to their allelic number and PIC (Agrama et al., 2012; Yue et al., 2012). The discriminatory power of markers is determined by their efficiency in correct population assignments and is inversely correlated with their propensity for causing false assignments (Banks et al., 2003).

Successive assignment trials were performed in simulated populations based on the allele frequencies of markers (Banks et al., 2003). Thus, the *D* score of a marker is logically and positively associated with the corresponding genetic diversity, PIC, and number of alleles. In the “overall” panel analysis, six markers with high *D* scores presented high PICs and high allele numbers. Similarly, SSR68, SSR54, and SSR16, which showed low *D* scores, corresponded to low polymorphism and diversity. The correlation coefficient between *D* score and PIC and that between *D* score and number of alleles are 0.795 and 0.924, respectively. These results are consistent with those reported in a previous study (Agrama et al., 2012).

The order sorted by *D* scores and PICs was not exactly the same. This difference may be due to allelic distribution among

and within populations. In critical population analysis for each species, specific markers containing a minimum of four or five alleles would be ranked as high for particular species, such as in the case of SSR01 in *C. goeringii* panel, which amplified *C. goeringii*-specific alleles. Similar results were also reported previously (Banks et al., 2003). *C. goeringii* was the least differentiated species in this study (Table 5) and required six markers for accurate assignment. This high requirement could be due to the polyphyletic quality of *C. goeringii* accessions. For example, numerous *C. goeringii* accessions are clustered with *C. faberi* or *C. sinense*, as reported in previous studies (Li et al., 2014a, 2014b).

**EVALUATION AND IMPROVEMENT OF PANELS FOR POPULATION CLASSIFICATION.** To evaluate the capability of panels for population identification, we compared the results obtained using panels with those derived from total markers. Both results showed similar population structure patterns, as indicated by the high correlation coefficient ( $R = 0.8068$ ) among genetic distance matrices. Nevertheless, the panels are considered more appropriate for population assignment because of their potential to present clear genetic differentiation with few markers. This ability could be attributed to the allelic distribution of panel markers and is highly consistent with the population structure. Among the “critical” panels, the *C. ensifolium* panel was found to be the most efficient because it can explain the highest variation among populations with the least number of markers (four). The *C. goeringii* panel explained relatively low variation and used the most markers (six). Considering the possibility of misassignment and uncertainty of the genetic background, we proposed an approach for correcting the population structure, that is, several reference accessions are included when using the panels for population analysis.

Genic SSR panels with the strongest discriminatory power were used to characterize Chinese cymbidium germplasm and explore their population structure. In practice,

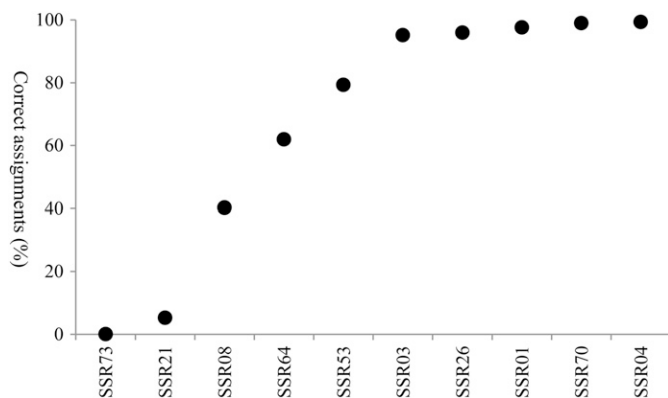


Fig. 2. Percentage of correct assignments and corresponding number of selected simple sequence repeats (SSRs). SSRs were ranked according to their discriminatory scores (*D* scores) from high to low, as calculated by WHICHLOCI (Banks et al., 2003). *D* scores was assessed by conducting 10,000 simulations (population size  $N = 2000$ ) based on the allele frequency of each species. The threshold of the assignment accuracy was set to 95%, and the stringency of limit of detection was designated as 3.0 for all simulations.

Table 4. Estimation of correct assignments of the five *Cymbidium* species and “critical” species by using “overall” or “critical” panels.

Marker	Panels (including markers) with the corresponding <i>D</i> scores <sup>z</sup>					
	Overall	<i>Cymbidium ensifolium</i>	<i>Cymbidium faberi</i>	<i>Cymbidium goeringii</i>	<i>Cymbidium sinense</i>	<i>Cymbidium kanran</i>
SSR01				1.57	1.79	
SSR03	3.57			1.52	2.57	2.26
SSR08	4.51	2.67	2.13	2.05	2.18	1.94
SSR21	4.51	2.07	2.77	2.17	2.01	2.60
SSR22			1.83			
SSR53	3.64		2.25		1.81	2.79
SSR64	3.75	2.67		1.58		
SSR73	4.94	2.27	2.58	2.67	2.36	2.05
SSR76			1.71			
Mean of <i>D</i> score	4.15	2.42	2.31	2.00	2.19	2.33
Correct assignment (%)	95.16	98.05	95.90	95.15	96.35	96.10

<sup>z</sup>*D* score = discriminatory score.

Table 5. Genetic variation among *Cymbidium* species explained by panels and all 25 markers (genic simple sequence repeats).

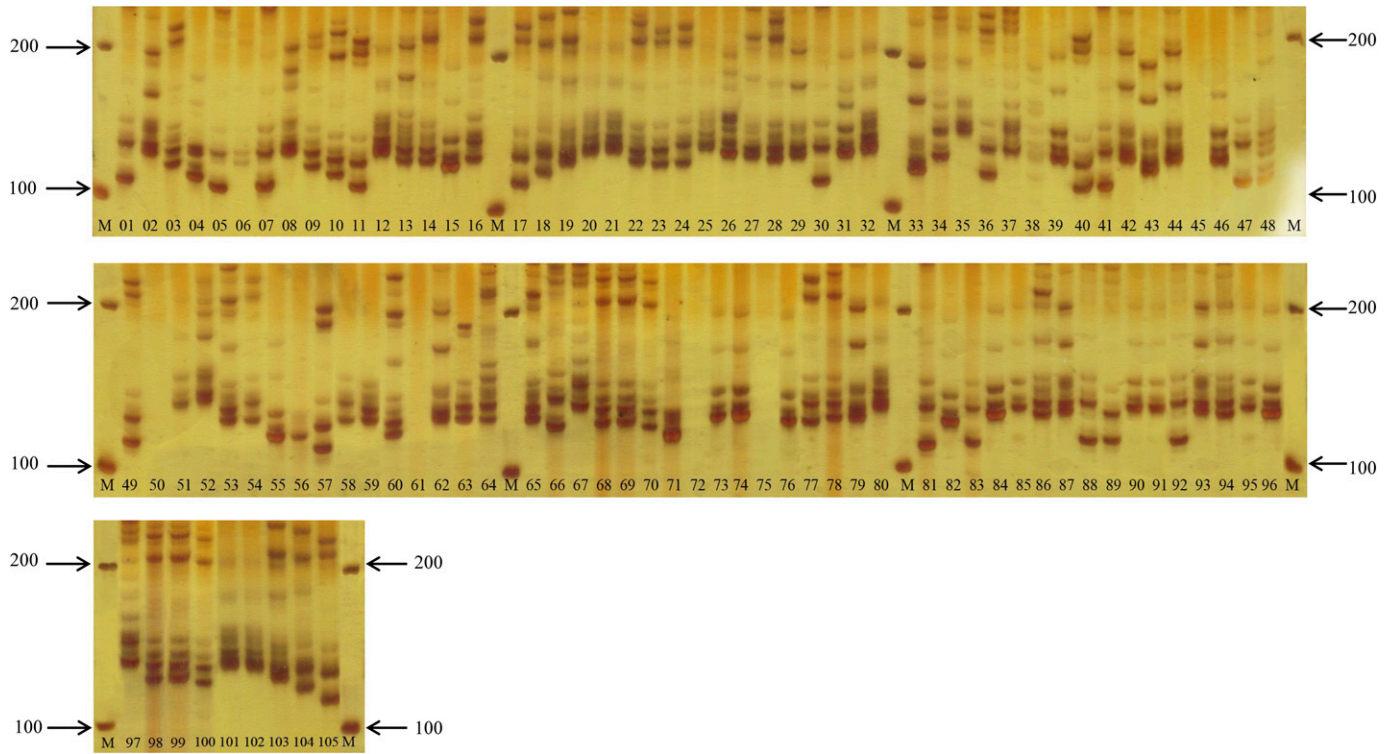
Marker combination	Genetic variation (%)					
	Overall	<i>Cymbidium ensifolium</i>	<i>Cymbidium faberi</i>	<i>Cymbidium goeringii</i>	<i>Cymbidium sinense</i>	<i>Cymbidium kanran</i>
Panels <sup>z</sup>	29.92	30.46	21.96	14.64	18.20	20.05
All 25 markers	19.98	14.00	12.82	10.86	13.37	15.66

<sup>z</sup>Markers were selected by WHICHLOCI (Banks et al., 2003) based on their discriminatory powers in population assignment.

these panels can facilitate rapid screening of large germplasm banks and species classification at a relatively low cost.

### Literature Cited

- Agrama, H.A., A.M. McClung, and W.G. Yan. 2012. Using minimum DNA marker loci for accurate population classification in rice (*Oryza sativa* L.). *Mol. Breed.* 29:413–425.
- Banks, M.A., W. Eichert, and J.B. Olsen. 2003. Which genetic loci have greater population assignment power? *Bioinformatics* 19:1436–1438.
- Capesius, I. 1976. Isolation and characterization of native AT-rich satellite DNA from nuclei of the orchid *Cymbidium*. *FEBS Lett.* 68:255–258.
- Choi, H., M.J. Kim, J.S. Lee, and K.H. Ryu. 2006. Genetic diversity and phylogenetic relationships among and within species of oriental *cymbidiums* based on RAPD analysis. *Sci. Hort.* 108:79–85.
- Ellis, J.R. and J.M. Burke. 2007. EST-SSRs as a resource for population genetic analyses. *Heredity* 99:125–132.
- Greig, C., D.P. Jacobson, and M.A. Banks. 2003. New tetranucleotide microsatellites for fine-scale discrimination among endangered chinook salmon (*Oncorhynchus tshawytscha*). *Mol. Ecol. Notes* 3:376–379.
- Hayes, B., A.K. Sonesson, and B. Gjerde. 2005. Evaluation of three strategies using DNA markers for traceability in aquaculture species. *Aquaculture* 250:70–81.
- Honma, T. and K. Goto. 2001. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409:525–529.
- Li, X., H. Cui, and M. Zhang. 2006. Molecular markers derived from EST: Their development and applications in comparative genomics. *Biodiversity Sci.* 14:541–547.
- Li, X., H. Cui, and M. Zhang. 2007. Detecting the genetic diversity of *Brassica napus* by EST-SSRs. *J. Agron. Biotechnol.* 15:661–667.
- Li, X., F. Jin, L. Jin, A. Jackson, C. Huang, K. Li, and X. Shu. 2014a. Development of *Cymbidium ensifolium* genic-SSR markers and their utility in genetic diversity and population structure analysis in cymbidiums. *BMC Genet.* 15:124.
- Li, X., J. Luo, T. Yan, L. Xiang, F. Jin, D. Qin, C. Sun, and M. Xie. 2013. Deep sequencing-based analysis of the *Cymbidium ensifolium* floral transcriptome. *PLoS One* 8:e85480.
- Li, X., L. Xiang, Y. Wang, J. Luo, C. Wu, C. Sun, and M. Xie. 2014b. Genetic diversity, population structure, pollen morphology and cross-compatibility among Chinese cymbidiums. *Plant Breed.* 133:145–152.
- Li, X., W. Xu, M. Chowdhury, and F. Jin. 2014c. Comparative proteomic analysis of labellum and inner lateral petals in *Cymbidium ensifolium* flowers. *Intl. J. Mol. Sci.* 15:19877–19897.
- Li, X., W. Yan, H. Agrama, B. Hu, L. Jia, M. Jia, A. Jackson, K. Moldenhauer, A. McClung, and D. Wu. 2010. Genotypic and phenotypic characterization of genetic differentiation and diversity in the USDA rice mini-core collection. *Genetica* 138:1221–1230.
- Li, X., W. Yan, H. Agrama, L. Jia, A. Jackson, K. Moldenhauer, K. Yeater, A. McClung, and D. Wu. 2012. Unraveling the complex trait of harvest index with association mapping in rice (*Oryza sativa* L.). *PLoS One* 7:e29350.
- Li, X., W. Yan, H. Agrama, L. Jia, X. Shen, A. Jackson, K. Moldenhauer, K. Yeater, A. McClung, and D. Wu. 2011. Mapping QTLs for improving grain yield using the USDA rice mini-core collection. *Planta* 234:347–361.
- Li, Z.L., R.S. Jakkula, J.P. Hussey, and H.R. Boerma. 2001. SSR mapping and confirmation of the QTL from PI96354 conditioning soybean resistance to southern root-knot nematode. *Theor. Appl. Genet.* 103:1167–1173.
- Liu, K. and S.V. Muse. 2005. PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* 21:2128–2129.
- Lu, J., X. Hu, J. Liu, and H. Wang. 2011. Genetic diversity and population structure of 151 *Cymbidium sinense* cultivars. *J. Hort. For.* 3:104–114.
- Moe, K.T., W. Zhao, H.S. Song, Y.H. Kim, J.W. Chung, Y.I. Cho, P. Park, H.S. Park, S.C. Chae, and Y.J. Park. 2010. Development of SSR markers to study diversity in the genus *Cymbidium*. *Biochem. Syst. Ecol.* 38:585–594.
- Moea, K.T. and Y. Parka. 2012. Analysis of population structure revealed apparent genetic disturbance in Korea *Cymbidium* collection. *Sci. Hort.* 134:157–162.
- Nei, M. and N. Takezaki. 1983. Estimation of genetic distances and phylogenetic trees from DNA analysis. *Proc. 5th World Congr. Genet. Appl. Livestock Prod.* p. 405–412.
- Obara-Okeyo, P. and S. Kako. 1998. Genetic diversity and identification of *Cymbidium* cultivars as measured by random amplified polymorphic DNA (RAPD) markers. *Euphytica* 99:95–101.
- Okada, M., C. Lanzatella, and C.M. Tobias. 2011. Single-locus EST-SSR markers for characterization of population genetic diversity and structure across ploidy levels in switchgrass (*Panicum virgatum* L.). *Genet. Resources Crop Evol.* 58:919–931.
- Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Rosenberg, N.H., L.M. Li, R. Ward, and J.K. Pritchard. 2003. Informativeness of genetic markers for inference of ancestry. *Amer. J. Hum. Genet.* 73:1402–1422.
- Schneider, S. and L. Excoffier. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: Application to human mitochondrial DNA. *Genetics* 152:1079–1089.
- Schneitz, K. 1999. The molecular and genetic control of ovule development. *Curr. Opin. Plant Biol.* 2:13–17.
- Schwarz-Sommer, Z., P. Huijser, W. Nacken, H. Saedler, and H. Sommer. 1990. Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* 250:931–936.
- Scott, K.D., P. Eggler, G. Seaton, M. Rossetto, E.M. Ablett, L.S. Lee, and R.J. Henry. 2000. Analysis of SSRs derived from grape ESTs. *Theor. Appl. Genet.* 100:723–726.
- Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24:1596–1599.
- Tsai, W.C., C.S. Kuoh, M.H. Chuang, W.H. Chen, and H.H. Chen. 2004. Four DEF-like MADS box genes displayed distinct floral morphogenetic roles in *Phalaenopsis* orchid. *Plant Cell Physiol.* 45:831–844.
- Van de Putte, A.P., J.K.J. Van Houdt, G.E. Maes, K. Janko, P. Koubbi, J. Rock, and F.A.M. Volckaert. 2009. Species identification in the trematomid family using nuclear genetic markers. *Polar Biol.* 32:1731–1741.
- Venkateswarlu, M., S. Raje Urs, B. Surendra Nath, H.E. Shashidhar, M. Maheswaran, T.M. Veeraiah, and M.G. Sabitha. 2006. A first genetic linkage map of mulberry (*Morus* spp.) using RAPD, ISSR, and SSR markers and pseudotestcross mapping strategy. *Tree Genet. Genomes* 3:15–24.
- Wang, H.Z., Y.D. Wang, X.Y. Zhou, Q.C. Ying, and K.L. Zheng. 2004. Analysis of genetic diversity of 14 species of *Cymbidium* based on RAPDs and AFLPs. *Acta Biol. Expt. Sinica* 37:482–486 (in Chinese).
- Xu, Y., L.L. Teo, J. Zhou, P.P. Kumar, and H. Yu. 2006. Floral organ identity genes in the orchid *Dendrobium crumenatum*. *Plant J.* 46:54–68.
- Yue, G.H., J.H. Xia, P. Liu, F. Liu, F. Sun, and G. Lin. 2012. Tracing asian seabass individuals to single fish farms using microsatellites. *PLoS One* 7:e52721.



Supplemental Fig. 1. Polyacrylamide gel electrophoresis profile of SSR08 tested among 105 Chinese cymbidiums accessions. Fifty-three of 105 accessions were selected based on their ancestry for this study. Marker was labeled with M; *Cymbidium goeringii* accessions were shown as 1 to 25; *Cymbidium ensifolium* accessions were shown as 26–47; *Cymbidium faberi* accessions were shown as 48–66; *Cymbidium sinense* accessions were shown as 67–85; and *Cymbidium kanran* accessions were shown as 86–105.