Development of *Cymbidium ensifolium* genic-SSR markers and their utility in genetic diversity and population structure analysis in cymbidiums

Xiaobai Li1*, Feng Jin2, Liang Jin1, Aaron Jackson3, Cheng Huang4, Kehu Li5 and Xiaoli Shu6*

**Abstract**

**Background:** *Cymbidium* is a genus of 68 species in the orchid family, with extremely high ornamental value. Marker-assisted selection has proven to be an effective strategy in accelerating plant breeding for many plant species. Analysis of cymbidiums genetic background by molecular markers can be of great value in assisting parental selection and breeding strategy design, however, in plants such as cymbidiums limited genomic resources exist. In order to obtain efficient markers, we deep sequenced the *C. ensifolium* transcriptome to identify simple sequence repeats derived from gene regions (genic-SSR).

**Result:** The 7,936 genic-SSR markers were identified. A total of 80 genic-SSRs were selected, and primers were designed according to their flanking sequences. Of the 80 genic-SSR primer sets, 62 were amplified in *C. ensifolium* successfully, and 55 showed polymorphism when cross-tested among 9 *Cymbidium* species comprising 59 accessions. Unigenes containing the 62 genic-SSRs were searched against Non-redundant (Nr), Gene Ontology database (GO), eukaryotic orthologous groups (KOGs) and Kyoto Encyclopedia of Genes and Genomes (KEGG) database. The search resulted in 53 matching Nr sequences, of which 39 had GO terms, 18 were assigned to KOGs, and 15 were annotated with KEGG. Genetic diversity and population structure were analyzed based on 55 polymorphic genic-SSR data among 59 accessions. The genetic distance averaged 0.3911, ranging from 0.016 to 0.618. The polymorphic index content (PIC) of 55 polymorphic markers averaged 0.407, ranging from 0.033 to 0.863. A model-based clustering analysis revealed that five genetic groups existed in the collection. Accessions from the same species were typically grouped together; however, *C. goeringii* accessions did not always form a separate cluster, suggesting that *C. goeringii* accessions were polyphyletic.

**Conclusion:** The genic-SSR identified in this study constitute a set of markers that can be applied across multiple *Cymbidium* species and used for the evaluation of genetic relationships as well as qualitative and quantitative trait mapping studies. Genic-SSR's coupled with the functional annotations provided by the unigenes will aid in mapping candidate genes of specific function.

**Keywords:** *Cymbidium ensifolium*, Genic-SSR, Genetic diversity, Population structure

**Background**

*Cymbidium* is a genus of 68 species in the orchid family [1]. *Cymbidium* species are mainly distributed in the tropical and subtropical regions of Asia, including northwest India, China, Japan, Korea, the Malay Archipelago, and north and east Australia [2,3]. A total of 49 species can be found in China, including five famous species, i.e., *C. goeringii*, *C. faberi*, *C. ensifolium*, *C. kanran*, and *C. sinense*. These cymbidiums comprise some of the rarest plant species, with only a few surviving original populations and some reintroduced plants in the south of China, including Yunnan and Taiwan [4]. The fascinating varieties and shapes of their flowers endow these species with extremely high ornamental value that has attracted the world’s attention.
Knowledge of the genetic diversity and population structure of germplasm collections is an important foundation for plant improvement [5]. Estimation of genetic distance among germplasm is helpful in selecting parental combinations for creating segregating populations so as to maintain genetic diversity in a breeding program. However, genetic diversity may appear spatially structured at different scales, such as population, subpopulation or among neighboring individuals [6]. Population genetic analyses can provide important parameters including levels of genetic variation and the partitioning of this variability within/between populations [7]. The genetic diversity or population structure of C. ensifolium and other cymbidiums have been measured by using different molecular tools, including restriction enzyme polymorphism (RFLP) markers [3], random amplified polymorphic DNA (RAPD) markers [3,4,8], amplified fragment length polymorphism (AFLP) markers [4], polymorphisms of internal transcribed spacers (ITS) of nuclear ribosomal DNA and plastid, inter-simple sequence repeats (ISSR) markers [4,9], and SSRs [10,11]. Compared with RAPD, ISSR and ITS, SSR markers are more reliable, locus-specific, codominant, highly polymorphic, and well distributed throughout the genome [12]. Moreover, SSRs only require polymerase chain reaction (PCR), which is a big advantage over RFLP and AFLP. These features make SSRs well suited for marker-assisted selection, genetic diversity analysis, population genetic analysis, genetic mapping, and genetic map comparison in various species [13,14].

The number of SSR is very limited for C. ensifolium, due to limited sequence resources. Until now, the National Center for Biotechnology Information (NCBI) contained very limited Cymbidium sequence information, i.e., 692 nucleotide sequences and 78 expressed sequence tags (ESTs) (http://www.ncbi.nlm.nih.gov/nucest?term=cymbidium%5BOrganism%5D, verified 2014). RNA-seq provides a fast, cost-effective, and reliable approach for generating large-scale transcriptome data in non-model species, and also offers an opportunity to identify and develop genic-SSRs by transcriptome data mining [15]. Compared with traditional anonymous SSRs from genomic DNA, these new genic-SSR markers have two advantages, i.e. a wealth of functional annotations and high transferability across taxa [15,16]. Herein, we extracted the total mRNA from C. ensifolium flower buds for RNA-seq, which resulted in 9.52 Gb of transcriptome data. From the Censifolium transcriptome, we obtained 55 new polymorphic microsatellite loci after testing their transferability across 59 Cymbidium accessions.

Methods

Plant materials

A total of 11 C. ensifolium accessions were employed to test genic-SSRs and additional 47 accessions from C. lancifolium, C. floribundum, C. suavissimum, C. cyperifolium, C. qiubeiense, C. faberi, C. goeringii and C. sinense were used to cross-test these markers among multiple species. The plants were grown and maintained in a greenhouse at the Zhejiang University under natural light (Table 1). Fresh leaf samples were collected from two or three seedling of each accession for genomic DNA extraction.

Genic-SSR search and primer design

Total RNA was isolated from native cultivar of C. ensifolium Tiegusu using TRIzol? reagent (Invitrogen, CA, USA) and treated with RNase-free DNase I (TaKaRa Bio, Dalian, China) for 45 min according to the manufacturer's protocol. The RNA was used in cDNA library construction and Illumina deep sequencing [17]. The raw sequencing reads were stringently filtered, and high-quality reads were assembled de novo using Trinity with an optimized k-mer length of 25 [18], MSATCOMMANDER V. 0.8.2 [19] was used to analyze SSR distribution. The minimum number of repeats for SSR detection was as follows: six for di-SSRs, and four for tri-, tetra-, penta-, and hexa-SSRs. The open reading frame (ORF) and untranslated region (UTR) within unigenes were identified using Trinity [18]. Software Primer3.0 [20] was used to design primers for genic-SSR loci with sufficient flanking sequences.

Unigenes containing genic-SSRs were compared with protein databases, including the non-redundant (Nr) database (http://www.ncbi.nlm.nih.gov/), using BLASTX with a significance cut-off E-value of 1e^-5 [17]. For the non-redundant annotations, BLAST2GO V. 2.4.4 was used to obtain Gene Ontology (GO) annotations of unique transcripts [21]. Metabolic pathway analysis were performed based on the pathways of Oryza sativa in the Kyoto Encyclopedia of Genes and Genomes (KEGG) [22,23]. The unigene sequences were also aligned to the KOG (Eukaryotic Orthologous Groups) database to predict and classify possible functions [24].

Genotyping

Genomic DNA was extracted from leaf samples as previously described [25]. PCR primers were synthesized by Life Technologies (AB & Invitrogen, Shanghai, China). PCR reactions were conducted based on a previously published protocol [26]. The PCR products were separated through polyacrylamide gel electrophoresis using 8% bis-acrylamide, 0.5% TBE buffer, 0.07% APS, and 0.035% TEMED. The gel was run at constant 120 V for approximately 3 h in 1% TBE buffer. The gel was silver-stained according to Lis procedure [27], and was then documented using a scanner. The genotype was determined by analysis of the bands pattern, dependent on the number and the position of bands.
Statistical analysis
Genetic distance was calculated using Nei's distance [28]. Phylogenetic reconstruction was based on the unweighted pair-group method that utilizes the arithmetic average (UPGMA) method implemented in PowerMarker version 2.7 [29]. The tree that was used to visualize the phylogenetic distribution of accessions and ancestry groups was constructed using MEGA version 4 [30]. A model-based program structure [31] was used to infer population structure with 5,000 burn-in and run length. The model allowed for admixture and correlated allele frequencies. The number of groups (K) was set from 1 to 10, each with 10 independent runs. The most probable structure number (K) was determined through log probability [32]. Principal component analysis (PCA), which summarizes the major patterns of variation in a multi-locus data set, was performed using NTSYSpc version 2.11 V [33]. Two principal components were used to represent the dispersion of the collection accessions graphically [34]. PowerMarker was used to calculate the average number of marker alleles and the polymorphism information content (PIC) values. Fixation index ($F_{ST}$), which indicates the differentiation among genetic groups, was calculated using an Analysis of Molecular Variance (AMOVA) approach in Arlequin V2.000 [35].

Results
Genic-SSR search and primer design
In C. ensifolium transcriptome, 98,819,349 reads, (9.52 Gb), were obtained after removal of adaptor sequences, ambiguous reads, and low-quality reads (Q-value <25). These reads

Table 1 Fifty nine cymbidium accessions used for genetic analysis

| Accession | Name          | Group | Species          |
|-----------|---------------|-------|------------------|
| 1         | Tiegusu       | 4     | C. ensifolium    |
| 2         | Qingshanyuquan| 4     | C. ensifolium    |
| 3         | Jinsimawei    | 4     | C. ensifolium    |
| 4         | Jinhe         | 4     | C. ensifolium    |
| 5         | Yinsimawei    | 4     | C. ensifolium    |
| 6         | Dayibai       | 4     | C. ensifolium    |
| 7         | Dahongzhusha  | 2     | C. ensifolium    |
| 8         | Qiuhong       | 4     | C. ensifolium    |
| 9         | Baodao        | 4     | C. ensifolium    |
| 10        | Jinhe         | 2     | C. ensifolium    |
| 11        | Tianhe        | 4     | C. ensifolium    |
| 12        | Shiantsaihao  | 4     | C. ensifolium    |
| 13        | TuerA         | 2     | C. lancifolium   |
| 14        | TuerB         | 2     | C. lancifolium   |
| 15        | DuohualanA    | 5     | C. floribundum   |
| 16        | GuoxianglanA  | 2     | C. sauvissimum   |
| 17        | ShyelanA      | 1     | C. openifolium   |
| 18        | ShyelanB      | 1     | C. openifolium   |
| 19        | ShyelanC      | 1     | C. openifolium   |
| 20        | ShyelanD      | 1     | C. openifolium   |
| 21        | QiubeidonghuiA| 2     | C. qiubeiense    |
| 22        | ShyelanE      | 1     | C. openifolium   |
| 23        | LvlanA        | 1     | C. faberi        |
| 24        | GuoxianglanB  | 5     | C. sauvissimum   |
| 25        | LvlanB        | 1     | C. faberi        |
| 26        | DuohualanB    | 5     | C. floribundum   |
| 27        | Yuhudie       | 2     | C. goeringii     |
| 28        | Yinhe         | 5     | C. goeringii     |
| 29        | Silan         | 2     | C. goeringii     |
| 30        | Hexingmei     | 5     | C. goeringii     |
| 31        | Dasongmei     | 2     | C. goeringii     |
| 32        | Yipin         | 2     | C. goeringii     |
| 33        | Huangmei      | 2     | C. goeringii     |
| 34        | Puchunhong    | 2     | C. goeringii     |
| 35        | Chunjiansuxin | 2     | C. goeringii     |
| 36        | Hongmeigui    | 2     | C. goeringii     |
| 37        | Wenyi         | 2     | C. goeringii     |
| 38        | Jixianrudan   | 2     | C. goeringii     |
| 39        | Dayipin       | 3     | C. faberi        |
| 40        | Ruyisu        | 2     | C. faberi        |
| 41        | Jiepeimei     | 3     | C. faberi        |
| 42        | Xinhanghaimei | 3     | C. faberi        |
| 43        | Laoranzi      | 3     | C. faberi        |
| 44        | Xinshanglei   | 3     | C. faberi        |
| 45        | Guienmei      | 3     | C. faberi        |

Table 1 Fifty nine cymbidium accessions used for genetic analysis (Continued)

46| Xiashanjiujielan | 3 | C. faberi |
47| Guifei          | 3 | C. faberi |
48| Mingyue         | 3 | C. faberi |
49| Xiyang (Qingxiang) | 3 | C. faberi |
50| Yuchan          | 3 | C. faberi |
51| QiubeidonghuiB  | 2 | C. qiubeiense |
52| DuohualanC      | 5 | C. floribundum |
53| DuohualanD      | 5 | C. floribundum |
54| QiubeidonghuiC  | 2 | C. qiubeiense |
57| Wuzicui         | 2 | C. sinense |
58| Jinhuaishan     | 2 | C. sinense |
59| Rixiang         | 2 | C. sinense |
60| Qihei           | 2 | C. sinense |
61| Darno           | 2 | C. sinense |
62| Hongmeiren      | 2 | C. sinense |
63| Baimo           | 2 | C. sinense |

aFive groups indicated by population structure analysis.
motif, which are often associated with polymorphism [37]. The length of the motif and the number of the repeat are the two factors that influence the likelihood of being polymorphic [36]. Another two factors considered were the locations of SSRs relative to ORFs. SSRs within UTR are exposed to lower selective pressure than those in coding regions and have a higher likelihood of being polymorphic [37].

In the present study, 7,936 genic-SSRs were identified, with one SSR locus for every 17.56 kb (kb/SSR). Estimated locations (coding, 5′UTR or 3′UTR) were obtained for 5,524 genic-SSRs. Sequence information could not be determined for the remaining 2,412 genic-SSR regions, because the locations were extended over both estimated coding and non-coding regions. Given such high numbers of SSR, we analyzed the sequence data to isolate high quality SSR loci for further testing. An important factor considered was the locations of SSRs relative to ORFs. SSRs within UTR are exposed to lower selective pressure than those in coding regions and have a higher likelihood of being polymorphic [36]. Another two factors are the length of the motif and the number of the repeat motif, which are often associated with polymorphism [37]. Thus, SSRs within UTR, with short motifs and high repeat number would be the best marker candidates. Herein, we selected 80 genic-SSRs and designed primers based on their motifs, sizes and locations.

Genic-SSRs profile
All primer sets were initially tested among 12 C. ensifolium accessions, and then were cross-tested among other 47 Cymbidium accessions (Table 1). Of the 80 genic-SSR primers, 62 amplified within C. ensifolium accessions successfully, and 55 showed polymorphism when cross-tested among all 9 cymbidium species (Additional file 1: Figure S1). These accessions belonged to 9 cymbidium species i.e. C. ensifolium, C. lancifolium, C. suavissimum, C. cyperfolium, C. qiubeiensis, C. floribundum, C. goeringii, C. faberi and C. sinense. Among the 55 polymorphic markers, the PIC averaged 0.407, ranging from 0.033 (for both SSR29 and SSR31) to 0.863 (for SSR73). Similarly, allele number averaged 5.75, ranging from 2 (for SSR06, SSR24, SSR29, SSR31, SSR46, SSR55, SSR71, SSR75 and SSR79) to 16 (for SSR73) (Table 2). These results suggested that genic-SSR markers had a broad applicability within Cymbidium genus.

Genetic diversity and population structure
These genic-SSRs revealed genetic variation among accessions. The genetic distance among accessions ranged from 0.016 to 0.618, with an average of 0.391. The model-based clustering method revealed five groups (Figure 1A and B). Group 2 had the most accessions (26), with the highest mean genetic distance (MGD) of 0.431 among these accessions; Group 4 had 10, with an average distance of 0.236; Group 5 had 7, with MGD of 0.332; Group 1 and Group 5 both had 7 accessions, with MGD of 0.155 and 0.332, respectively; Group 3 had 9, with MGD of 0.213. Genetic distance among five groups was from 0.340 (between group 1 and group 5) to 0.176 (between group 2 and group 4, with average of 0.248) (Table 3).

The five groups revealed by the model-based clustering analysis consisted of different species. Three groups comprised more than one species, whereas the other two only comprised one species. Group 1 included two species i.e. C. cyperfolium and C. goeringii; Group 2 included C. ensifolium, C. lancifolium, C. suavissimum, C. qiubeiensis, C. goeringii, C. faberi, and C. sinense; Group 5 included C. floribundum, C. suavissimum and C. goeringii. Group 3 and Group 4 included only C. faberi and C. ensifolium, respectively (Figure 2).

The first two components in PCA (47.87% and 21.59% of total variation, respectively) discriminated the five groups at a certain level. Basically, accessions in group 1 and group 3 stayed alone, whereas group 2 overlapped with group 4 and group 5 (Figure 1C). In the phylogenetic tree, group 2 and group 4 were genetically close, while group 5 was relatively distant from the other groups (Figure 1A). In addition, a few accessions in group 2 had admixture ancestry from group 3 and group 4, while accessions in group 3 and group 1 had less admixture ancestry (Figure 1B). AMOVA results showed that 25.34% of the total variation was among groups, while 74.66% of the variation was within groups. The $F_{ST}$ was 0.25, as indicated by the AMOVA approach.

Genic-SSR annotation
Annotations of these unigenes provide biological information for 62 genic-SSRs, such as KOG clusters, GO, and KEGG pathway information. Distinct gene sequences were first searched using BLASTX against the Nr database. The results showed that 53 unigenes had hits that exceeded the E-value threshold. In the present study, 39 unigenes were categorized into 25 GO terms in three GO ontologies (Figure 3A). Two groups membrane and nucleus, one group binding, and one group cellular process comprised the most representative genes found in cellular components, molecular function, and biological processes, respectively. Out of 53 hits in the Nr databases, 18 sequences were classified into 9 KOG categories (Figure 3B). Among the 9 KOG categories, General function prediction only and Post-translational modification, protein turnover, chaperones were the two largest groups. When referenced to rice (Oryza sativa), 15 unigenes were found to be involved in 14 pathways (Figure 3C). The most highly representative one was metabolic pathways, where unigenes shared similarity with 18 rice sequences.
| Name   | Product size (bp) | SSR   | SSR location | Primer                              | Homologs in non-redundant database (accession in Genbank)                                                                 | GO annotation | KOG annotation | KEGG annotation | Allele number | PIC   |
|--------|------------------|-------|---------------|-------------------------------------|--------------------------------------------------------------------------------------------------------------------------|---------------|----------------|-----------------|---------------|-------|
| SSR01  | 400-500          | (AC)8 | utr5          | F: AACGCCCATCTCAATACC<br>R: GGAAGGCTATTGCAACG<br>PREDICTED: probable transcription factor KAN2-like (XP_002278005.2) | GO: 0003677 KOG1601 NULL | 5       | 0.552          |
| SSR02  | 300-400          | (AC)8 | utr5          | F: CTCCTTCAAGCTTCTGCC<br>R: GACCGCCAGCTTAAGACC<br>PREDICTED: histone-lysine N-methyltransferase, H3 lysine-9, H3 lysine-27, H4 lysine-20 and cytosine specific SUVH2 (XP_002282386.1) | GO: 0042393 NULL NULL NA NA |           |
| SSR03  | 400-500          | (AC)8 | utr3          | F: CTCGGGTCAATTGCAAGCC<br>R: GGGTGGGTATGGCGAAATC<br>PREDICTED: mitochondrial import receptor subunit TOM20 (XP_002269795.1) | GO: 0045040 NULL NULL 7 0.690 |
| SSR04  | 400-500          | (AC)8 | utr3          | F: AGAATCTGCCAACACCTTGATAC<br>R: GCAGATGCCAGTTAGAATGG<br>PREDICTED: protein CbbY, chromosomal-like isoform 1 (XP_003574671.1) | NULL NULL NULL NULL 6 0.657 |
| SSR05  | 1000             | (AC)8 | utr3          | F: AGAATCTGCCAGCTTGAAAC<br>R: GCAGATGCCAGTTAGAATGG<br>PREDICTED: histone-lysine N-methyltransferase, H3 lysine-9, H3 lysine-27, H4 lysine-20 and cytosine specific SUVH2 (XP_002282386.1) | GO: 0016787 NULL NULL 3 0.125 |
| SSR06  | 600              | (AC)9 | utr3          | F: GCCTTGAAGTCGCGATAACC<br>R: AAGACGCGCATTCATACC<br>PREDICTED: predicted protein (XP_002298559.1) | GO: 0016020 KOG0157 K09587 2 0.063 |
| SSR07  | 300-400          | (AC)9 | utr3          | F: AGCGCTGATCCATCTAC<br>R: CAGTCTGTTAGGAAGCCG<br>PREDICTED: uncharacterized protein LOC100243361 (XP_002276849.2) | GO: 0008987 NULL K03517 4 0.180 |
| SSR08  | 100-200          | (AC)10| utr3          | F: TOCTTGAAACTATCGGCAGAC<br>R: GTTGGCCGAGGCAGTCG<br>Predicted protein (XP_002298559.1) | GO: 0016020 KOG0157 K09587 2 0.063 |
| SSR11  | 600              | (AG)10| utr3          | F: AACTGACAGCTTGAGCAG<br>R: CGTGCGATTGCCCTATCC<br>Uncharacterized protein LOC100273319 precursor (NP_001141232.1) | GO: 0005774 NULL NULL 6 0.477 |
| SSR12  | 300              | (AG)11| utr5          | F: TCAGCGCGGAGGTGATATAC<br>R: CTTGCCGATCGGAGCTC<br>PREDICTED: phosphatidylinositol-4-phosphate 5-kinase 9-like (XP_002265706.1) | GO: 0016020 KOG0229 K00889 NA NA |
| SSR13  | 400-500          | (AG)11| utr5          | F: GCCTGTCGAGTGGGAAAC<br>R: GCCGTGCTGTTAGTGTCG<br>Predicted protein (XP_0022317724.1) | GO: 0005488 NULL NULL 6 0.343 |
| SSR14  | 300              | (AG)11| utr5          | F: CACACGCACTCACATCTCAC<br>R: TACAGCCCTGGTACCC<br>Uncharacterized protein product (CB020568.3) | GO: 0006099 KOG1257 K00029 8 0.467 |
| SSR15  | 100-200          | (AG)11| utr3          | F: CCTTTCAGCCGTTAAGCAC<br>R: CTTCCGTGTTGGCGTTATAGGG<br>PREDICTED: uncharacterized protein LOC100825549 (XP_003558805.1) | GO: 0005783 NULL NULL 4 0.339 |
| SSR16  | 300-400          | (AG)11| utr5          | F: GCCCCACGAACTTATCCATCT<br>R: GCAGTCAAGAAGAAGCTGG<br>PE repeat family protein (XP_003014087.1) | NULL NULL NULL NULL 7 0.348 |
| SSR17  | 400              | (AG)11| utr5          | F: GGATCCACCAACACGAC<br>R: TCCACAAAGAAGAAGG<br>Transcription factor (ADGS7844.1) | GO: 0003677 NULL K09060 4 0.417 |
| SSR   | Range | Length | Forward Primer | Reverse Primer | Function                                                                 | ORF ID | Length | Function | Length | Identity |
|-------|-------|--------|----------------|----------------|--------------------------------------------------------------------------|--------|--------|----------|--------|----------|
| SSR18 | 300-500 | 300   | F: TGAAACGTGGCCCTAGTTC | R: AGCAAGCCTCAGCTCGAAAC | Conserved hypothetical protein (XP_002527260.1)                           | NULL   | 13     | 0.519    |        |          |
| SSR21 | 300-500 | 200   | F: TGCCCGCATGTCGTTGCTG | R: AGCTATGCTGGGCTTGGTG | Hypothetical protein OsU_08996 (EAZ25197.1)                               | NULL   | 15     | 0.794    |        |          |
| SSR22 | 200-300 | 14    | F: TATCGGTCTCCTCCAACCG | R: ACATGCTTCTCAATGTCGG | 14-3-3-like protein B-like (ACQ45020.1)                                  | GO: 0019904 | NULL   | NULL     |        |          |
| SSR23 | 100-200 | 300   | F: CGCCGATGCTGTTGACGG | R: CGATGCTTCGTAGCTGGG | PREDICTED: beta-amyrase 1, chloroplastic isoform 1 (XP_002285569.1)     | GO: 0005634 | NULL   | NULL     |        |          |
| SSR24 | 200-300 | 200   | F: TCCGTAACCTGTTGCAAGG | R: ACCTGGAAGCTACCCAGAC | PREDICTED: flavin-containing monoxygenase YUCCA6-like (XP_003550114.1) | GO: 0050661 | NULL   | NULL     |        |          |
| SSR25 | 100-250 | 250   | F: GAATCTCTCGCACCCGAAG | R: TGGACAACATGAAGTACGC | Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B1, putative  | NULL   | 11     | 0.638    |        |          |
| SSR26 | 200-300 | 100   | F: GCTTTATGCGCATGCTGC | R: CGTCGCTCCTACGCAATAC | Unnamed protein product (CB025980.3)                                     | GO: 0005634 | NULL   | NULL     |        |          |
| SSR27 | 100-250 | 600   | F: CTGCCTCACAGCTAAATG | R: GCATGCTGACGTCACGAC | Os04g0512400 (NP_001053298.1)                                            | GO: 0046872 | NULL   | NULL     |        |          |
| SSR28 | 200-300 | 200   | F: AGCAACCGCAAGTGATCG | R: ATTCGACTACGCCGCAATGAC | RING finger protein 113A, putative (XP_002522169.1)                      | NULL   | 11     | 0.486    |        |          |
| SSR31 | 100-200 | 500   | F: AAAAGGCGGTTGGTACGC | R: TTTGACTACGGAAATGGGC | Protein MSF1, putative (XP_002535293.1)                                  | NULL   | 11     | 0.601    |        |          |
| SSR32 | 1000-1100 | 900 | F: CTCCGGGCTTCTGCTTCTC | R: AGTGGATGAGCTTGAGCGC | PREDICTED: HVA22-like protein j (XP_002281038.1)                        | NULL   | 7      | 0.695    |        |          |
| SSR34 | 700-900 | 400   | F: GAGAGGAAATTGCAATGAC | R: ACGGAGCTAGACCTGACTC | Hypothetical protein (BA168347.1)                                       | NULL   | 6      | 0.695    |        |          |
| SSR35 | 500-700 | 900   | F: AGAGTTGATTGTCAGCTCCTC | R: TGCTCTTCTGTTGATGTTGCC | PREDICTED: diacylglycerol kinase-like (XP_003534537.1)                  | GO: 0009395 | NULL   | NULL     |        |          |
| SSR36 | 400-500 | 200   | F: AGATTGACCGACCAGCCG | R: AGAGGATCATGGTGAGGCGC | NULL                                                                     | NULL   | 5      | 0.536    |        |          |
| SSR37 | 200-300 | 200   | F: GGCGCTAGCGACCCCGTC | R: ATTTGCTGACCCACAAAGCGG | NULL                                                                     | NULL   | 3      | 0.205    |        |          |
Table 2 List of the 62 C. ensifolium genic-SSR primers including their unigenes annotation (Continued)

| SSR  | Reaction Range | Unigene Type | Unigene Name | Protein Name | Gene Ontology | GO ID | Gene Ontology | GO ID |
|------|----------------|--------------|--------------|--------------|---------------|-------|---------------|-------|
| SSR38| 200-300        | UTR3         | F: TAGCCCATGCCCAGGTGTCG  | LOC10028537 (NP_001151738.1) | 10028537     |   | 10028537      | 10028537 |
| SSR39| 1000-1100      | UTR3         | F: ACAAGCAGTCCAGGTGTCG  | unnamed protein product (CBI38283.3) | 38283.3     |   | 38283.3       |   |
| SSR40| 400            | UTR5         | F: ACAAGCAGTCACTCCAAAATTC  | predicted: probably inactive leucine-rich repeat receptor-like protein kinase At2g25790-like (XP_002267653.1) | 25790       |   | 25790         |   |
| SSR42| 200-300        | UTR5         | F: GAGACACATACCTCGTCG  | unnamed protein product (CBI18667.3) | 18667.3     |   | 18667.3       |   |
| SSR43| 500            | UTR5         | F: GGAGCTGCACTCGGAAATG  | glycinebetaine/proline transporter (BAJ07206.1) | 07206.1     |   | 07206.1       |   |
| SSR44| 300-400        | UTR5         | F: CGTCGACTCCTCCGAGATC  | predicted protein (BAJ93650.1) | 93650.1     |   | 93650.1       |   |
| SSR45| 400-500        | UTR5         | F: GCTTACACATCCCAAATTCC  | predicted protein (XP_002267653.1) | 2267653.1   |   | 2267653.1     |   |
| SSR46| 200-300        | UTR5         | F: CCTTCGTGGACTCAACAGC  | hypothetical protein SORBIDRAFT_01g031510 (XP_002465065.1) | 031510.1    |   | 031510.1      |   |
| SSR47| 400-500        | UTR3         | F: GCAGGTGTCCTCATGGGAG  | CONSTANS-like protein (ADN97077.1) | 97077.1     |   | 97077.1       |   |
| SSR49| 300            | UTR3         | F: AGAGGGCGCCACTGTTTCC  | predicted protein (XP_002312577.1) | 2312577.1   |   | 2312577.1     |   |
| SSR52| 400-500        | UTR5         | F: AAGAGGCGACTGCAAAGCC  | hypothetical protein SORBIDRAFT_01g031070 (XP_002465040.1) | 031070.1    |   | 031070.1      |   |
| SSR53| 100-200        | UTR5         | F: GCTGAAAGGTGGCCCTGTC  | PREDICTED: uncharacterized protein LOC100830480 (XP_003580351.1) | 100830480   |   | 100830480     |   |
| SSR54| 200-300        | UTR5         | F: ATCTGCGCTCCACATGGGC  | hypothetical protein MTR_1g083540 (XP_003591171.1) | 083540.1    |   | 083540.1      |   |
| SSR55| 200-300        | UTR5         | F: TGGAGCTTTGGCTAGGTC  | hypothetical protein (CCA65980.1) | 65980.1     |   | 65980.1       |   |
| SSR56| 900-1000       | UTR5         | F: TGGTCTTATTTGGGAGAGGC  | predicted protein (XP_002324427.1) | 2324427.1   |   | 2324427.1     |   |
| SSR59| 200-300        | UTR3         | F: GTTCCACACCTCGTACGCTG  | leucine-rich repeat transmembrane protein kinase family protein (NP_177007.1) | 177007.1    |   | 177007.1      |   |

Li et al. BMC Genetics 2014, 15:124 Page 7 of 14
http://www.biomedcentral.com/1471-2156/15/124
| SSR  | 200-300 | (CGG)5 | utr5 | F: TACGGTTTCGACCACGCTC R: CCATCGACAGTCGCCGCAAAG | Unnamed protein product (CB41056.3) GO: 0005634 KOG0265 K10143 4 0.203 |
| SSR62 | 300-400 | (CGG)6 | utr5 | F: GGTGGGTAGACCCAGCTC R: TCTCTCAAGAACAACTCC | Hypothetical protein Os1_29809 (EA207555.1) GO: 0005634 NULL NULL 6 0.570 |
| SSR63 | 100-200 | (CGG)6 | utr5 | F: CTTCCCTCACCTGGATCGC R: CTGGCGATCAATCCGAGAC | Uncharacterized protein LOC100277749(NP_001414941.1) GO: 0008270 NULL NULL 4 0.308 |
| SSR64 | 400-500 | (CGG)6 | utr3 | F: CGCTCAAGAGATGGGACG R: TACGGTTTCGACCACGCTC | Os01g0226200 (NP_001042462.1) NULL NULL NULL 11 0.627 |
| SSR66 | 300-400 | (CGG)7 | utr3 | F: CATCTCTCTTGCCTCGATGC R: CCCCACAAAATTCGAGACC | PREDICTED: pentatricopeptide repeat-containing protein At5g42310, mitochondrial (XP_002272226.1) NULL KOG4197 NULL 4 0.126 |
| SSR68 | 100-200 | (GAT)5 | utr5 | F: CCAGATCGAATGGCTACG R: CAAGGAGCTGTCGAGGG | Hypothetical protein VITISV_010525 (CAN79523.1) GO: 0003723 NULL NULL 4 0.211 |
| SSR69 | 200-300 | (GAT)5 | utr5 | F: GTTTAGGCTAGCAGTGCG | NULL NULL NULL NULL 3 0.149 |
| SSR70 | 200-300 | (GAT)7 | utr3 | F: CCAAACAGCAGACAGATGC R: CGTGGCACAAAATGGAAGG | NULL NULL NULL NULL 5 0.529 |
| SSR71 | 400-500 | (GAT)5 | utr5 | F: GCATCGAAACACCTGTCGG R: CCCTACCGAGGTCTCAAC | Hypothetical protein SORIDRAFTH_09g018170 (XP_002439663.1) NULL NULL NULL 2 0.262 |
| SSR73 | 100-200 | (GAT)5 | utr3 | F: GGACACAATGGAGAGCGGG R: TGACGAAACATGGCAGG | T4.15 (CCH50976.1) GO: 0044238 NULL NULL 16 0.863 |
| SSR75 | 400-500 | (GAT)6 | utr3 | F: GCCTTTCAGACCTCCGTG R: GGCGGCCATAG TAGGAAGA | Mitogen-activated protein kinase 1 (AEQ28763.1) GO: 0043622 KOG0660 K04371 2 0.118 |
| SSR76 | 500-600 | (GAT)6 | utr5 | F: AGACAGAGAGTCCCAAAGGG R: CAGGGATGTAAATGCGGG | NULL NULL NULL NULL 7 0.519 |
| SSR77 | 300-400 | (GAT)7 | utr3 | F: TTGTGGGCTGAGAAAGCG R: TGATCACGATGCGAGGG | NULL NULL NULL NULL 5 0.470 |
| SSR79 | 200-300 | (GAT)7 | utr5 | F: AGATTGATGATGAGCGACCTC R: TCCCTGAAGGGGCGAAAAC | Hypothetical protein Os1_35425 (EE67831.1) NULL NULL K10728 2 0.207 |
| SSR80 | 400-500 | (GAT)7 | utr3 | F: GCACCCACCTTGGGAGGG R: CCCTACATTACAGGGAAAAG | NULL NULL NULL NULL 8 0.626 |

Note: A total of 62 genic-SSR markers successfully amplified were listed, however SS polymorphic markers were used in subsequent population analysis or cross species comparison. NULL: no annotation. NA: monomorphic marker.
Discussion

Diversity
Because genic-SSR markers are derived from transcribed regions of DNA, they are expected to be more conserved and have a higher rate of transferability than anonymous SSR markers [38]. Herein, 55 *C. ensifolium* polymorphic genic-SSR markers exhibited 100% transferability across the 59 accessions of the 9 *Cymbidium* species tested. It is common that genic-SSRs possess a high potential for inter-specific transferability [39,40]. Other markers such as RAPDs, ISSRs and non-genic SSRs have also been used with success among *C. ensifolium* and the *Cymbidium* species reflecting the genetic similarity among many members of the genus [8,11,15].

The conserved nature of the genic-SSRs may limit their polymorphism relative to randomly selected SSRs.

Table 3 Pairwise comparison of Nei's genetic distance among groups and mean of genetic distance within group based on 55 polymorphic genic-SSRs

| Group      | No. of accessions | Mean genetic distance (MGD) | Group 1 | Group 2 | Group 3 | Group 4 | Group 5 |
|------------|-------------------|----------------------------|---------|---------|---------|---------|---------|
| Group 1    | 7                 | 0.155                      | 0.000   |         |         |         |         |
| Group 2    | 26                | 0.430                      | 0.217   | 0.000   |         |         |         |
| Group 3    | 9                 | 0.213                      | 0.252   | 0.190   | 0.000   |         |         |
| Group 4    | 10                | 0.236                      | 0.212   | 0.176   | 0.234   | 0.000   |         |
| Group 5    | 7                 | 0.332                      | 0.340   | 0.263   | 0.298   | 0.293   | 0.0000  |
Figure 2 Phylogenetic tree of cymbidiums. A: Unweighted pair-group method tree of 59 accessions based on 55 polymorphic SSRs, and B: Morphology of cymbidium's flower. Group 1 indicated by red dot, Group 2 by green dot, Group 3 by blue dot, Group 4 by yellow dot, Group 5 by purple dot.

Note: accessions were listed in table 1.
In this study, PIC of genic-SSR markers averaged 0.407, lower than 0.782 [5] and 0.639 [11] of anonymous SSRs tested on Chinese cymbidiums in other studies. The pairwise genetic distance averaged 0.391 among 59 accessions, which is also lower than that from previous studies conducted on Chinese Cymbidiums using other molecular markers [3,8,41-44]. Even though genic-SSRs revealed less variability than SSRs, these markers still reveal sufficient levels of variation for population genetic analysis.

Population structure

One of the biggest advantages for genic-SSRs is that they allow one to make direct comparisons among taxa without running the risk that locus-specific differences might mask true species-level differences, such as overall levels of genetic diversity, the extent of population structure, and so on. However, the greatest concern with the utilization of genic-SSRs in genetic studies is that selection on these loci might influence the estimation of population genetic parameters. While a recent study by Woodhead et al. [45] revealed that estimates of population differentiation based on genic-SSRs are comparable to those based on both SSRs and AFLPs in ferns, and large-scale comparative analysis suggest that only a very small percentage of all genes has experienced positive selection [46,47], a small fraction of SSRs will be inevitably subject to selection. The view is consistent with the theory that most mutations are neutral, or nearly neutral, [48] or, at least, do not change the function of gene products appreciably [49].

In the population genetic analysis, almost all accessions from the same species clustered together. *C. suavissimum* and *C. floribundum* were clustered into one brand, and clearly distinguished from other cymbidiums. Two of them belong to Section *Floribundum*, and have a distant relationship with other cymbidiums. However, the genetic relationship between *C. goeringii* and *C. sinensis* was close, which was congruent with the previous reports [5,11]. The close relationship was also found between *C. ensifolium* and *C. cyperifolium*. In the intersection level, we discovered that two accessions of *C. faberi* were clustered with *C. cyperifolium*, and accessions of *C. lancifolium* and *C. ensifolium* were scattered among ones of *C. goeringii*. The splitting feature of these clusters might be linked to the non-homologous synapomorphy, even though accessions belonged to different species. The accessions of *C. goeringii* did not always form a separate cluster in the phylogenetic tree or were not grouped together in structure analysis, suggesting that they were polyphyletic. Previous morphologic, cytogenetic, and

![Figure 3 Functional annotations of unigenes containing SSR. A: KOG prediction and possible function, B: GO functional classifications, and C: KEGG pathways involved.](http://www.biomedcentral.com/1471-2156/15/124)
molecular studies have shown that the major lineages of Chinese cymbidiums are ambiguous. *C. ensifolium* and *C. sinense* are classified in section *Jensoa*; *C. faberi* and *C. goeringii*, are classified in section *Maxillarianthae*; *C. faberi*, *C. kanran*, and *C. longibracteatum* are classified in one group; *C. ensifolium*, *C. goeringii*, and *C. sinense* are categorized into another group [44].

**Genic-SSR annotation**

Putative functions were assigned to those unigenes containing SSRs by sequence similarities. These unigenes were involved in a wide range of functions, which indicated that these genic-SSRs were likely important biologically characters. For example, unigene containing SSR47 shares homology with CONSTANS-like protein. In Arabidopsis, the CO (CONSTANS) gene has an important role in the regulation of flowering by photoperiod [50]. Unigene containing SSR43 has homology with a glycinebetaine/proline transporter. The accumulation of glycinebetaine (GB) is one of the adaptive strategies to adverse salt stress conditions [51]. The transporters mediate the uptake of GB and/or proline in many plant species e.g. *Arabidopsis thaliana* [52], tomato (*Solanum lycopersicum*) [53], rice (*Oryza sativa*) [54], barley [55]. Unigene having SSR75, was annotated as mitogen-activated protein kinase (MAPK). MAPK cascades function as key signal transducers that use protein phosphorylation/dephosphorylation cycles to channel information [56]. In the plant, MAPKs have been shown to regulate numerous cellular processes, including biotic stress relief [57,58]. Although some unigenes with SSRs had no match to known genes in current gene database, they will likely gain functional annotations as the knowledge of plant genes increases. Compared with anonymous SSRs, genic-SSR markers have a higher probability of being functionally associated with differences in gene expression, which may be in identifying associations between genotype and phenotype. Mapping of genic-SSRs will also provide a map location, in many cases, for genes with known functions.

**Conclusion**

In this work, 7,936 genic-SSRs were identified in *C. ensifolium* transcriptome and their characterizations were further analyzed. A total of 80 genic-SSRs were chosen for validation, and 55 markers successfully yielded polymorphism across 9 *Cymbidium* species including 59 accessions. The high transferability of genic-SSR will be a powerful resource for molecular taxonomic studies and construction of a reference molecular map of the *Cymbidium* genome. Since genic-SSR markers belong to generic regions of the genome, some of these can be exploited for use in marker-assisted breeding of *Cymbidium*. Therefore, the set of genic-SSR markers developed here is a promising genomic resource.

**Additional file**

**Additional file 1: Figure S1.** Polyacrylamide gel electrophoresis profile of SSR62 a and SSR73 b. M: Maker DL2000; 1: 63: cymbidium accession listed in Table 1.

**Competing interests**

The authors declare that they have no competing interests.

**Authors contribution**

XL and FJ carried out the molecular genetic studies, participated in the sequence alignment and drafted the manuscript. LJ, AJ and CH participated in the sequence alignment. XL, FJ, and KL participated in the design of the study and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

**Acknowledgments**

The authors thank Lin Biao for critical review, the lab of Professor Diannong Wu for materials supply and technical support, and Chongbdo Sun for a part of materials supply. This research was supported by the National Basic Research Program funded by the Nature Science Foundation of China (No. 31201648), the Postdoctoral Science Foundation of China (No. 2012 MS21023), the Special Postdoctoral Science Foundation of China (No. 2013 T60607), and the Foundation for Selected Postdoctoral project of Zhejiang (Bs1201032), the Qianjiang talents project (No. 2013R10081), and Scientific and technical innovation promotion project of ZAAS (2012R05Y01E04).

**Author details**

1. Zhejiang Academy of Agricultural Sciences, Shiqiao Road 139, Hangzhou 310021, People’s Republic of China. 2. Hubus University, College of Life Sciences, Wuhan 430602, People’s Republic of China. 3. USDA-ARS, Dale Bumpers National Rice Research Center, Stuttgart, Arkansas 72160, USA. 4. Agricultural Technology Extension Stations, Shaoxing County Agricultural Bureau, Shaoxing 312000, Peoples Republic of China. 5. School of Biological Sciences, The University of Hong Kong, Pokfulam Road, Hong Kong, SAR, People’s Republic of China. 6. State Key Lab of Rice Biology, International Atomic Energy Agency Collaborating Center, Zhejiang University, Hangzhou 310029, Peoples Republic of China.

Received: 14 May 2014 Accepted: 30 October 2014
Published online: 05 December 2014

**References**

1. Liu ZJ, Chen SC, Ru ZZ, Chen LJ: Chinese Cymbidium plants. Beijing: Science; 2006.
2. Dupuy D, Cribb PJ: The genus Cymbidium. London: Christopher Helm; 1988.
3. Obara-Okeyo P, Kako S: Genetic diversity and identification of Cymbidium cultivars as measured by random amplified polymorphic DNA (RAPD) markers. Euphytica 1998, 99:95-101.
4. Wang L-m, Wang S-q, Yang Y-f: Germ Plasm Resources and Breeding of Orchids. J Anyang Institute Technol 2005, 21: 14.
5. Li X, Xiang L, Wang Y, Luo J, Wu C, Sun C, Xie M: Genetic diversity, population structure, pollen morphology and cross- compatibility among Chinese Cymbidiums. Plant Breed 2014, 133:145-152.
6. Escudero A, Iriondo JM, Torres ME: Spatial analysis of genetic diversity as a tool for plant conservation. Biol Conserv 2003, 113:351-365.
7. Ellis JR, Burke JM: EST-SSRs as a resource for population genetic analyses. Heredity 2007, 99:125-132.
8. Choi H, Kim MJ, Lee JS, Ryu KH: Genetic diversity and phylogenetic relationships among and within species of oriental cymbidiums based on RAPD analysis. Sci Hort 2006, 108:79-85.
9. Lu J, Hu X, Liu J, Wang H: Genetic diversity and population structure of 151 Cymbidium sinense cultivars. J Horticulture Forestry 2011, 3:104-114.
10. Capesius I: Isolation and characterization of native AT-rich satellite DNA from nuclei of the orchid Cymbidium, FEBS Lett 1976, 68:255-258.
11. Moe KT, Zhao W, Song HS, Kim YJ, Chung JW, Cho YP, Park PK, Park HS, Chae SC, Park YJ: Development of SSR markers to study diversity in the genus Cymbidium. Biochem Syst Ecol 2010, 38:385-394.
12. Venkateswarlu M, Raje U, S, Surenda Nath B, Shashidhar HE, Maheswaran M, Veeraiah TM, Sathiga M. A first genetic linkage map of mulberry (Morus spp) using RAPD, ISSR, and SSR markers and pseudotetrasomic mapping strategy. Tree Genet Genomes 2006, 3:15. 24.

13. Kenis K, Keulemans J. Genetic linkage maps of two apple cultivars (Malus x domestica Borkh.) based on AFLP and microsatellite markers. Mol Breed 2005, 12(2):15. 19.

14. Rl R, Angiolillo A, Guerrero C, Pellegrini M, Rallo L, Berndar G, Bervall A, Martin A, Baldoni L. A first linkage map of olive (Olea europaea L.) cultivars using RAPD, AFLP, FLPFL and SSR markers. Theor Appl Genet 2003, 106:21. 1282.

15. Li X, Xiang L, Luo J, Hu B, Tian S, Xie M, Sun C. The strategy of RNA-seq, application and development of molecular marker derived from RNA-Seq, Chinese J Cell Biol 2013, 33:55.

16. Li X, Cui H, Zhang M. Molecular markers derived from EST: Their development and applications in comparative genomics. Biodiversity Sci 2006, 14:541. 547.

17. Li X, Luo J, Yani T, Xiang L, Jia F, Qin D, Sun C, Xie M. Deep sequencing-based analysis of the Cymbidium ensifolium floral transcriptome. PLoS One 2012, 8(12):e54860.

18. Grabherr MG, Haja BL, Yassour M, Levin JZ, Thompson DA, Lake I, Adiconis X, Mcclung A, Wu D. Diversity and applications in comparative genomics. MolEcolResour 2008, 8(92): 385.

19. Faircloth BC. Msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. Mol Ecol Resour 2008, 8(92): 292.

20. Rozen S, Skeatsky H. From methods in molecular Biology: Primer3 on the www for general users and for biologist programmers. In Bioinformatics Methods and Protocols volume 32. Edited by Miniseri S, Krawetz SA, New Jersey: Humana Press; 2000:365.

21. Koonin EV, Fedorova ND, Jacobs AR, Krylov DM, Makarova KS, Durek V, Letunic I, Bork P, Ponting CP, Fleischmann D, Wolf YI, Prochnik SE, Koonin EV. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol 2004, 5(7):

22. Li XR, Zhang ML, Cui HR. Data mining for SSRs in ESTs and development of EST-SSR marker in oilseed rape. J MolEcol Cell Biol 2007, 40:137. 144.

23. Li XR, Cui H, Zhang M. Detecting the genetic diversity of Brassica napus by EST-SSRs. J Agric Biotechnol 2007, 15:661. 667.

24. Li ZL, Jakkula RS, Hussey JP, Boerma HR. Deep sequencing-based Data mining for SSRs in ESTs and development of EST-SSR marker in oilseed rape. J MolEcol Cell Biol 2007, 40:137. 144.

25. Koonin EV, Fedorova ND, Jacobs AR, Krylov DM, Makarova KS, Letunic I, Bork P, Ponting CP, Fleischmann D, Wolf YI, Prochnik SE, Koonin EV. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol 2004, 5(7):

26. Li XR, Cui H, Zhang M. Detecting the genetic diversity of Brassica napus by EST-SSRs. J Agric Biotechnol 2007, 15:661. 667.

27. Li ZL, Jakkula RS, Hussey JP, Boerma HR. SSR mapping and confirmation of the QTL from Pyrex135 conditioning soybean resistance to southern root-knot nematode. Theor Appl Genet 2003, 103:1167. 1173.

28. Nei M, Satta K. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. Genetics 1999, 152:1079. 1089.

29. Gupta PR, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Genet Genomics 2003, 270(4):315. 323.

30. Hu H, Li K, Chen K. Analysis of diversity and relationships among Chinese orchid cultivars using EST-SSR markers. Biochem Syst Ecol 2010, 38:93. 102.

31. Wang HZ, Wu ZX, Luji JI, Shi NN, Zhao Y, Zhang ZT, Liu JJ. Molecular diversity and relationships among Cymbidium goeringii cultivars based on inter-simple sequence repeat (SSR) markers. Genetica 2009, 136:391. 399.

32. Nei M, Takezaki N. Evolutionary analysis of the human mitochondrial DNA. MolEcolRessour 1997, 4:1173.

33. Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 2005, 21:128. 129.

34. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 2007, 24(4):1596. 1599.

35. Schneider S, Excoffier L. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. Genetics 1999, 152:1079. 1089.

36. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 2005, 21:3674. 3676.

37. Rl R, Angiolillo A, Guerrero C, Pellegrini M, Rallo L, Berndar G, Bervall A, Martin A, Baldoni L. A first linkage map of olive (Olea europaea L.) cultivars using RAPD, AFLP, FLPFL and SSR markers. Theor Appl Genet 2003, 106:21. 1282.

38. Li X, Xiang L, Luo J, Hu B, Tian S, Xie M, Sun C. The strategy of RNA-seq, application and development of molecular marker derived from RNA-Seq, Chinese J Cell Biol 2013, 33:55.

39. Li X, Cui H, Zhang M. Molecular markers derived from EST: Their development and applications in comparative genomics. Biodiversity Sci 2006, 14:541. 547.

40. Grabherr MG, Haja BL, Yassour M, Levin JZ, Thompson DA, Lake I, Adiconis X, Mcclung A, Wu D. Diversity and applications in comparative genomics. MolEcolResour 2008, 8(92): 385.

41. Koonin EV, Fedorova ND, Jacobs AR, Krylov DM, Makarova KS, Letunic I, Bork P, Ponting CP, Fleischmann D, Wolf YI, Prochnik SE, Koonin EV. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol 2004, 5(7):

42. Li XR, Zhang ML, Cui HR. Data mining for SSRs in ESTs and development of EST-SSR marker in oilseed rape. J MolEcol Cell Biol 2007, 40:137. 144.
56. Hamel LP, Nicole MC, Duplessis S, Ellis BE: Mitogen-activated protein kinase signaling in plant-interacting fungi: distinct messages from conserved messengers. *Plant Cell* 2012, 24(4):1327–1351.

57. Pitzschke A, Schikora A, Hirt H: MAPK cascade signalling networks in plant defence. *Curr Opin Plant Biol* 2009, 12(4):421–426.

58. Andreasson E, Ellis B: Convergence and specificity in the Arabidopsis MAPK nexus. *Trends Plant Sci* 2010, 15(2):106–113.