Scientia Horticulturae 134 (2012) 157–162

Contents lists available at SciVerse ScienceDirect

Scientia Horticulturae

journal homepage: www.elsevier.com/locate/scihorti

Analysis of population structure revealed apparent genetic disturbance in Korea *Cymbidium* collection

Kyaw Thu Moe [a], Yong-Jin Park [a,b,*]

[a] Department of Plant Resources, College of Industrial Sciences, Kongju National University, Yesan 340-702, Republic of Korea
[b] Legume Bio-Resource Center of Green Manure (LBRCGM), Kongju National University, Yesan 340-702, Republic of Korea

ARTICLE INFO

Article history:
Received 22 March 2011
Received in revised form 30 June 2011
Accepted 14 October 2011

Keywords:
Cymbidium
Phylogenetic relationship
Population structure
SSR
HWE
LD

ABSTRACT

The Oriental genus *Cymbidium* contains some of the most important and popular species of ornamental orchids in Korea. Genetic characterization of *Cymbidium* is vital for its conservation and management, as well as for understanding the genetic relationships among accessions. Among 100 tested accessions, 226 alleles were detected with an average of 16.1 alleles per simple sequence repeat locus; the number of alleles ranged from 28 for the KNU-CC-32 locus to 7 for KNU-CC-25. The allele size ranged from 103 to 380 bp. Thirteen loci were deviated from Hardy–Weinberg equilibrium, and showed highly significant linkage disequilibrium ($P < 0.01$). These results indicate that influencing disturbance in the *Cymbidium* population, such as natural selection and/or human intervention (i.e., plant breeding), are taking place among the species in Korea. The values for heterozygosity ranged from 0.000 to 0.969, with a mean value of 0.402. The average gene diversity and polymorphism information content values were 0.679 and 0.656, respectively, and ranged from 0.223 (KNU-CC-52) to 0.936 (KNU-CC-32) and from 0.219 (KNU-CC-52) to 0.933 (KNU-CC-32), respectively. All *Cymbidium* accessions were put into three main groups, and no evidence of mixed population ancestry was observed among the three populations identified. *Cymbidium sinensis* is not as well distributed and abundant as *C. goeringii* in Korea. *Cymbidium goeringii* is endemic in Korea. The genetic diversity in *Cymbidium* is not related to geographical area, which indicates that the species are randomly distributed in Korea. Our finding helps explain the genetic relationships and the population structure of *Cymbidium* species in Korea.

1. Introduction

The genus *Cymbidium* is a member of the Orchidaceae, one of the largest families of flowering plants (Dressler, 1993; Zhu et al., 2004). It comprises 44 species which are distributed throughout northwestern India, China, Japan, Korea, the Malay Archipelago, and northern and eastern Australia (Du Puy and Cribb, 1988; Obara-Okeyo and Kako, 1998). One category of *Cymbidium* has large, luxuriant, and colorful flowers with no fragrance and originates from tropical countries. The other category is the Oriental *Cymbidium*, or Chinese *Cymbidium*, which is found throughout East Asia (China, Korea, and Japan) and some regions of Southeast Asia. The Oriental *Cymbidium* orchids comprise popular ornamental plants and are one of the most important orchids in Korea. *Cymbidium* species show tremendous variations in floral morphology, pollinator relationships, and diversity in ecological habitats (Arditti, 1992; Judd et al., 1999). Furthermore, as they are among the most fragrant of plants, high ornamental and economic value (Chen and Tang, 1982; Wolff, 1999), these orchids have been cultivated for several centuries. These have magnificent inflorescences and leaves, with a wide array of colors and a fascinating variety of shapes and sizes. Despite their considerable features and economic importance, many species have become extraordinarily rare and endangered (Chung and Chung, 1999; Xia et al., 2008). Additionally, the genetic relationships among many of the major lineages of *Cymbidium* species still remain unclear (Choi et al., 2006). Therefore, for an effective conservation and sustainable use, one must properly characterize and evaluate the phylogenetic relationships of *Cymbidium*.

Morphological characteristics have been widely used in the characterization of *Cymbidium* orchids, but their morphological characteristics vary under different agroclimatic conditions. In recent years, biochemical markers such as isozymes (Chaparro et al., 1987; Chung and Chung, 1999; De Loose, 1979; Obara-Okeyo et al., 1998; Obara-Okeyo and Kako, 1997), and molecular markers such as random amplified polymorphic DNA (RAPD), amplified fragment lengths polymorphisms (AFLP), intersimple sequence repeats (ISSR), and the internal transcribed spacer (ITS) sequence

* Corresponding author at: Department of Plant Resources, College of Industrial Sciences, Kongju National University, Yesan 340-702, Republic of Korea.
Tel.: +82 41 330 1201; fax: +82 41 330 1209.
E-mail address: yjpark@kongju.ac.kr (Y.-J. Park).

0304-4238/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.scienta.2011.10.012

(Choi et al., 2006; Obara-Okeyo et al., 1998; Wang et al., 2004, 2009; Wu et al., 2008; Zhang et al., 2002, 2006) have been used to study genetic diversity and/or phylogenetic relationships in *Cymbidium*.

Simple sequence repeats (SSRs) are the markers of choice for crop improvement in many species because they are reliable and easy to score (Gupta and Varshney, 2000; Park et al., 2009). The SSR markers are codominant, multi-allelic, and require only a small amount of DNA for scoring. The characterization of close siblings by SSR markers is also straightforward because this analytical method simplifies breed confirmation (Garris et al., 2005; Panaud et al., 1996). Moreover, SSRs present a higher level of polymorphism and greater information content than AFLPs and RAPDs (Belaj et al., 2003). Owing to these considerations, and because they are Mendelian inherited and amenable to semi-automation through the use of fluorescence-labeled primers with high reproducibility and reliability, the use of SSRs has been increasing (Brondani et al., 1998; Jatoi et al., 2006). Agrama et al. (2007) pointed out that SSRs are useful molecular markers for marker-assisted selection, analysis of genetic diversity, and population genetic analysis in various species (Agrama et al., 2007). In this study, we used SSR markers to assess the level of genetic diversity and fine population structure and to explore the phylogenetic relationships among collected *Cymbidium* accessions in Korea.

2. Materials and methods

2.1. Plant materials and DNA extraction

To analyze the genetic diversity and population structure of *Cymbidium* species in Korea, 100 *Cymbidium* accessions, including 76 accessions collected from different geographical regions and 24 accessions from the National Institute of Horticultural and Herbal Science of the Rural Development Administration (RDA), Republic of Korea, were used (Supplementary table S1). 76 accessions were collected according to its density at designed geographical regions. Stratified random design was used to select suitable location. First, we stratified the location according the density of flower production. Then the accessions were collected from the randomly selected geographical regions. The accessions collected from different geographical regions were kept in Kongju National University, Republic of Korea.

DNA was extracted from fresh leaves using a DNA extraction kit (Qiagen, Hilden, Germany). The relative purity and concentration of extracted DNA was estimated with NanoDrop ND-1000 (NanoDrop Technologies, Inc., Wilmington, DE, USA). The final concentration of each DNA sample was adjusted to 20 ng/μl.

2.2. SSR genotyping

Fourteen SSR markers developed by our group (Moe et al., 2010) were used, and the M13-tailed PCR method of Schuelke (2000) was used to measure the size of the PCR products (Schuelke, 2000). Conditions of the PCR amplification were as follows: 94 °C for 3 min, 30 cycles each at 94 °C for 30 s, 55 °C (varied with different annealing temperature requirements of primers) for 45 s, 72 °C for 1 min, followed by 10 cycles at 94 °C for 30 s, 53 °C for 45 s, 72 °C for 1 min, and a final extension at 72 °C for 10 min. The SSR alleles were resolved on an ABI-3500 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) using GeneMapper 4.1 software (Applied Biosystems) and sized precisely using the GeneScan Installation Kit DS-33 (Applied Biosystems) and the GeneScan 600 LIZ Size Standard v2.0 (Applied Biosystems). The GeneScan Installation Standard DS-33 consists of pooled PCR products labeled with 6-FAM, VIC, NED and PET dyes.

2.3. Data analysis

The variability at each locus was measured in terms of the number of alleles, heterozygosity (H), major allele frequency (M_{AF}), gene diversity (GD), and polymorphic information content (PIC) using PowerMarker 3.25 (Liu and Muse, 2005). The UPGMA algorithm of MEGA4 software embedded in PowerMarker was used to construct an unrooted neighbour-joining tree of each accession based on the shared allele distances (Tamura et al., 2007). The genetic distance (Nei, 1978) between each pair of accessions was determined using the genetic analysis package POPGENE Version 1.32 (Yeh et al., 1999). The same program was used to test the Hardy–Weinberg equilibrium (HWE) and pairwise linkage disequilibrium (LD).

Population structure was determined using the model-based software program, Structure 2.2 (Falush et al., 2003; Pritchard et al., 2000). In this model, several populations (K) are assumed to be present, each of which is characterized by a set of allele frequencies at each locus. Individuals in the sample are assigned to populations (clusters) or jointly to more populations if their genotypes indicate that they are admixed. All loci are assumed to be independent, and each K population is assumed to follow HWE. Posterior probabilities were estimated using a Markov chain Monte Carlo (MCMC) method. The MCMC chains were run for 100,000 burn-in period lengths, followed by 200,000 iterations using a model allowing for admixture and correlated allele frequencies. At least three runs of the Structure software were performed with K ranging from 2 to 10, and an average likelihood value, L (K), across all runs was calculated for each value of K. The model choice criterion to detect the most probable value of K was ΔK, which is an ad hoc quantity related to the second-order change in the log probability of data (Ln $P(D)$) with respect to the number of clusters inferred by Structure (Evanno et al., 2005). An individual was assigned to a group if more than 75% of its genome fraction value derived from that group.

3. Results

3.1. SSR polymorphism

In total, 226 alleles were detected with an average of 16.1 alleles per locus among 100 tested accessions. Locus KNU-CC-32 had the highest number (28) of alleles, whereas only seven alleles were observed at KNU-CC-25 (Table 1). The allele size ranged from 103 to 380 bp. The allele frequency data showed that rare alleles (with a frequency < 0.05) comprised 74.8% of all alleles, while intermediate (0.05 < frequency < 0.50) and abundant alleles (frequency > 0.50) comprised 23.5% and 1.8% of all detected alleles, respectively; thus, most of the alleles were of low frequency (Fig. 1; Table 1). The values for heterozygosity ranged from 0.000 to 0.969 with a mean value of 0.402. A Bonferroni correction for multiple comparisons was applied to the HWE at a significance level of $P < 0.05$, and 13 loci (except KNU-CC-52) deviated from HWE (Table 1). Highly significant ($P < 0.01$) LD values were observed among 13 pairs of loci (data not shown). The average gene diversity and PIC values were 0.679 and 0.656, respectively, and ranged from 0.223 (KNU-CC-52) to 0.936 (KNU-CC-32), and from 0.219 (KNU-CC-52) to 0.933 (KNU-CC-32), respectively.

When the data were reanalyzed according to the populations determined by structural analysis, Nei's genetic identity ranged from 0.340 (between Pop-1 and Pop-3) to 0.400 (between Pop-1 and Pop-2). Pop-2 revealed the highest heterozygosity (0.512), gene diversity (0.692), and polymorphic information content (0.648) while Pop-3 showed the lowest heterozygosity (0.343) and genetic diversity (0.541), and Pop-1 showed the lowest PIC (0.519) (Table 1). The genetic distance between populations was